

Seminario sobre Curación de Datos de Investigación

CAPACITACIÓN

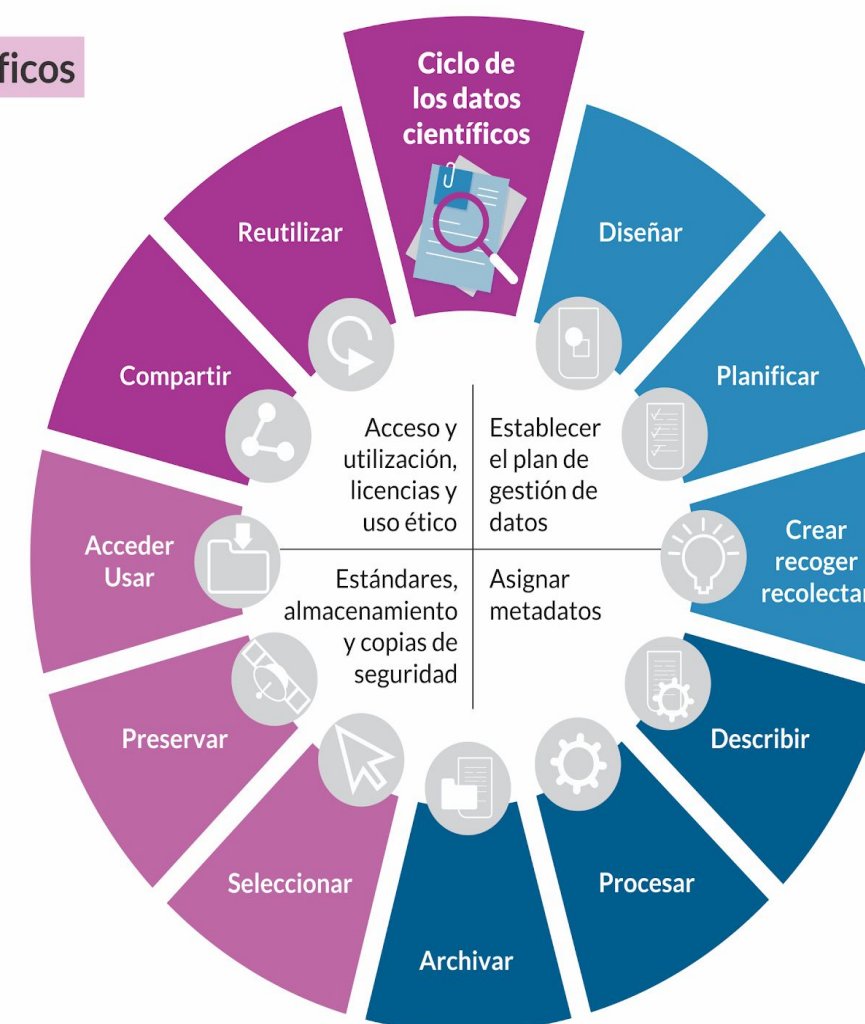
¿Cómo publicar Datos de investigación en el
REPOSITORIO DE DATOS ACADÉMICOS RDA-UNR?

Paola Bongiovani

Facultad de Ciencias Bioquímicas y Farmacéuticas

5 de octubre de 2023

El ciclo de los datos científicos



RDA-UNR Primer Repositorio de Datos en Universidades Argentinas

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

Repositorio
institucional
para datos de
investigación



The screenshot shows the homepage of the RDA-UNR 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 green card with the heading "¿Qué es un repositorio de datos de investigación?". Below the heading, there is a paragraph explaining that it is a technological infrastructure for data preservation and free access. To the right of the text are icons for a computer, a speech bubble with a question mark, and an open book. Below the paragraph is a link labeled "Infografía". On the right side of the page, there is a vertical menu with five buttons: "Buscar datos", "Depositar datos", "Guía de uso", "Buenas prácticas", and "Políticas".

Inicio Acerca de Equipo Sobre datos abiertos Historias de datos Novedades

¿Qué es un repositorio de datos de investigación?

Es una **infraestructura tecnológica** que permite la **preservación** y el **acceso libre** a los datos a través de Internet.

