RDA-UNR | Repositorio de Datos Académicos

UNR Universidad Nacional de Rosario

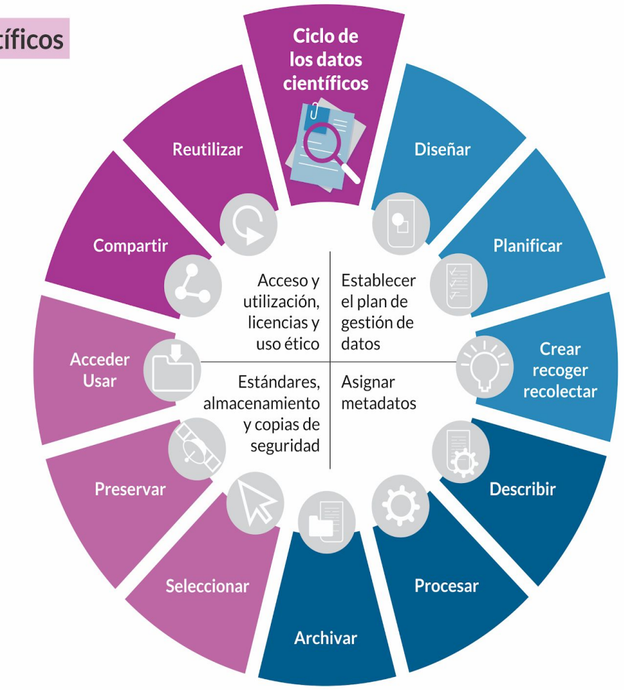
El Impacto de los Repositorios de Datos en la Investigación Científica

CAPACITACIÓN

Curación de Datos en el REPOSITORIO DE DATOS ACADÉMICOS UNR

Plan de Gestión de Datos (PGD)

El ciclo de los datos científicos



Basado en REBIUN

26 de septiembre de 2023

Comité de Acceso Abierto | Unidad de Gestión de Acceso Abierto | Universidad Nacional de Rosario

Ciencia Abierta

Recomendación de Ciencia Abierta de UNESCO (2021)

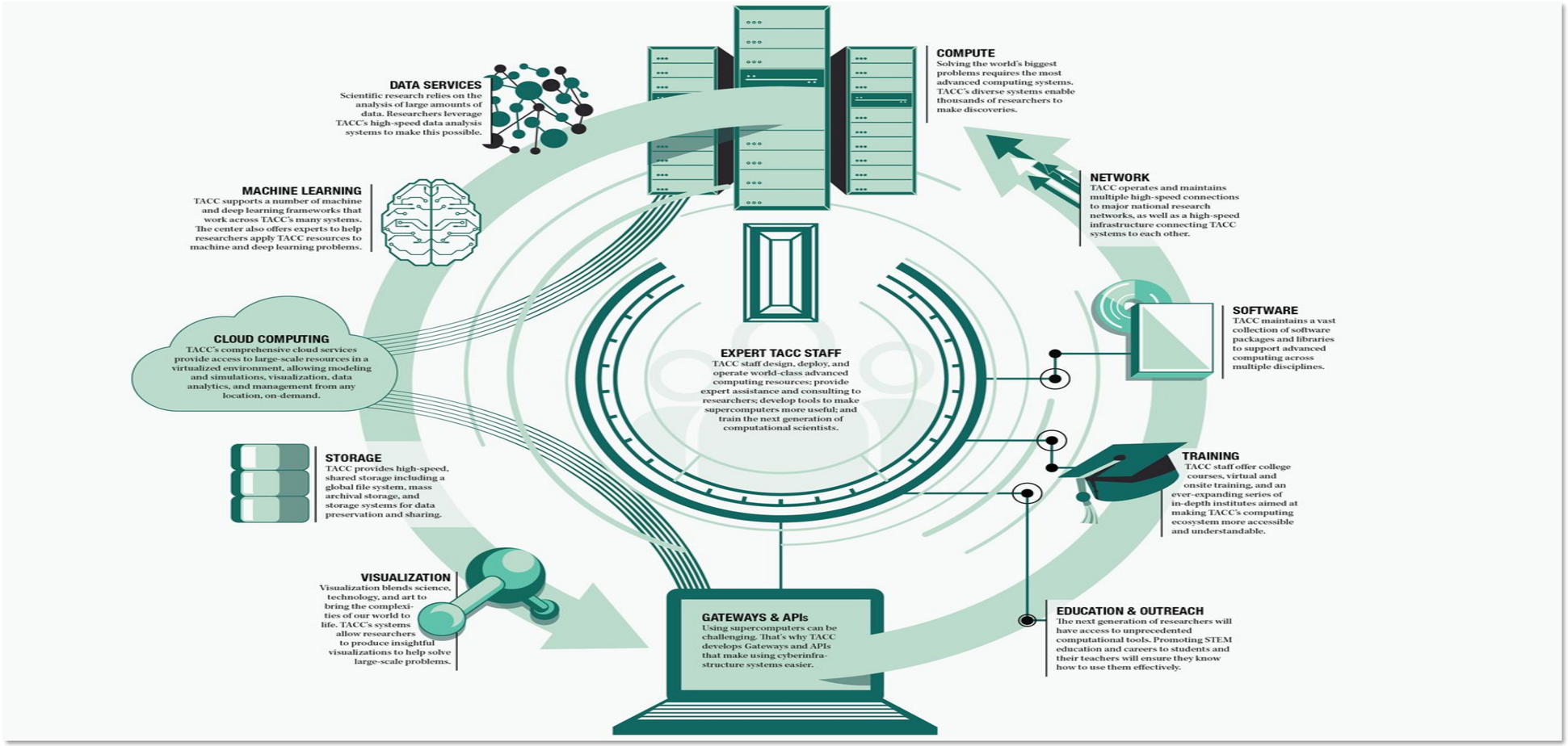
MINCyT (2022). Diagnóstico y lineamientos para una política de ciencia abierta en Argentina

