

El Impacto de los Repositorios de Datos en la Investigación Científica

CAPACITACIÓN

Curación de Datos en la Facultad
de Ciencias Bioquímicas y Farmacéuticas
REPOSITORIO DE DATOS ACADÉMICOS UNR



2 de junio de 2023



Recomendación de Ciencia Abierta de UNESCO (2021)

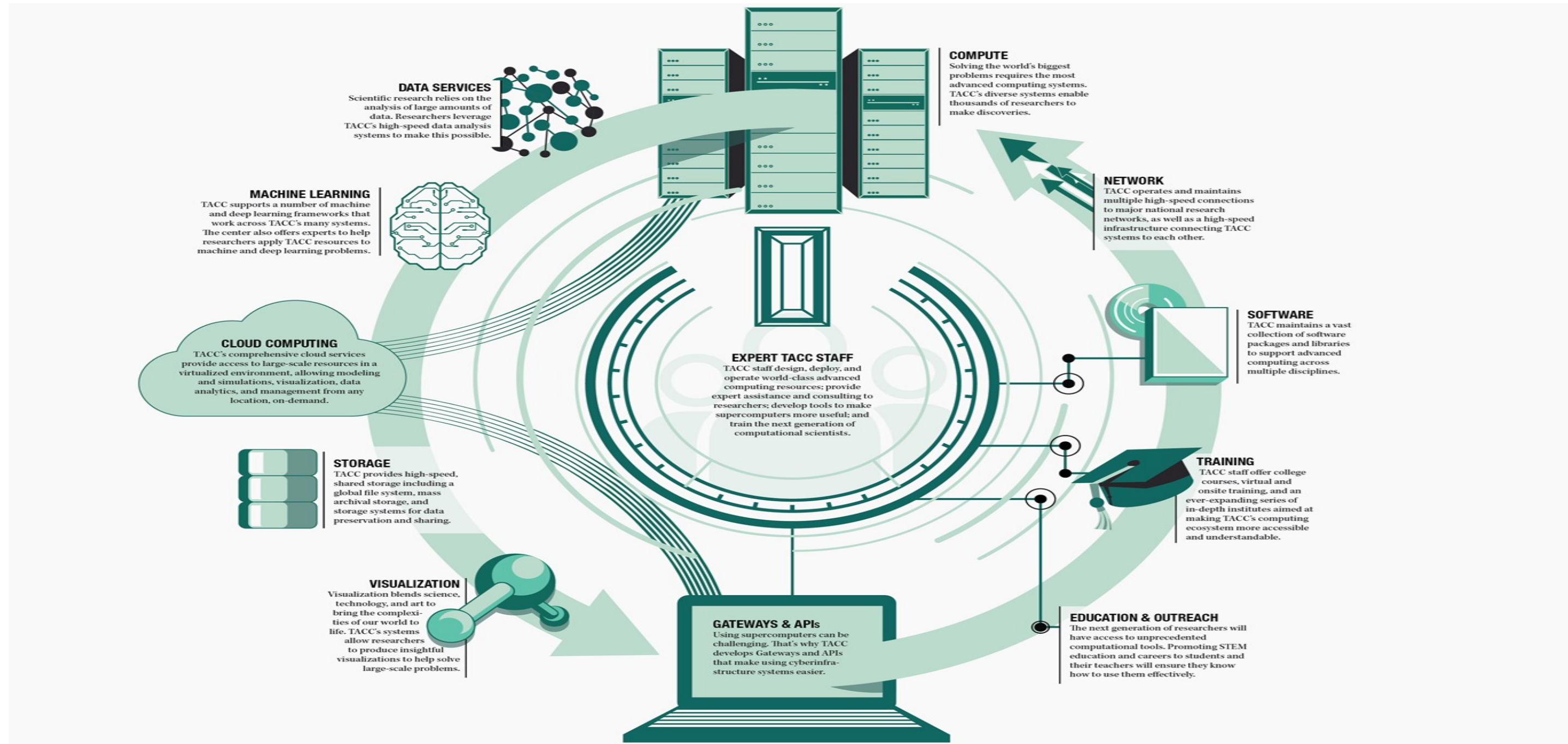
MINCyT (2022). Diagnóstico y lineamientos para una política de ciencia abierta en Argentina

¿Qué son los datos de investigación?



