



Evento paralelo

Gestión de datos de investigación en repositorios institucionales. Experiencias y plan de trabajo del SEDICI repositorio institucional de la Universidad Nacional de La Plata Argentina

Dra. Marisa R. De Giusti
PREBI-SEDICI Universidad Nacional de La Plata
CESGI Comisión de Investigaciones Científicas

26 de abril de 2021

<http://sedici.unlp.edu.ar/>



UNIVERSIDAD
NACIONAL
DE LA PLATA



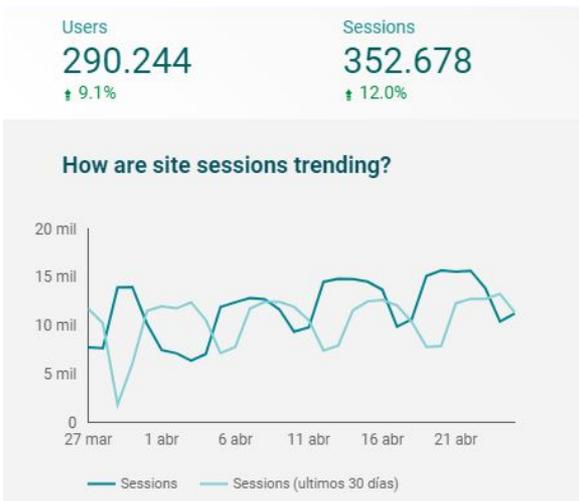
Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Logros

El repositorio SEDICI de la UNLP tiene casi 20 años desde su creación y lentamente se trata de adaptar a los nuevos actores y materiales.

El repositorio aloja **110.000** obras de todo tipo de las cuales **solamente** 300 son datos:

- [47 datasets](#) de fuentes diversas (Geografía (6), Humanidades (5), Botánica (2), Ciencias Naturales (1), Paleontología (1), Zoología (1))
- [255 Placas espectrográficas y placas fotográficas](#) de la Facultad de Ciencias Astronómicas.





Problemas frecuentes con los autores

- Extensas y difíciles reuniones con investigadores para
 - relevar los datos usados y generados en el marco de cada proyecto de investigación,
 - acordar formatos aptos para la visualización y eventual reúso.
- Celo sobre el dato para evitar que otro lo use y publique.
- Predisposición a cumplir con las obligaciones de depósito en repositorios externos y no en los institucionales.
- Falta de conocimiento del repositorio, sus funciones y capacidad para aumentar la visibilidad y el impacto.

Problemas frecuentes en la institución



- Políticas institucionales de acceso abierto incompletas o inexistentes, aún cuando es un requerimiento por ley.
- Reglamentación interna ausente o imprecisa en cuanto al mandato de depósito de datos
- Falta de comunicación hacia dentro de la propia institución.
- Falta de conciencia sobre la importación de soberanía y la preservación.
- Ausencia de Planes de gestión de datos.
- No se provee infraestructura tecnológica para guardar datos durante los proyectos, lo que resulta en que *cada proyecto guarda sus datos en cuentas de la nube y equipos personales.*

Problemas frecuentes en el repositorio



- Determinar cuáles son los formatos más convenientes para un área específica.
- Determinar si dado un formato de datos nuevo, es válido para subir al repositorio:
 - ◆ Los registros de formatos tradicionales, como PRONOM, tienen poca cobertura de formatos de datos.
 - ◆ Se requiere registros de formatos aptos para conjuntos de datos con información confiable y actualizada.
 - ◆ Al momento se usa Pronom, fileformats.archiveteam.org y Wikipedia.
- Espacio insuficiente para guardar los datos.
 - ◆ es un problema dependiente del tipo de datos. Los datos tabulares o de base de datos consumen en general muy poco mientras que los basados en imágenes y señales consumen mucho más.