¿Qué son los datos de investigación?

Información recolectada, observada o generada durante el proceso de investigación, que fundamenta las afirmaciones de la investigación y sirve para **validar los resultados**.

Todo dato sobre el que se basa una investigación y que fundamenta un nuevo conocimiento **(Ley 26.899)**

Datos y avance científico



Ejemplo de infraestructura de investigación TACC, Texas Advanced Computing Center Universidad de Texas en Austin

Los repositorios de datos abiertos

RDA-UNR
dataverse-info.unr.edu.ar

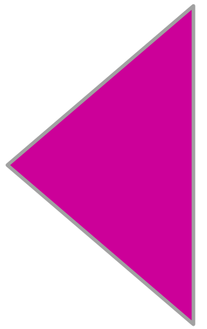
- ✔ Componente de **la infraestructura para investigación**
- ✔ Las bibliotecas académicas aceptaron el desafío
- ✔ Plataforma para datos: Software y almacenamiento seguro
- ✔ Gobernancia: compromiso institucional y personal especializado
- ✔ Estándares y buenas prácticas
- ✔ Distribución global de datasets y Marketing
- ✔ Preservación indefinida de los datos

Tipos de repositorios de datos abiertos

RDA-UNR
dataverse-info.unr.edu.ar

Implementación de repositorios de datos

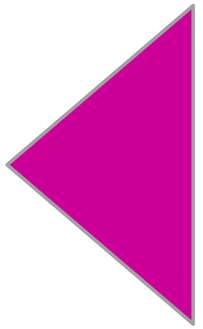
- Bibliotecas académicas
- Repositorios institucionales
- Modelo de auto publicación con guía