Evidencia generada durante el proceso de investigación que fundamenta las afirmaciones de la investigación y sirve para **validar los resultados.**

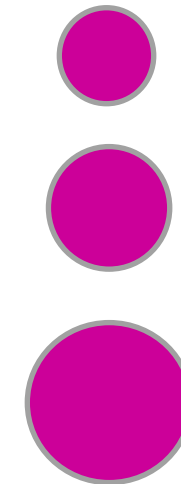
Todo dato sobre el que se basa una investigación y que fundamenta un nuevo conocimiento (**Ley 26.899**)



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

Los repositorios de datos abiertos

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



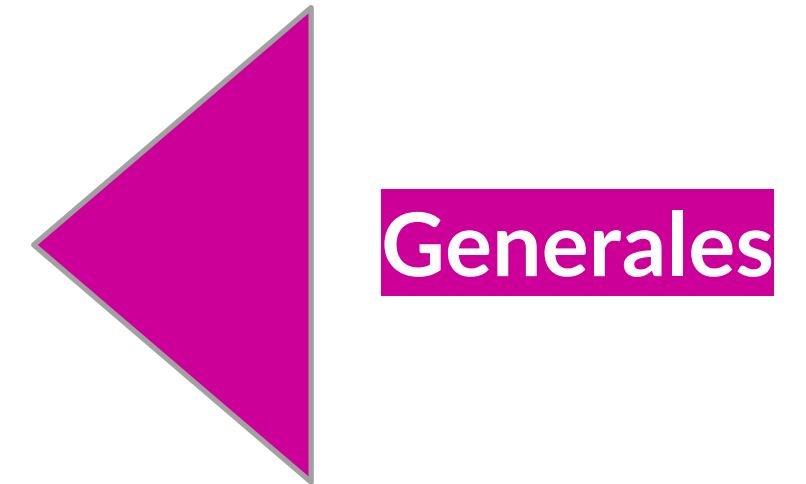
Tipos de repositorios de datos abiertos

Implementación de repositorios de datos

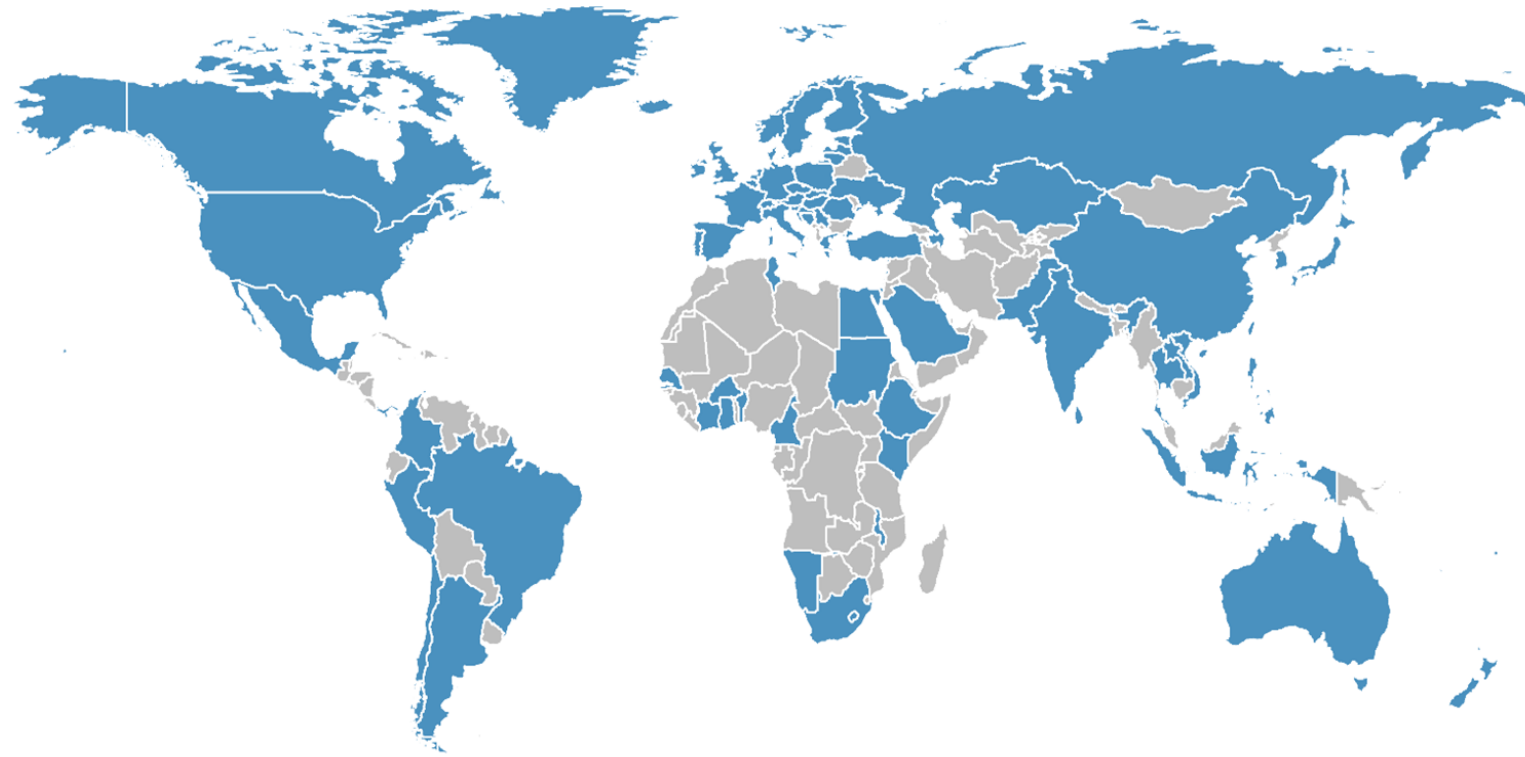
- Bibliotecas académicas
- Repositorios institucionales
- Modelo de auto publicación con guía