Debe cumplir con ciertos **requisitos**: que los datos se puedan encontrar, que sean accesibles, interoperables y reutilizables.

[Infografía](#)

Buscar datos

Depositar datos

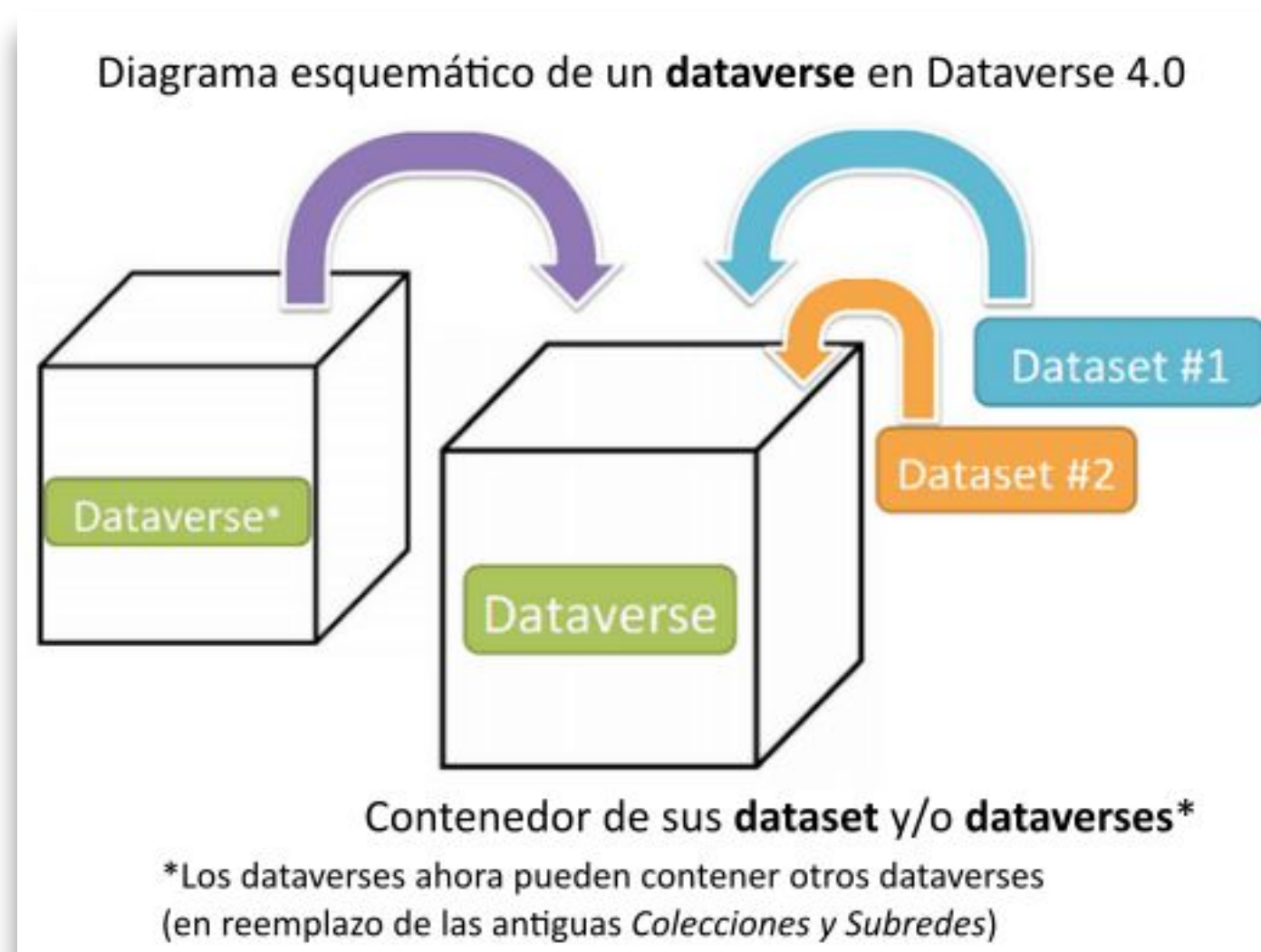
Guía de uso

Buenas prácticas

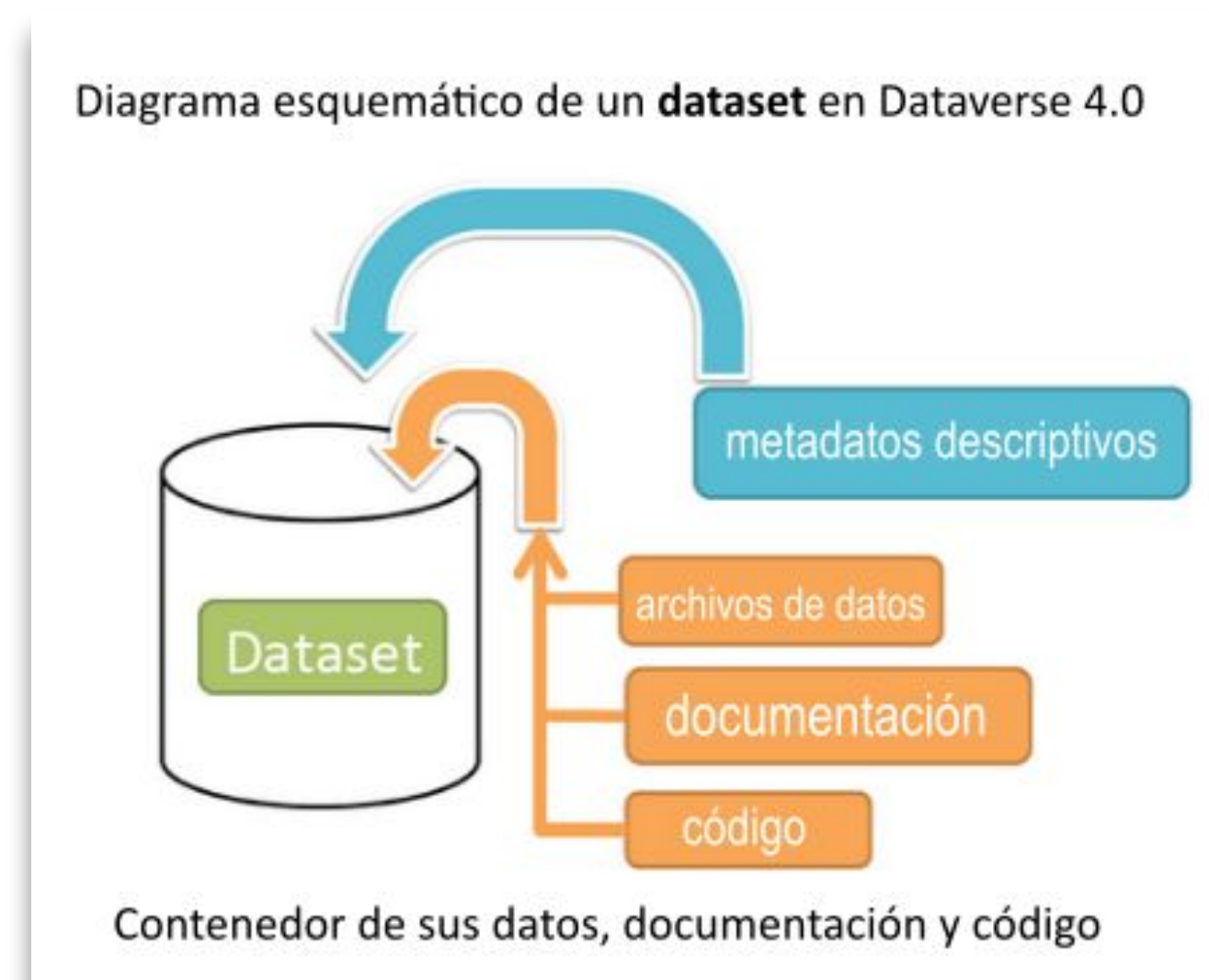
Políticas

¿Cómo se organiza RDA-UNR?

- ✓ **Dataverse** son las colecciones (Facultades, Institutos, etc.)



- ✓ **Datasets** son conjuntos de datos



Cumplimos con los Principios de datos FAIR

✓ Localizables

- Que tengan identificadores persistentes DOIs
- Metadatos estandarizados
- Que se puedan encontrar en buscadores (Google)

✓ Accesibles

- ¿Son gratuitos? ¿Tiene limitaciones de uso?
- ¿Se pueden descargar? ¿Se pueden abrir/ver?