Generales

Agencias gubernamentales

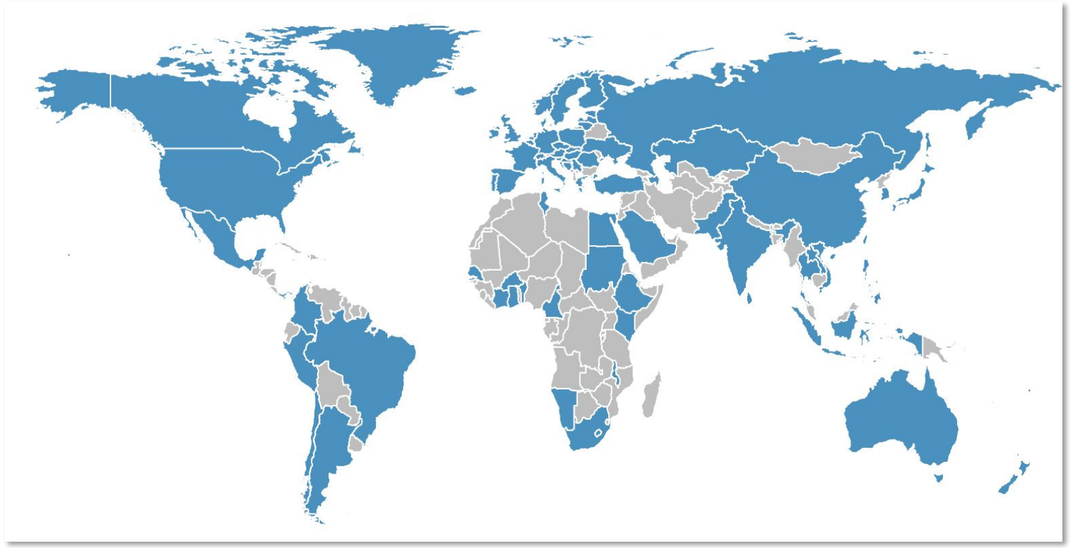
- Curación especializada
- Colecciones especializadas
- Financiamiento para colecciones de datos