Agencias gubernamentales

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



Repositorios de datos en diferentes disciplinas



re3data.org Search Browse Suggest Resources Contact

- Psychology (2)
- Social Sciences (1)
- Economics (1)
- Life Sciences (79)
- Biology (79)
- Basic Biological and Medical Research (79)
- Biochemistry (79)**
- Biophysics (4)
- Cell Biology (8)
- Structural Biology (4)
- General Genetics (20)
- Developmental Biology (3)
- Bioinformatics and Theoretical Biology (30)
- Anatomy (1)
- Plant Sciences (5)
- Plant Ecology and Ecosystem Analysis (1)
- Plant Biochemistry and Biophysics (2)
- Plant Cell and Developmental Biology (2)
- Plant Genetics (2)
- Zoology (8)

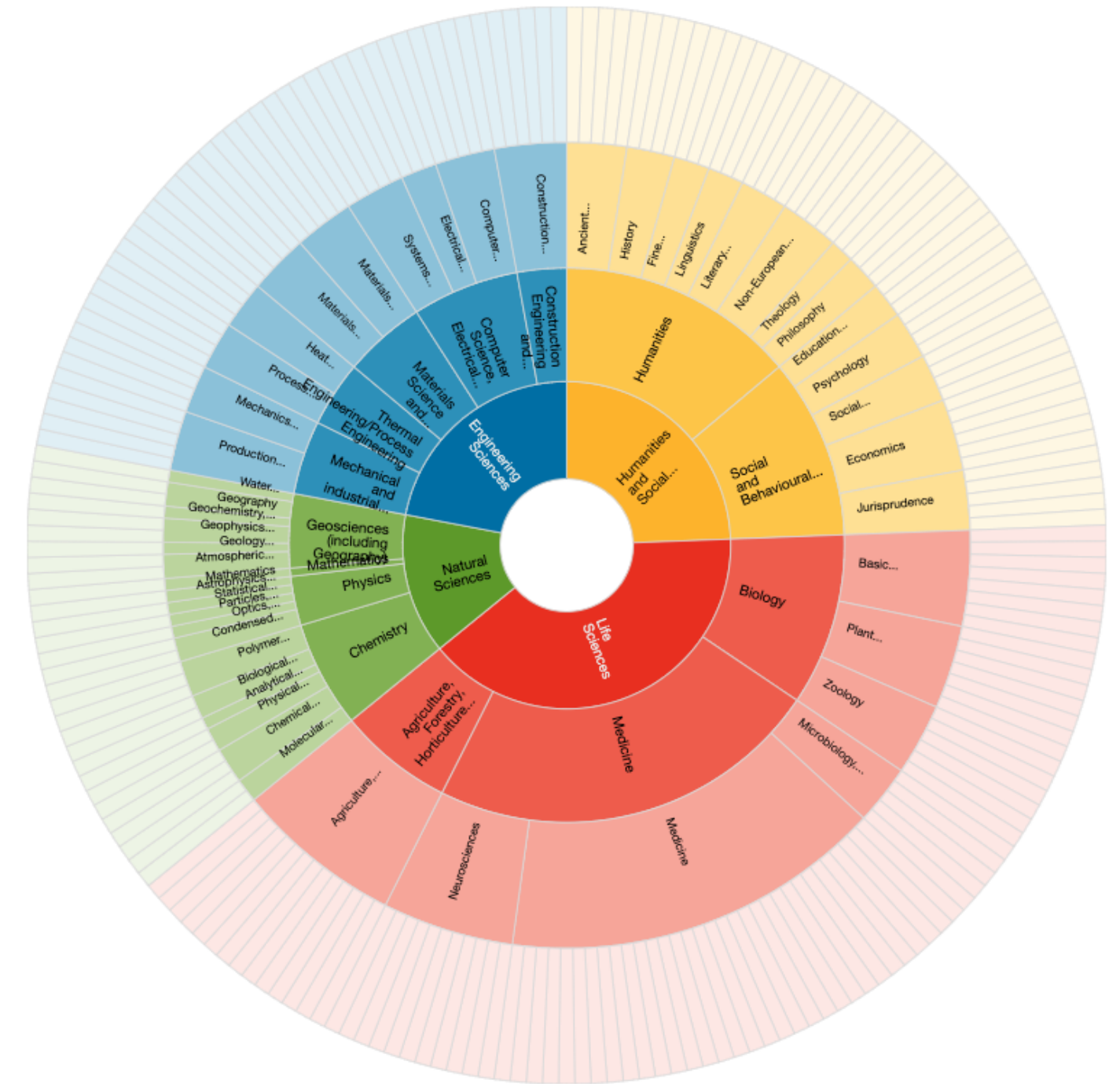
BRENDA
the comprehensive enzyme information system