Políticas y proyectos en curso

- Ley 26.899 (2013)
- Resolución 753 (2016)
- Sistema Nacional de Repositorios Digitales
- Comité de expertos en repositorios digitales
- Portal de datos abiertos DACYTAR
- Comité Asesor en Ciencia Abierta y Ciudadana



Ministerio de Ciencia,
Tecnología e Innovación Productiva
Presidencia de la Nación



Ley Nº 26.899
Creación de Repositorios Digitales
Institucionales de Acceso Abierto

[Infoleg.gob.ar](http://infoleg.gob.ar)



Resolución 753- E/2016 MINCYT
Reglamento operativo para la aplicación de la
ley Nº 26.899

boletinoficial.gob.ar



Redes de cooperación

- Los conjuntos de datos del repositorio son cosechados automáticamente desde portal de datos científicos nacional DACyTAR <https://dacytar.mincyt.gov.ar/>.
- Además, todas las obras son colectadas por el SNRD y luego por LaReferencia

The screenshot shows the DACyTAR website header with the Argentina.gov.ar logo and search bars. Below the header, the text reads: "El portal que te permite, de forma centralizada, buscar y acceder a todos los conjuntos de datos primarios de investigación disponibles en acceso abierto a través de los repositorios digitales institucionales que integran el Sistema Nacional de Repositorios Digitales (SNRD)".

Statistics displayed:

- 336 conjuntos de datos
- 7 repositorios cosechados

Below the statistics, there is a grid of category tiles with images and labels:

- Agricultura, silvicultura y pesca
- Artes (arte, historia del arte, artes escénicas, música)
- Biología agrícola
- Biología ambiental
- Biología industrial
- Biología médica



Cooperación: algunos ejemplos concretos

Recomendación y apoyo desde el **MINCYT** para que los datos generados en proyectos financiados en otros sistemas nacionales sean depositados en el repositorio institucional de cada entidad financiada:

- UNLP - Facultad de Ciencias Naturales y Museo (*en curso*)
 - 3 datasets con 60.000 observaciones de especímenes invertebrados
 - Formato de datos Darwin Core
 - Propuesta: depósito parcial de datos a medida que se desarrolla el proyecto y cosecha automática desde DACYTAR
- CEPAVE: bases de datos de mosquitos (*en curso*)
 - Normalización y georeferenciación de datos 6000 datos de ocurrencia entre 1900-2019 de mosquitos (culicidae)
 - Formato de datos Darwin Core
 - Propuesta: depósito de datos y cosecha automática desde DACYTAR

Siguientes pasos: servicios complementarios

- Visualizadores de datos para los formatos principales.
- Herramientas básicas para uso y explotación de datos en diversos formatos (ejemplo consultas en una grilla, navegación de mapas, extracción de subconjuntos de datos, entre otros).
- Generación automática de doi para datasets depositados en el repositorio.
- Interconectar obras a partir de metadatos de relación, en particular para versiones de un mismo dataset.
- Promoción del depósito como una vía para aumentar la visibilidad e impacto de la producción y de sus autores.

Recomendaciones: Requisitos mínimos y necesidades urgentes

- Apoyo institucional expresado a partir de la definición de políticas escritas y claras, actualización de contratos y procedimientos de becas y proyectos.
- Mejora de las prácticas de gestión de datos en proyectos a partir del uso de PGD y de la asistencia institucional para resguardo de datos primarios y derivados en proyectos.
- Mejora de la comunicación con investigadores y becarios para convencer el beneficio del depósito de datos, de FAIR, etcétera.
- Mejora de la infraestructura del repositorio para afrontar los datos que se vienen.
- Se proyecta una enorme diversidad de formatos, con lo cual la preservación y las migraciones de formatos tomarán aún mayor relevancia.
- **Tenemos ante todo que contar con el dato. Esa es la prioridad! Ese es el desafío!**



¡Muchas gracias!

Marisa R. De Giusti

marisa.degiusti@sedici.unlp.edu.ar

Nuestra colección en el RI: <http://sedici.unlp.edu.ar/handle/10915/25293>



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)