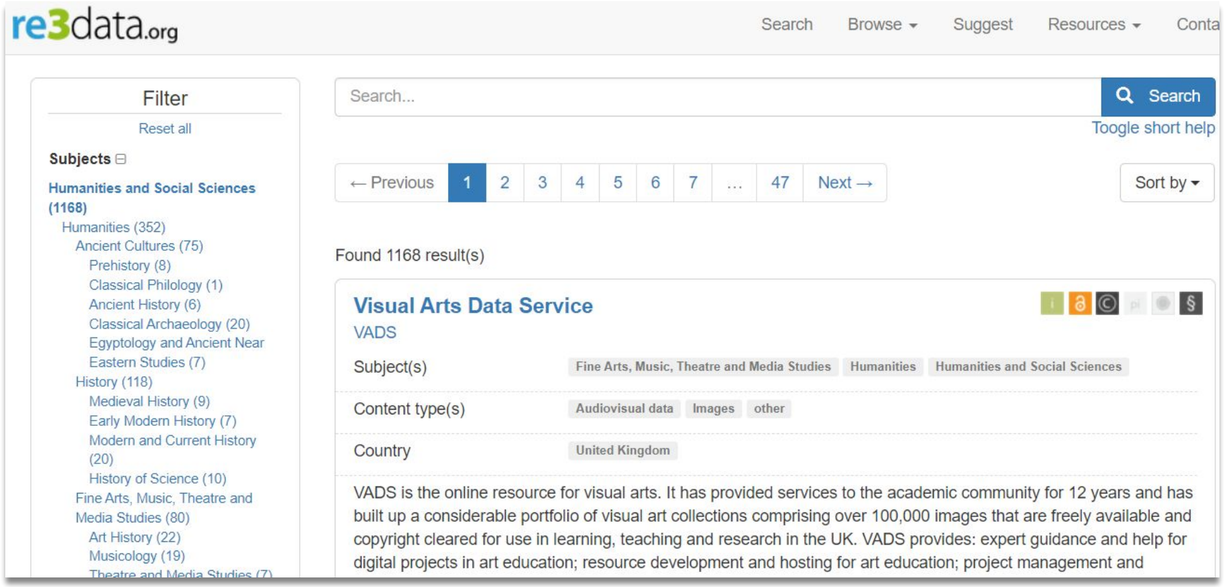


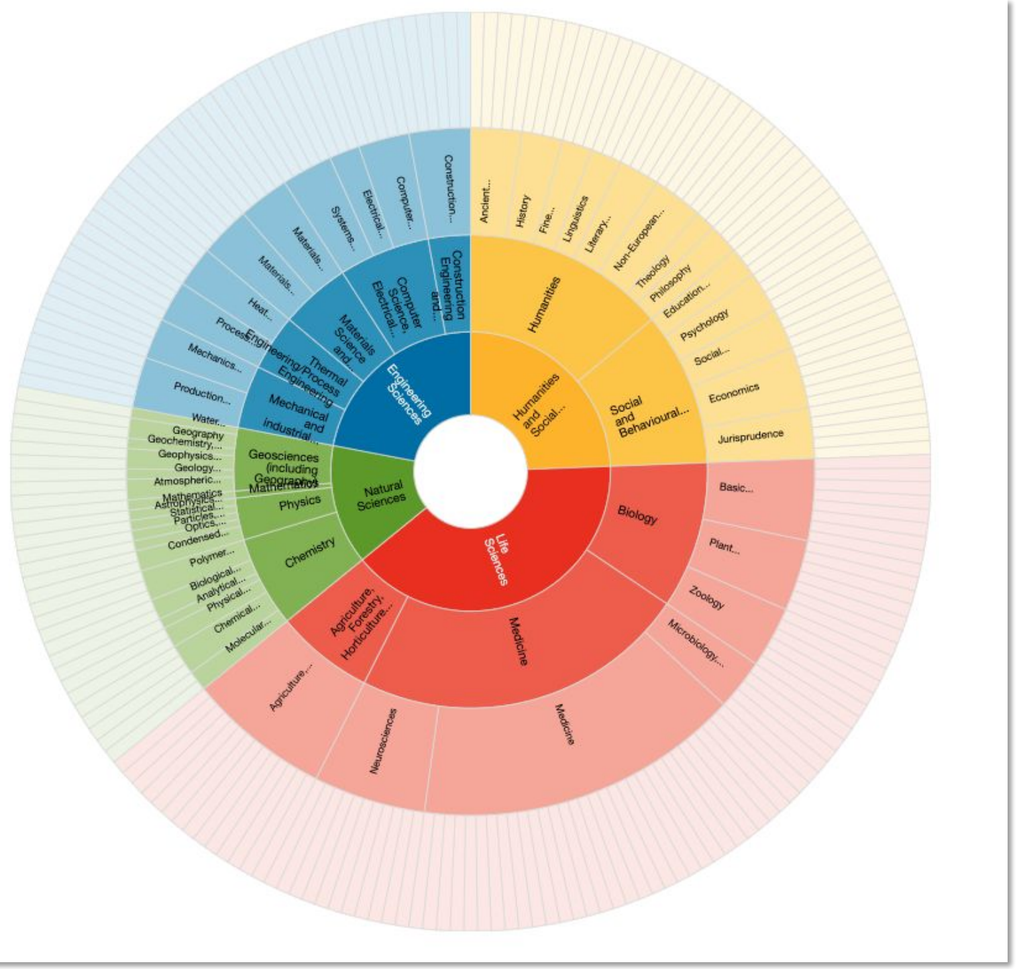
Disciplinarios

Repositorios de datos en diferentes países y disciplinas

RDA-UNR
dataverse-info.unr.edu.ar







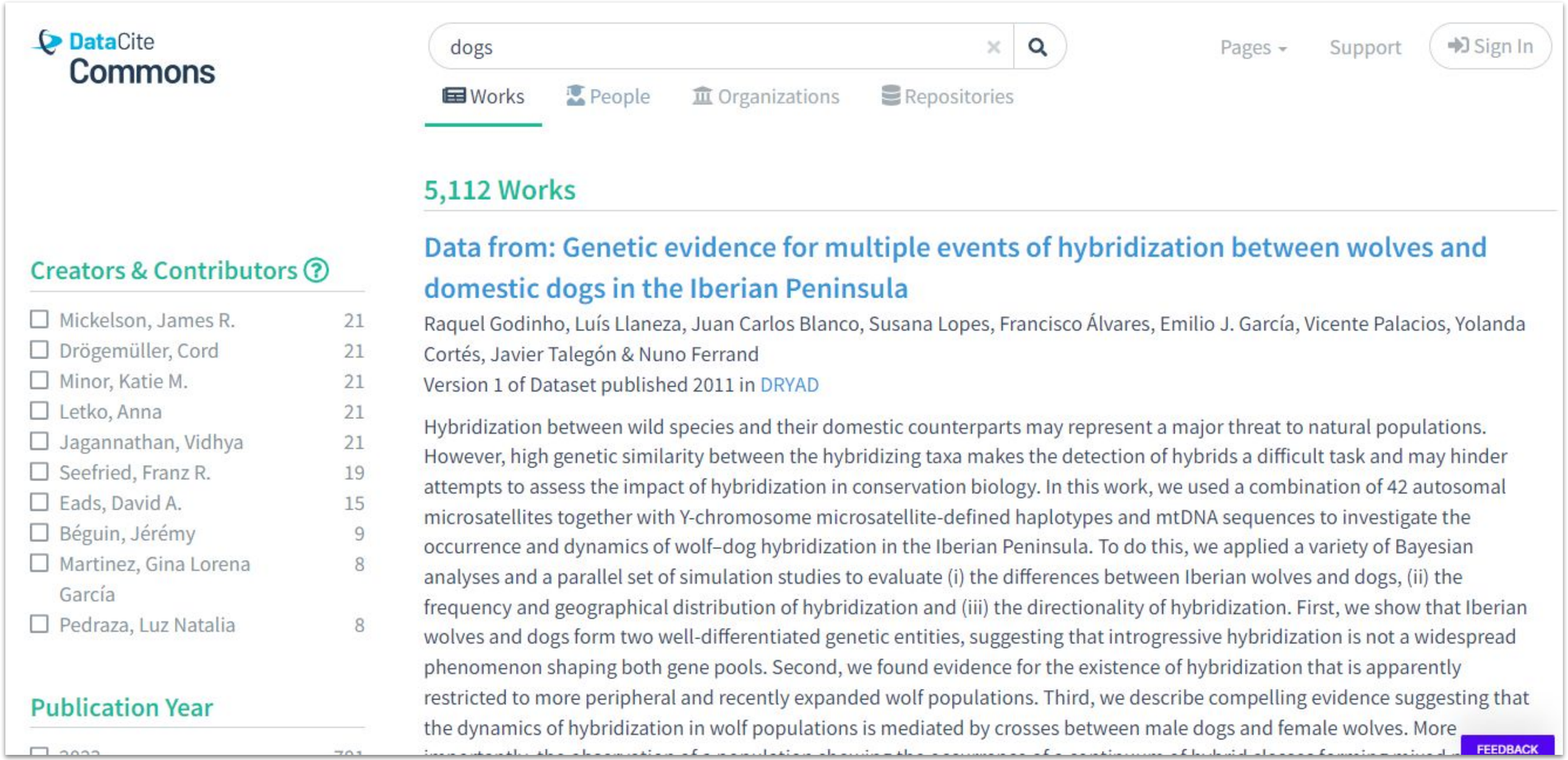
Fuente: re3data https://www.re3data.org/

Repositorios Específicos: OFA Records

- ✔ https://ofa.org/chic-programs/browse-by-breed/
  Gestionado por Orthopedic Foundation for Animals (OFA)
- ✔ Tiene como objetivo Recopilar y difundir información sobre enfermedades ortopédicas y genéticas de los animales.
- ✔ Keller, G. G., Dziuk, E., & Bell, J. S. (2011). How the Orthopedic Foundation for Animals (OFA) is tackling inherited disorders in the USA: Using hip and elbow dysplasia as examples. The Veterinary Journal, 189(2), 197-202. https://doi.org/10.1016/j.tvjl.2011.06.019