Subject(s): Basic Biological and Medical Research, Biochemistry, Bioinformatics and Theoretical Biology, Medicine, Biology, Life Sciences

Content type(s): Plain text, Structured graphics, other, Scientific and statistical data formats

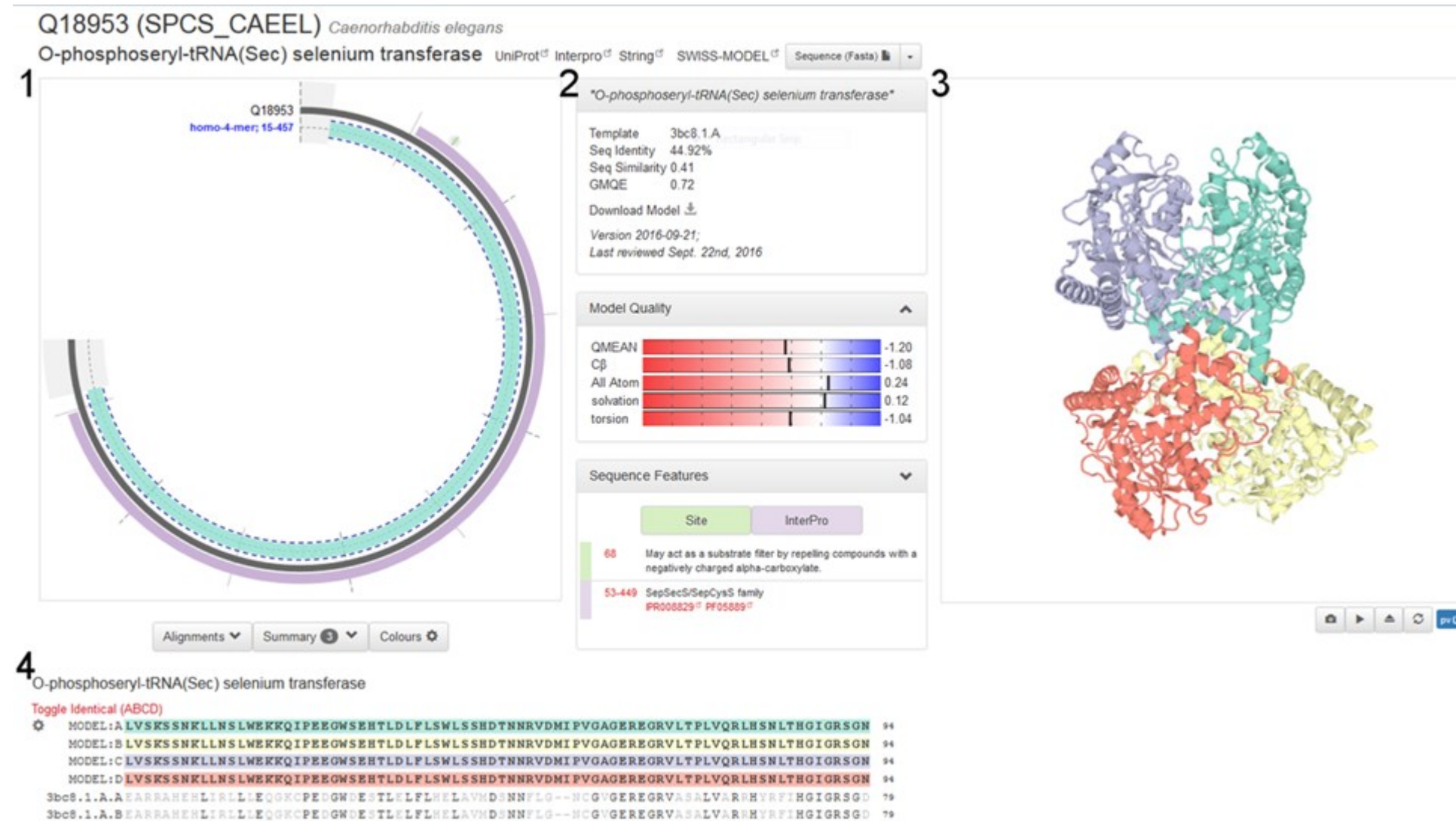
Country: Germany

BRENDA is the main collection of enzyme functional data available to the scientific community worldwide. The enzymes are classified according to the Enzyme Commission list of enzymes. It is available free of charge for via the internet (<http://www.brenda-enzymes.org/>) and as an in-house database for commercial users (requests to our distributor Biobase). The enzymes are classified according to the Enzyme Commission list of enzymes. Some 5000 "different" enzymes are covered. Frequently enzymes with very different properties