✓ Interoperables

- Protocolos de interoperabilidad
 - ¿Están en un formato abierto? ¿Pueden ser fácilmente utilizados por humanos y computadoras? ¿Requieren diccionarios para ser entendibles
- Los (meta) datos usan vocabularios que siguen los principios FAIR

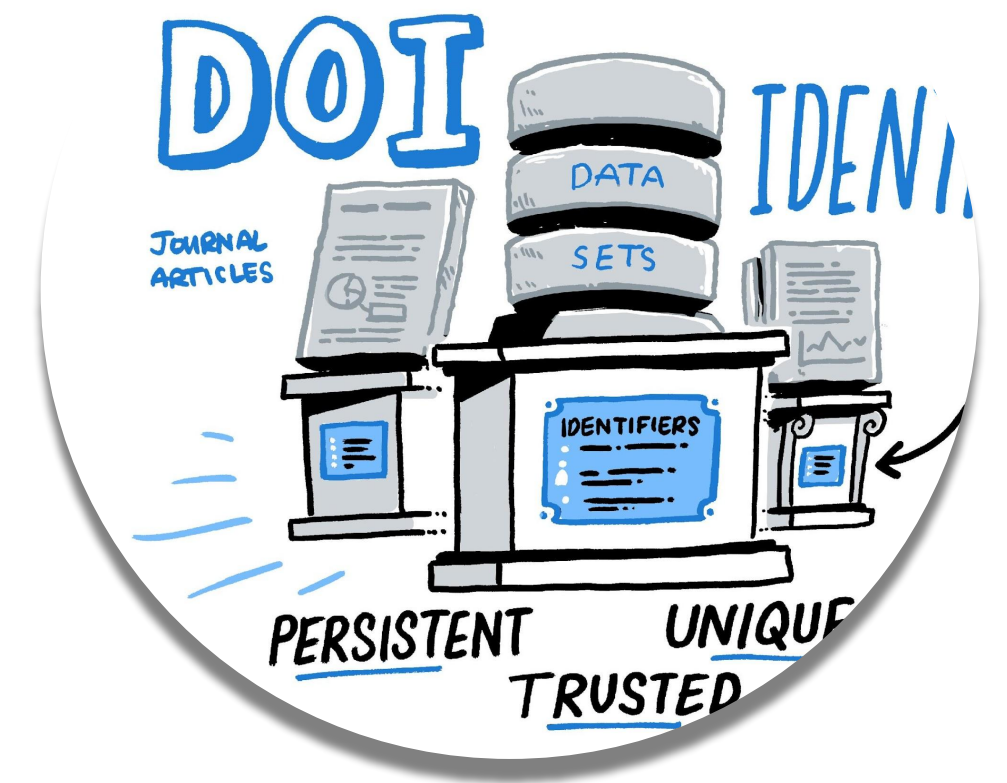
✓ Reutilizables

- Calidad de los datos
- Completos
- Entendibles
- Licencias abiertas



Acceso a los datos

- Los datos que se suben a RDA-UNR obtienen un DOI
 - identificador persistente
- Los datos se pueden citar. Ver “Compromiso de citar datos”
- Los datos se pueden encontrar en Google y otros buscadores específicos de datos.
- Los datos se pueden mantener en el repositorio sin publicar por un tiempo limitado (embargo).



¿Por qué debo tener un identificador ORCID?

- ✓ ORCID es un acrónimo de 'Open Researcher and Contributor ID'.
- ✓ ORCID es un **identificador único** de larga duración para investigadores, una especie de DNI de investigador.
- ✓ Permite **recopilar y encontrar** todos tus trabajos de investigación relacionados para que puedan ser usados y citados.
- ✓ Para registrarse ir a <https://orcid.org/signin>





Documentar la investigación

- ✓ La documentación explica **cómo y por qué** se realizaron ciertos pasos o se tomaron ciertas decisiones sobre otras.
- ✓ Lo que es obvio ahora **puede no ser intuitivo para otras personas** o incluso para nosotros en el futuro.
- ✓ Documentar estas decisiones es de gran ayuda en la **reutilización y validación** de datos.