Buscar datasets: DataCite Commons

RDA-UNR
dataverse-info.unr.edu.ar



https://commons.datacite.org/doi.org?query=dogs&resource-type=dataset

dataverse.unr.edu.ar

Primer repositorio de datos

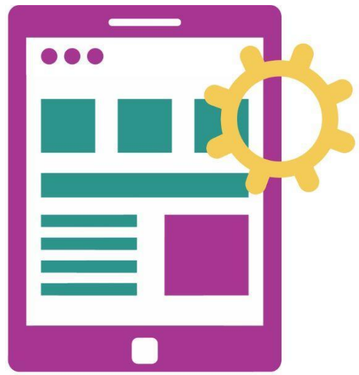
RDA-UNR
dataverse-info.unr.edu.ar

Única universidad en Argentina con Repositorio Institucional dedicado a datos de investigación.

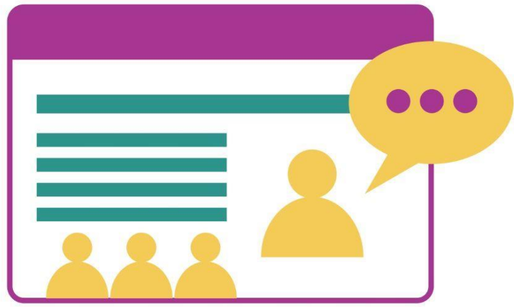
dataverse-info.unr.edu.ar/

Equipos de trabajo + asesoría externa

RDA-UNR
dataverse-info.unr.edu.ar

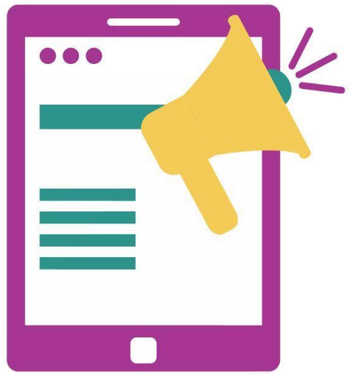


Manual de buenas prácticas de curación de datos



Capacitaciones

Estadísticas



Comunicación

¿POR QUÉ PUBLICAR MIS DATOS?