are included under the same EC number. BRENDA includes biochemical and molecular information on classification, nomenclature, reaction, specificity, functional parameters, occurrence, enzyme structure, application, engineering, stability, disease, isolation, and preparation. The database also provides additional information on ligands, which function as natural or in vitro substrates/products, inhibitors, activating compounds, cofactors, bound metals, and other attributes.



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

Repositorios Disciplinarios: SWISS-MODEL Repository



✓ <https://swissmodel.expasy.org/repository>

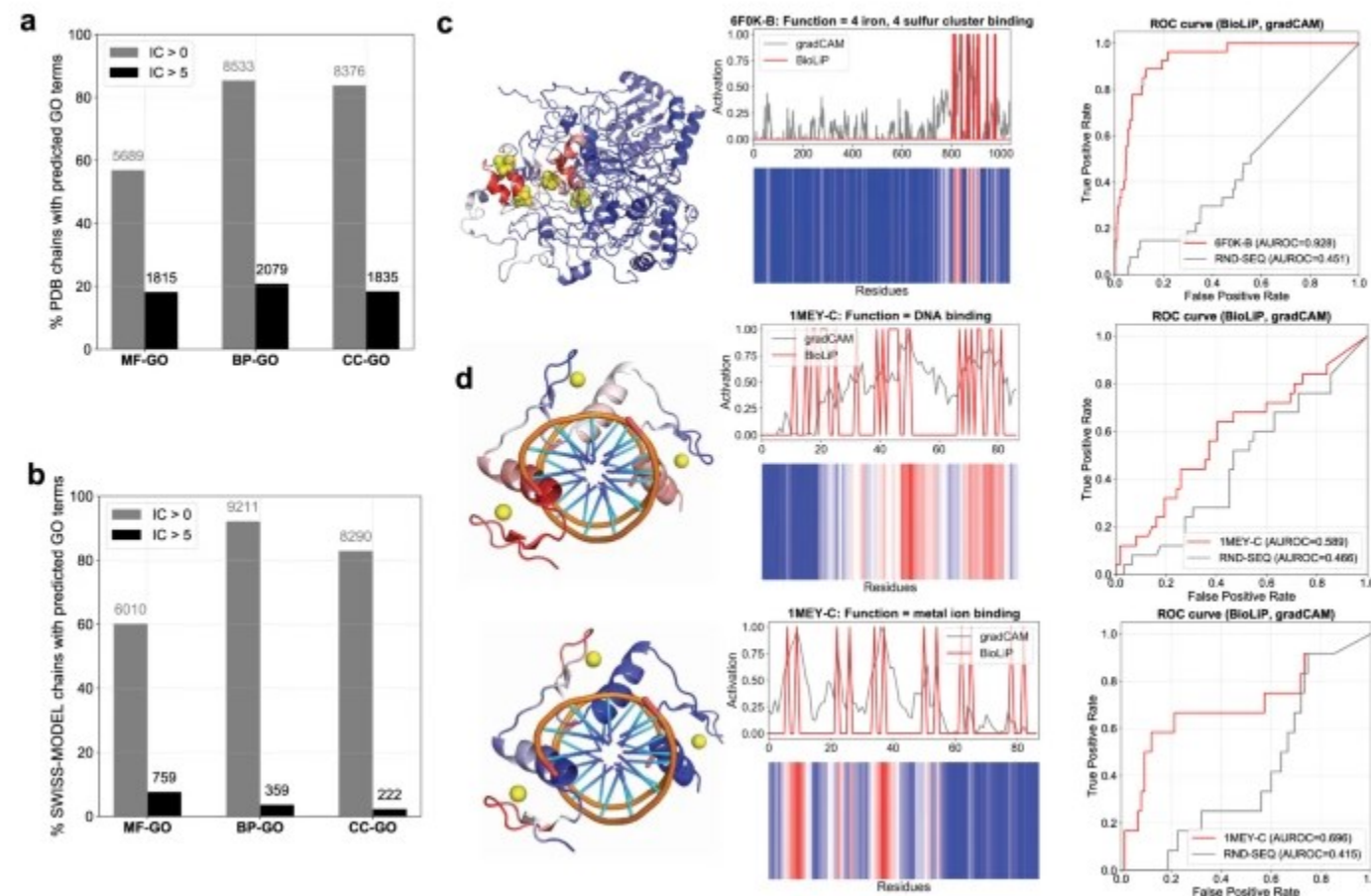
✓ Es una base de datos de **modelos de estructuras proteicas en 3D** anotados, generados por el proceso de modelado por homología de SWISS-MODEL.