Metadatos de citas ^

Título * ?

Título del Dataset

Subtítulo ?

Título alternativo ?

URL alternativa ?

Ingresar la URL completa: http://

Otro ID ?

Agencia ?

Identificador ?

Autor * ?

Nombre * ?

Afiliación ?

Apellido, Nombres

Universidad Nacional de Ros

Esquema de identificadores ?

Identificador ?

ORCID

https://orcid.org/0000-0000-0

Contacto * ?

Nombre ?

Afiliación ?

¿Qué son los Metadatos?

- ✓ “Es información descriptiva sobre los datos, su contexto, calidad, condición o características de un dato u objeto de información, que tiene la finalidad de facilitar su búsqueda, recuperación, identificación, autenticación, evaluación, preservación y/o interoperabilidad” Resolución 753 – E/2016. Reglamentación de la Ley 26.899

Guía para escribir una correcta descripción de datos (resumen)

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

¿Qué es una descripción de datos?

Es una **narración** con estilo directo, convincente y atractivo.

Escrita en inglés y/o español, con una extensión máxima de **300 palabras**.

Enfocada en **captar el interés** de otros usuarios para que profundicen en su estudio y posible reutilización.



¡IMPORTANTE! Documentación de los datos

- ✓ **Piensen en la información que necesitaría un investigador externo al proyecto para entender, replicar o reutilizar los datos.**
- ✓ **Documentación a nivel de proyecto:** incluye información sobre el estudio realizado, cuáles fueron las preguntas de investigación, el tipo de datos que se recopilaron para responder a estas preguntas (metadatos).
- ✓ **Documentación a nivel de los datos:** incluye información sobre los archivos de datos individuales, como la transcripción de una entrevista, las variables particulares (diccionario de datos, libro de códigos, etc.) junto con los datos.
- ✓ **Revisar la Guía de Curación/Buenas prácticas** para conocer los formatos recomendados, identificador ORCID y otros aspectos importantes.

¿Por qué usar formatos abiertos?

- ✓ Porque **potencian el acceso y la reutilización.**
- ✓ Porque **permiten la preservación a largo plazo al usar estándares.**
- ✓ Para los formatos abiertos existen **documentos de especificación disponibles gratuitamente.**
- ✓ Para **contribuir a la ciencia abierta.**

Ejemplos:

- Imágenes: TIFF, JPEG 2000, PDF, PNG, GIF, BMP, SVG
- Datos tabulares: CSV, TXT
- Texto: XML, PDF/A, HTML, ASCII, UTF-8



¿Qué incluimos junto con los datos? Diccionario de datos

✓ Los diccionarios de datos brindan información crítica sobre los datos al describir los nombres, definiciones y atributos de los elementos de datos en el archivo de datos.

✓ Libros de códigos (encuestas). Ejemplos

<https://ddialliance.org/resources/markup-examples/codebooks>

✓ Más información sobre tablas Broman, K. W., & Woo, K. H. (2018). Data organization in spreadsheets. *The American Statistician*, 72(1), 2-10.

<https://doi.org/10.1080/00031305.2017.1375989>

	Variable	Variable Label
1	ANTIFUNGICODIACERO	use of antifungals at day 0
2	ANTIFUNGICOEV	use of antifungals during evolution
3	ANTIVIRALESDIACERO	use of antiviral at day 0
4	ANTIVIRALESSEV	use of antiviral during evolution
5	ATBEMPIRICODIACERO	Inicial antibiotic schedule other than vancomicine
6	CATETERCENTRALCERO	CVC-associated infection at onset
7	CATETERCENTRALEV	CVC-associated infection durin evolution
8	CATETEROVIACENTRAL	Central nervous catheter at the onset of FNE
9	CausaNeutropeniaDesconocida	unknown cause of neutropenia (unralated to chemotherapy)
10	CausaOncologica	cancer patients
11	COMORBILIDADDBT	Diabetes
12	COMORBILIDADEPOC	chronic obstructve pulmonary disease
13	COMORBILIDADHEMODIALISIS	hemodialysis
14	COMORBILIDADIC	congestive heart failure
15	COMORBILIDADIRC	chronic renal insufficiency
16	COMORBILIDADTTOGCTCCRONICO	glucocorticoid therapy