- ✔ Importancia para la **investigación, la educación y la práctica profesional**
- ✔ **Registro científico** (archivo de ciencias)
- ✔ **Evidencia**
- ✔ **Antecedentes** académicos
- ✔ **Citas** de datos y de artículos
- ✔ **Servicio** público
- ✔ **Economía** de recursos
- ✔ **Formación** de nuevos investigadores

¿QUÉ DATOS VOY A PUBLICAR?

- ✔ Mis datos tienen un **valor** de **reutilización (validar, integrar) y reproducción (valor evidencial)**
- ✔ **Analizar, con ayuda de un curador, el proceso de investigación y determinar cuáles son los fundamentales para reutilizar y/o reproducir.**
- ✔ Los datos deben ser **funcionalmente utilizables**. ¿Se pueden leer y utilizar los datos? ¿Están disponibles las descripciones y son suficientes para permitir que los futuros usuarios comprendan sus datos?
- ✔ ¿Existen aspectos legales o éticos que impidan la publicación de los datos?
- ✔ De acuerdo a la Ley 26.899 hay que publicar los datos luego de 5 años salvo excepciones.

¿Qué piensan los investigadores?

Principales motivaciones para compartir sus datos (n=6.104)

- ✔ **Citación de sus artículos** de investigación (67 %)
- ✔ **Mayor impacto y visibilidad** de sus artículos (61 %)
- ✔ Alguna forma de **beneficio público** (56 %)
- ✔ **Mandato de la revista**/editor (56 %).
- ✔ **Citación de los datos** (54%)
- ✔ **Mayor transparencia y reutilización** (52%)

Science, Digital; Goodey, Gregory; Hahnel, Mark; Zhou, Yuanchun; Jiang, Lulu; Chandramouliswaran, Ishwar; et al. (2022):
The State of Open Data 2022. Digital Science. Report. https://doi.org/10.6084/m9.figshare.21276984.v1

Cada vez más revistas requieren datos abiertos

Las revistas también dan reconocimiento a los artículos que proveen datos abiertos

https://bmcmicrobiol.biomedcentral.com/open-data-badge-articles

Data Availability

The following policy applies to all PLOS journals, unless otherwise noted.

Introduction

PLOS journals require authors to make all data necessary to replicate their study's findings publicly available without restriction at the time of publication. When specific legal or ethical restrictions prohibit public sharing of a data set, authors must indicate how others may obtain access to the data.

When submitting a manuscript, authors must provide a Data Availability Statement describing compliance with PLOS' data policy. If the article is accepted for publication, the Data Availability Statement will be published as part of the article.

Acceptable data sharing methods are listed below, accompanied by guidance for authors as to what must be included in their Data Availability Statement and how to follow best practices in research reporting.

PLOS believes that sharing data fosters scientific progress. Data availability allows and facilitates:

- Validation, replication, reanalysis, new analysis, reinterpretation or inclusion into meta-analyses;
- Reproducibility of research;
- Efforts to ensure data are archived, increasing the value of the investment made in funding scientific research;
- Reduction of the burden on authors in preserving and finding old data, and managing data access requests;
- Citation and linking of research data and their associated articles, enhancing visibility and ensuring recognition for authors, data producers and curators.

Open data badge

BMC Microbiology participated in a pilot project in which all accepted papers were evaluated for eligibility to receive an Open data badge.

Open data badges were assessed based on the content of the article's data availability statement. All relevant information regarding data which has been used or reused to support the central findings for the study should be described in the data availability statement within the manuscript.

The criteria for receiving a badge were as follows:

1. A data availability statement is included with the manuscript, stating how the data can be accessed.
2. The dataset (or part of the dataset) is deposited in a public repository.
3. A DOI, Accession Number or other appropriate persistent identifier is supplied for the dataset.
4. The dataset provided is relevant to the related paper.