✓ Contiene **2.0273.936** modelos de SWISS-MODEL para objetivos de UniProtKB, así como 191,970 estructuras del PDB con mapeo a UniProtKB.

Bienert, S., Waterhouse, A., De Beer, T. A., Tauriello, G., Studer, G., Bordoli, L., & Schwede, T. (2017). The SWISS-MODEL Repository—new features and functionality. *Nucleic acids research*, 45(D1), D313-D319.

<https://doi.org/10.1093/nar/gkw1132>

Fig. 6: Predicting and mapping function to unannotated PDB & SWISS-MODEL chains.



- ✓ Patel, G. B., Rakholiya, P., Shindhal, T., Varjani, S., Tabhani, N. M., & Shah, K. R. (2021). Lipolytic Nocardiosis for reduction of pollution load in textile industry effluent and SWISS model for structural study of lipase. *Bioresource Technology*, 341, 125673. <https://doi.org/10.1016/j.biortech.2021.125673>
- ✓ Saha, O., Islam, I., Shatadru, R. N., Rakhi, N. N., Hossain, M. S., & Rahaman, M. M. (2021). Temporal landscape of mutational frequencies in SARS-CoV-2 genomes of Bangladesh: Possible implications from the ongoing outbreak in Bangladesh. *Virus genes*, 57, 413-425. <https://doi.org/10.1007/s11262-021-01860-x>
- ✓ Gligorijević, V., Renfrew, P. D., Kosciolk, T., Leman, J. K., Berenberg, D., Vatanen, T., ... & Bonneau, R. (2021). Structure-based protein function prediction using graph convolutional networks. *Nature communications*, 12(1), 3168. <https://doi.org/10.1038/s41467-021-23303-9>

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.

The screenshot shows the homepage of the dataverse.unr.edu.ar website. At the top, there is a navigation menu with the following items: Inicio, Acerca de, Equipo, Sobre datos abiertos, Historias de datos, and Novedades. A search icon is located in the top right corner. The main content area features a large infographic with a dark green background. The infographic contains the text: "¿Qué es el Repositorio de Datos Académicos UNR?" and "dataverse.unr.edu.ar" with the UNR logo. Below the infographic, the word "Infografía" is displayed. To the right of the infographic, there is a vertical stack of five blue buttons: "Buscar datos", "Depositar datos", "Guía de uso", "Buenas prácticas", and "Políticas". At the bottom right, there is a purple button that says "CALENDARIO - Consultas de curación powered by Calendly".

dataverse-info.unr.edu.ar/

Equipos de trabajo + asesoría externa



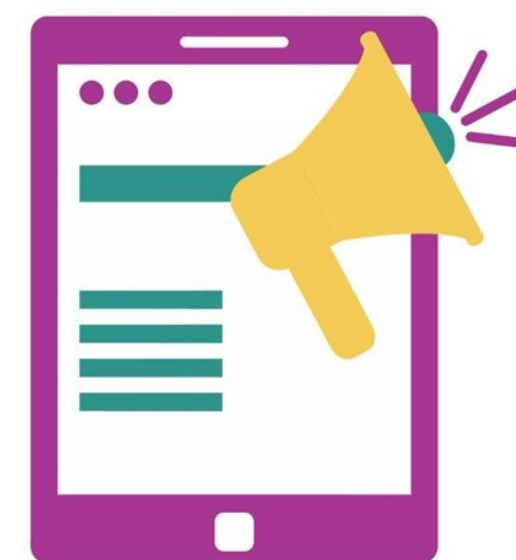
Manual de buenas prácticas de curación de datos



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**



CONSULTA

- ✓ ¿Conocías los repositorios de datos?
- ✓ ¿Publicaste datos en algún repositorio de datos?
- ✓ <https://www.menti.com/aljk64syhwi8>



¿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 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.