Parodi, Roberto Leandro; Lagrutta, Mariana; Tortolo, Mauro; Navall, Estefanía; Rodríguez, María S.; Sasia, Gervasio Flavio; De Candia, Lucas F.; Gruvman, Matias A.; Bottasso, Oscar; Greca, Alcides Alejandro, 2022, "Replication Dataset for: A multi-center prospective study of 515 febrile neutropenia episodes in Argentina during a 5-year period", <https://doi.org/10.57715/UNR/KNSPG1>, RDA UNR, V1,

¿Qué incluimos junto con los datos? Readme file

- ✓ README file tiene el propósito de brindar un panorama sobre los datos en este dataset. Se adjunta junto con los datos (readme.txt)
- ✓ Un archivo README es recomendado para una colección de datos, como un directorio para un proyecto o experimento específico, una herramienta de software o cualquier dato que esté relacionado entre sí "lógicamente".

```
dataset-jaw-movements
=====

*Title:* Audio recordings dataset of grazing jaw movements in dairy
cattle\
*Authors:* Sebastián R. Vanrell, José O. Chelotti, Leandro A. Bugnon, H.
Leonardo Rufiner, Diego H. Milone, Emilio A. Laca and Julio R. Galli\
*Journal:* Data in Brief\
*DOI:* https://doi.org/10.1016/j.dib.2020.105623

sinc(i) (Research institute for signals, systems and computational
intelligence) - http://www.sinc.unl.edu.ar

\* Corresponding author: srvanrell@sinc.unl.edu.ar

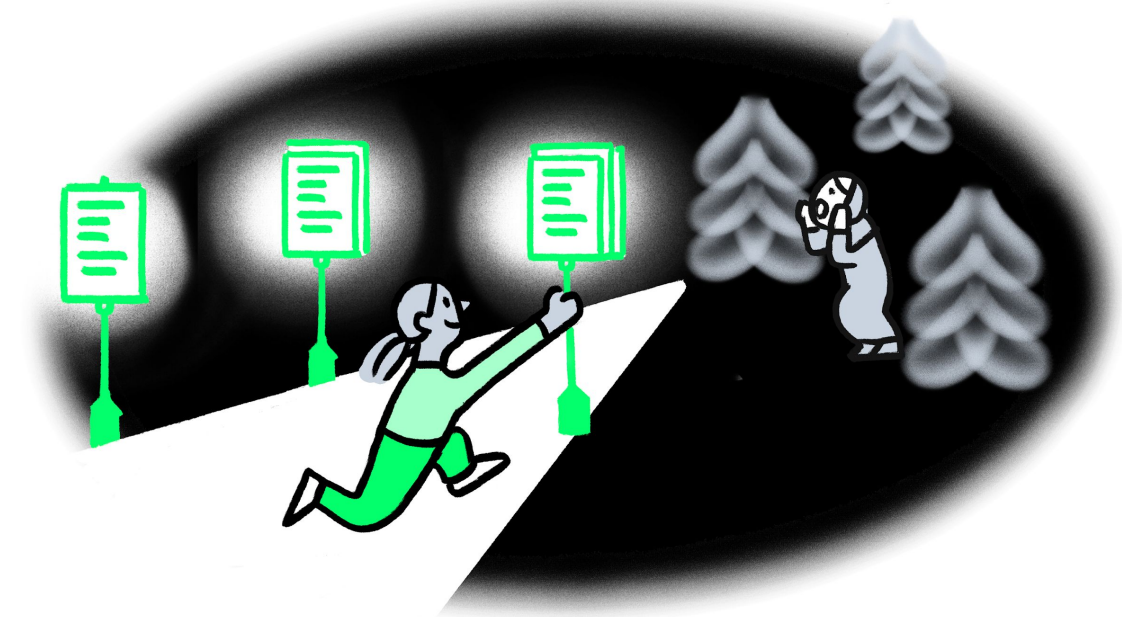
Description
-----

This dataset is composed of correlated audio recordings and labels of
ingestive jaw movements performed during grazing by dairy cattle. Using
a wireless microphone, we recorded sounds of three Holstein dairy cows
grazing short and tall alfalfa and short and tall fescue. Two experts in
grazing behavior identified and labeled the start, end, and type of each
jaw movement: bite, chew, and chew-bite (compound movement). For each
```

Vanrell, Sebastián Rodrigo; Chelotti, José; Bugnon, Leandro; Rufiner, Hugo Leonardo; Milone, Diego; Laca, Emilio; Galli, Julio R., 2022, "Dataset: Audio recordings of grazing jaw movements in dairy cattle", <https://doi.org/10.57715/UNR/T7SDAX>, RDA UNR

¿Qué incluimos junto con los datos?