Other references to data in the citations or references, or within the body of the article, were not assessed for an Open data badge.

https://journals.plos.org/plosone/s/data-availability

Compartir los datos aumenta las citas

RDA-UNR
dataverse-info.unr.edu.ar

Publicar los datos en repositorios es el único método de compartir datos correlacionado significativamente con el impacto de las citas a los artículos

https://researchdata.springernature.com/posts/how-sharing-your-data-could-increase-your-citations

Colavizza G, Hrynaszkiewicz I, Staden I, Whitaker K, McGillivray B (2020) The citation advantage of linking publications to research data. PLoS ONE 15(4): e0230416. https://doi.org/10.1371/journal.pone.0230416

THE IMPACT OF DATA REUSE

Citations of DesignSafe (updated on March 31st 2023)

Another way of measuring the impact of DesignSafe is by identifying research papers that cite the use of DesignSafe or the data available at DesignSafe. Table 1 lists DesignSafe citations since 2018 as determined from papers identified via Google Alerts. The first column represents papers that make any reference to DesignSafe through citation of the DesignSafe marker paper (Rathje et al. 2017) or through the acknowledgements. The next column represents papers in which a researcher cites their own data in DesignSafe as a part of the original research project, and the third column represents papers that re-use data available in DesignSafe after the original project is over. Note that a paper may contribute to multiple columns in Table 1. For instance, a data re-use paper may also reference the marker paper, or a paper may cite more than one dataset. There is a meaningful number of total citations that reference the use of DesignSafe and the data published in DesignSafe. While Google Alerts may not capture all of the citations and mentions of DesignSafe datasets that are available in the literature, the positive trend highlights the value of publishing data, the importance of citing data in the references using DOIs, and the types of research being conducted using data published in DesignSafe.

Year	DesignSafe Citation	Primary Data Use	Subsequent Data Reuse	Totals
2023	19	42	37	98
2022	65	107	105	277
2021	42	89	60	191
2020	52	74	61	187
2019	21	25	30	76
2018	26	31	13	70

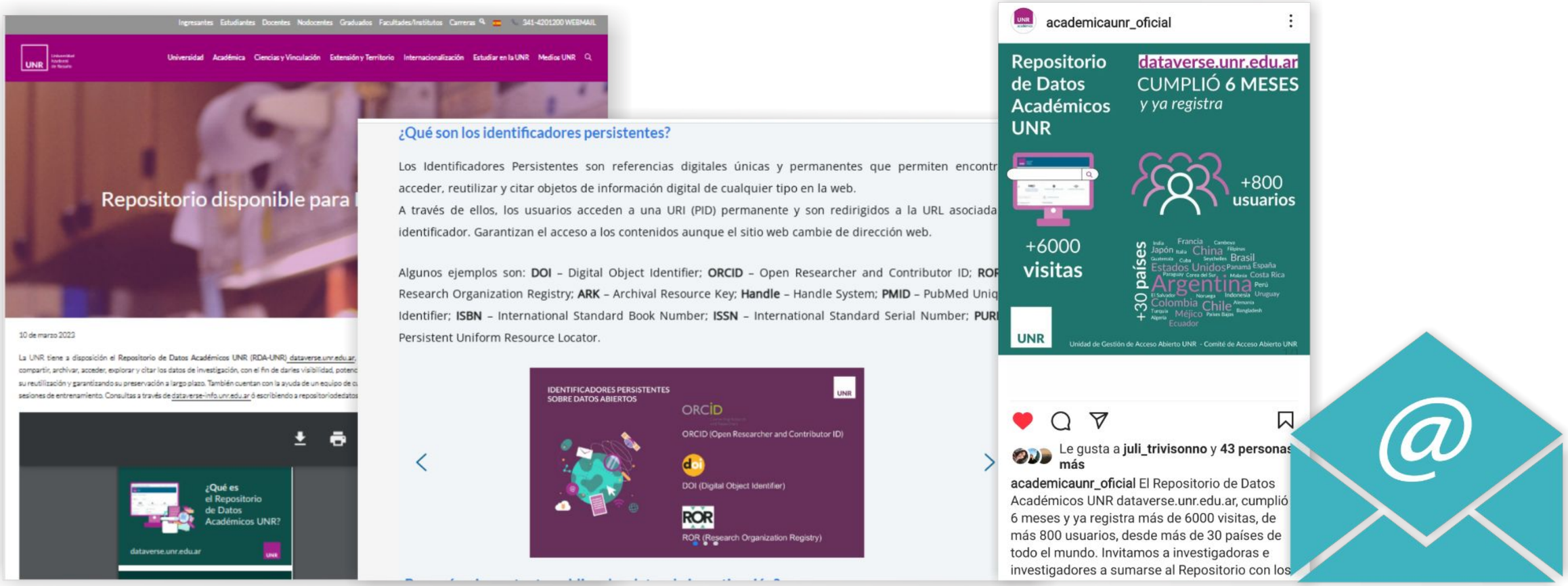
Funte: Repositorio DesignSafe https://www.designsafe-ci.org/rw/impact-of-data-reuse/