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

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.

Compartir los datos aumenta las citas



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/>

Repositorio
de Datos
Académicos
UNR

dataverse.unr.edu.ar
CUMPLIÓ 9 MESES
y ya registra



+1800
usuarios

+11500
visitas

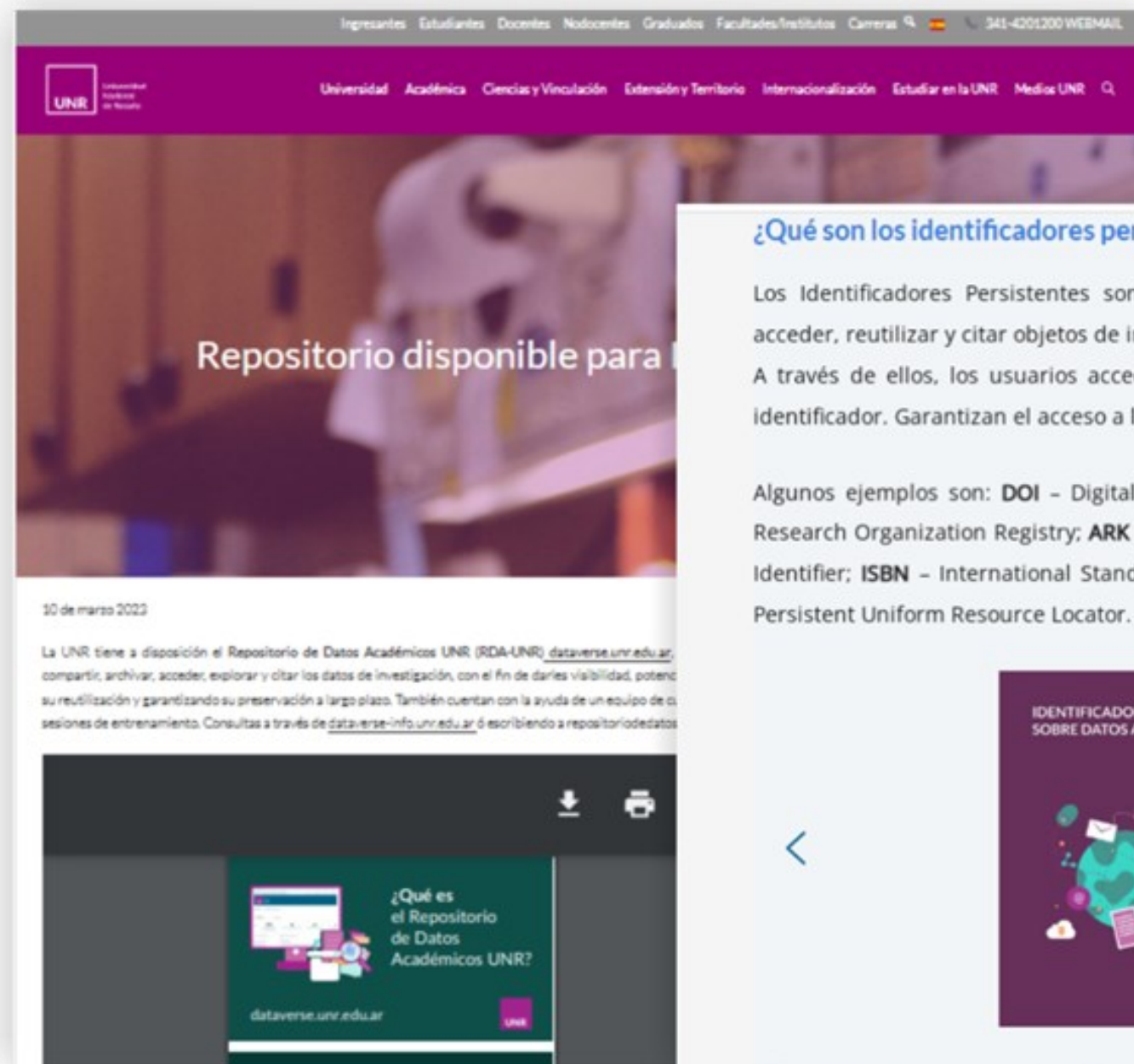
+30 países

India Francia Camboya
Japón Italia China Filipinas
Guatemala Cuba Seychelles Brasil
Estados Unidos Panamá España
Paraguay Corea del Sur Malasia Costa Rica
Argentina Perú
El Salvador Noruega Indonesia Uruguay
Colombia Chile Alemania
Turquía Méjico Países Bajos Bangladesh
Algeria Ecuador



Unidad de Gestión de Acceso Abierto UNR - Comité de Acceso Abierto UNR

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



¿Qué son los identificadores persistentes?