- ✓ Cuestionarios en blanco.
- ✓ Consentimiento informado.
- ✓ Listas de datos (por ejemplo listado de registros audiovisuales, fechas, lugares, etc.)



Scriberia

Cuestionario sobre Percepción e Implementación de Gobierno Abierto en Municipios de la República Argentina

El presente archivo contiene las 33 preguntas del cuestionario sobre la percepción e implementación del gobierno abierto en los Municipios de más de 50.000 habitantes de la República Argentina, respondido entre marzo de 2019 y mayo 2020 por referentes/responsables de gobierno abierto, cuyas referencias personales han sido anonimizadas.

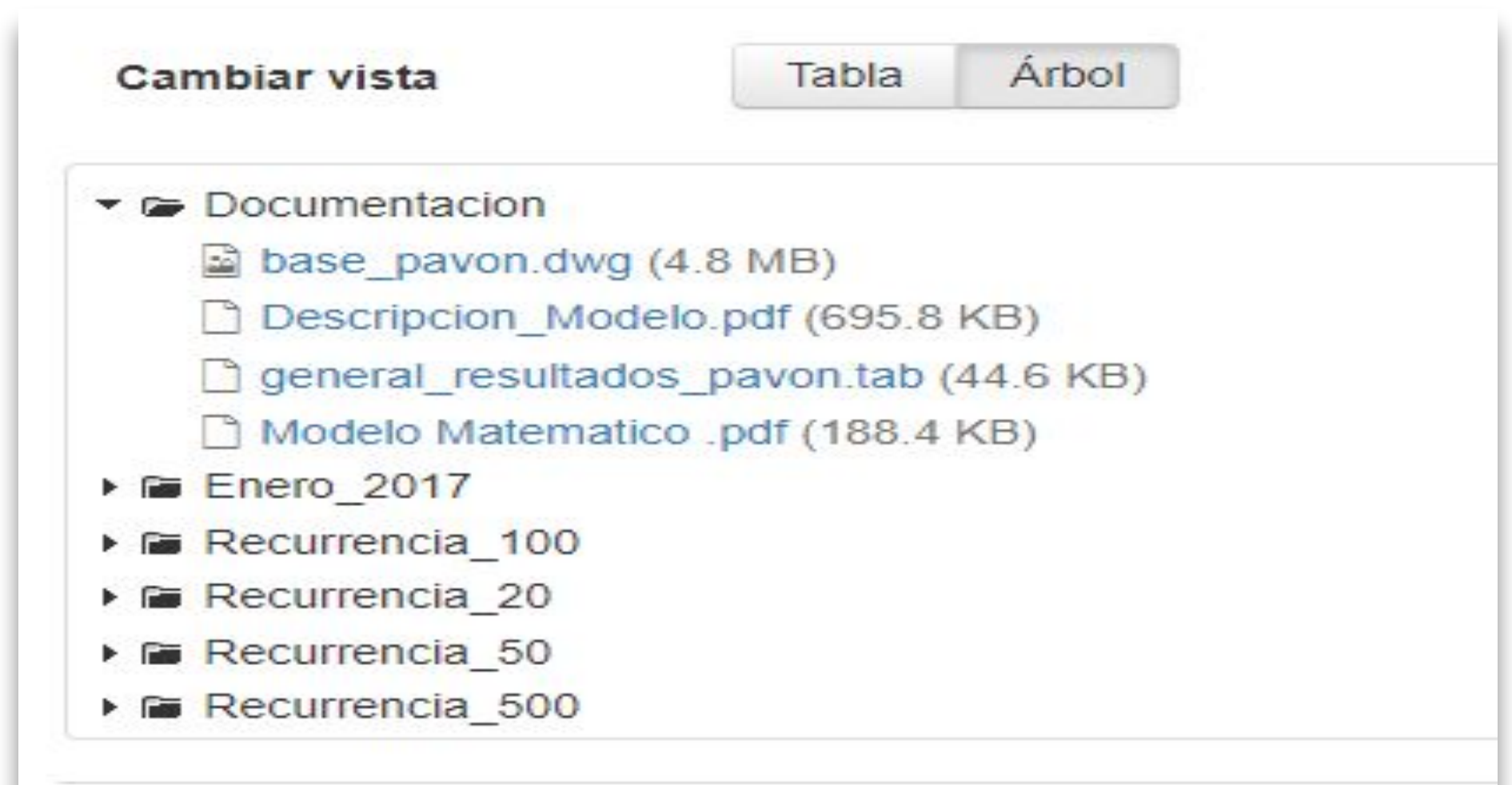
A través de la herramienta "formulario de Google" (Google Forms), se obtuvieron los datos sobre iniciativas de gobierno abierto de los municipios de las siguientes 28 ciudades, según orden temporal de respuesta:

La Plata (Buenos Aires), Rafaela (Santa Fe), Tandil (Buenos Aires), Luján de Cuyo (Mendoza), Oberá (Misiones), San Isidro (Buenos Aires), San Pedro (Santa Fe), Tigre (Buenos Aires), General Pueyrredón (Buenos Aires), Concordia (Entre Ríos), Olavarría (Buenos Aires), Santo Tomé (Santa Fe), Villa María (Córdoba), General San Martín (Buenos Aires), Rosario (Santa Fe), Guaymallén (Mendoza), Venado Tuerto (Santa Fe), Río Cuarto (Córdoba), Morón (Buenos Aires), San Carlos de Bariloche (Río Negro), Avellaneda (Buenos Aires), General Pico (La Pampa), Viedma (Río Negro), Maipú (Mendoza), Pilar

Miller, Ezequiel; Grandinetti, Rita María; Rodríguez, Ezequiel, 2022, "Respuestas de iniciativas de Gobierno Abierto en Municipios de Argentina", <https://doi.org/10.57715/UNR/UVRYFZ>, RDA UNR, V1

¿Cómo organizamos los datos?

- ✓ Si tenemos muchos datos o queremos mostrar la estructura de la investigación podemos organizarlos en carpetas.



Stenta, Hernán; Gerardo Riccardi; Pedro Basile; Carlos Scuderi, 2022, "Datos de: Modelación matemática hidrológica-hidráulica del escurrimiento superficial en la cuenca del A° Pavón (Santa Fe, Argentina)", <https://doi.org/10.57715/UNR/THCOKS>



Nos ponemos a trabajar en
<https://dataverse.unr.edu.ar/>

¿Qué revisamos los curadores?

- ✓ Que se cumplan las políticas y buenas prácticas.
- ✓ Que los metadatos estén completos y claros.
- ✓ Que se entienda el resumen.
- ✓ Que exista coherencia entre la descripción y los datos.
- ✓ Que se entienda la estructura de los datos (documentación).
- ✓ Que los datos estén en formatos abiertos.
- ✓ Que los datos tengan una licencia abierta.
- ✓ Que se explique cómo pueden ser reutilizados (Valor de los datos)



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**”



El Repositorio de Datos Académicos RDA-UNR CUMPLE 1 AÑO
dataverse.unr.edu.ar

 **15.438 VISITAS**

 **2.371 ACCESOS**

 **+ 30 PAÍSES**



RDA-UNR REPOSITORIO DE DATOS ACADÉMICOS

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

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

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



Ilustraciones:

The Turing Way Community, & Scriberia. (2023). Illustrations from The Turing Way: Shared under CC-BY 4.0 for reuse. Zenodo.

<https://doi.org/10.5281/zenodo.7587336>

Open Science Training Handbook CC0 1.0 Universal <https://doi.org/10.5281/zenodo.1212495>

Íconos de <https://www.freepik.com/>

UNR



**¡Muchas
gracias!**

repositoriodatos@unr.edu.ar