Visibilidad e impacto

https://www.instagram.com/academicaunr_oficial/

Difusión

RDA-UNR
dataverse-info.unr.edu.ar



Web UNR
unr.edu.ar
unr.edu.ar/servicio-disponible-para-investigadores-as/

Web RDA-UNR Info
dataverse-info.edu.ar
dataverse-info.unr.edu.ar/?page_id=129#identificadores

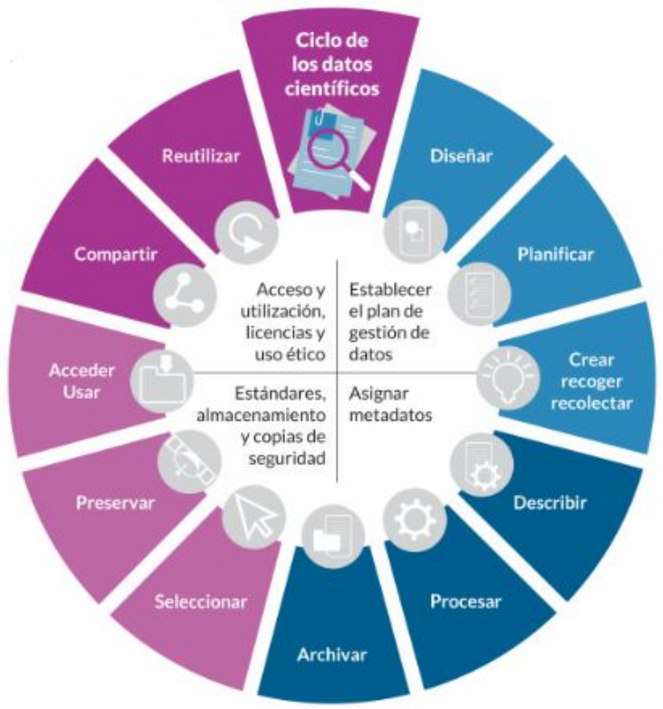
Instagram
@academicaunr_oficial
instagram.com/academicaunr_oficial

Correo Electrónico
repositoriodatos@unr.edu.ar

Google Dataset Search https://datasetsearch.research.google.com/

RDA-UNR
dataverse-info.unr.edu.ar

Plan de Gestión de Datos (PGD)



- ✔ Es un documento formal, elaborado por investigadores, que se presenta con el proyecto de investigación.
- ✔ Describe los aspectos de la gestión de datos, es decir, qué datos se van a generar y qué se hará con los datos durante y después del proyecto de investigación.
- ✔ Objetivo: asegurar que estos datos puedan ser publicados para su reutilización en un plazo no mayor a cinco (5) años a partir del momento de su recolección.
- ✔ En el sitio se puede generar el PGD.

https://dataverse-info.unr.edu.ar/pgd/

ACTIVIDAD

- ✔ ¿Conocías el RDA-UNR? https://www.menti.com/alywpzsef3mu
- ✔ ¿Qué beneficios tiene publicar los datos? Código 3741 9423
- ✔ Registrate en https://dataverse.unr.edu.ar/

UNR

RDA-UNR | REPOSITORIO DE DATOS ACADÉMICOS

https://dataverse-info.unr.edu.ar/

https://dataverse.unr.edu.ar/

http://doi.org/10.17616/R31NJN9K

@academicaunr_oficial

CC BY

Íconos de https://www.freepik.com/

UNR

¡Muchas gracias!

repositoriodatos@unr.edu.ar