Los Identificadores Persistentes son referencias digitales únicas y permanentes que permiten encontrar, acceder, reutilizar y citar objetos de información digital de cualquier tipo en la web. A través de ellos, los usuarios acceden a una URI (PID) permanente y son redirigidos a la URL asociada al identificador. Garantizan el acceso a los contenidos aunque el sitio web cambie de dirección web.

Algunos ejemplos son: **DOI** – Digital Object Identifier; **ORCID** – Open Researcher and Contributor ID; **ROR** – Research Organization Registry; **ARK** – Archival Resource Key; **Handle** – Handle System; **PMID** – PubMed Unique Identifier; **ISBN** – International Standard Book Number; **ISSN** – International Standard Serial Number; **PURL** – Persistent Uniform Resource Locator.



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

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

Instagram
[@academicaunr_oficial](https://www.instagram.com/academicaunr_oficial)
[instagram.com/academicaunr_oficial](https://www.instagram.com/academicaunr_oficial)

Correo Electrónico
repositoriodatos@unr.edu.ar



Transmembrane Histidine Kinase Functions



Última actualización

Formato de descarga

Derechos de uso

Tema

Gratis

Conjuntos de datos guardados

Se encontraron 14 conjuntos de datos

R Data from: A Transmembrane Histidine Kinase Functions as ...
dataverse.unr.edu.ar

jpeg, png, tiff, tsv

Última actualización: Aug 1, 2022

E Signal transduction histidine kinase, 5TM receptor LytS,...
ebi.ac.uk

Última actualización: Nov 20, 2015

E Histidine kinase CheA-like, homodimeric domain
ebi.ac.uk

Data from: A Transmembrane Histidine Kinase Functions as a pH Sensor



Artículo relacionado

Explorar en: [RDA UNR | dataverse.unr.edu.ar](https://dataverse.unr.edu.ar)

jpeg(41868), tiff(30201006), tsv(2423), tsv(892), png(286408), tsv(1356)

Identificador único

<https://doi.org/10.57715/UNR/PTDCEY>

Actualización del conjunto de datos:

Aug 1, 2022

Conjunto de datos proporcionado por

RDA UNR

Licencia

[Attribution 4.0 \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Se derivó automáticamente la información de la licencia

Testimonios



Dra. Mariana Lagrutta. Docente e investigadora en la UNR. Cátedra de Clínica Médica y Terapéutica y Carrera de Posgrado de Especialización en Clínica Médica



Doctora en Ciencias Biológicas y biotecnóloga, investigadora de UNR-CONICET,. Instituto de Inmunología Clínica y Experimental de Rosario (Idicer)

- ✓ “Permite mostrar que uno **trabaja con seriedad** y que todo lo que nosotros publicamos está **respaldado** en los datos que estamos mostrando.”
- ✓ “La principal ventaja, al tener el estímulo de compartirlo, es que nos obliga a tener sí o sí una **base de datos limpia, prolija y con todos los datos**, de tal manera que **cualquiera los pueda llegar a entender y a utilizar eventualmente**”
- ✓ “Para nosotros las **descargas**, las **citas de esos datos** son importantes, entonces la **visibilidad** es algo que nos beneficia mucho.”
- ✓ “La construcción de la **Historia de Datos** que es poder contar sobre la investigación en otro tipo de términos muy diferentes a lo que estamos acostumbrados pone la relevancia la problemática,... eso también le da **visibilidad y una relevancia social**”

RDA-UNR | REPOSITORIO DE DATOS ACADÉMICOS

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

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

<http://doi.org/10.17616/R31NJN9K>



ACTIVIDAD

- ✓ ¿Conocías el RDA-UNR? <https://www.menti.com/alqg5ddn2kqd>
- ✓ ¿Qué beneficios tiene publicar los datos?
- ✓ Regístrate en <https://dataverse.unr.edu.ar/>



¡Muchas gracias!

repositoriodatos@unr.edu.ar