

**Plan de Gestión de Datos UNR (PGD-UNR)
correspondiente a la convocatoria de
acreditación de Proyectos-UNR 2023**

Dirección General de Estadística Universitaria

Año 2024

Plan de Gestión de datos

El Plan de Gestión de Datos UNR (PGD-UNR) es un documento formal, elaborado por investigadores, que se presenta al mismo tiempo que 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. El objetivo del PGD-UNR es 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 este formulario podrá generar el PGD requerido por la Universidad Nacional de Rosario conforme a la Ley 26.899.

Para la elaboración de este informe se trabajó en conjunto con la Unidad de Gestión de Acceso Abierto con el fin de analizar PGD-UNR correspondiente a la convocatoria de acreditación de proyectos UNR durante el año 2023. En una primera instancia se eliminaron registros duplicados, registros de prueba o que no tenían respuestas. Además, se encontraron casos en que la misma persona respondía más de una vez con número diferente de identificador de proyecto pero el resumen y las respuestas eran exactamente iguales. A continuación, se analizan un total de 207 proyectos.

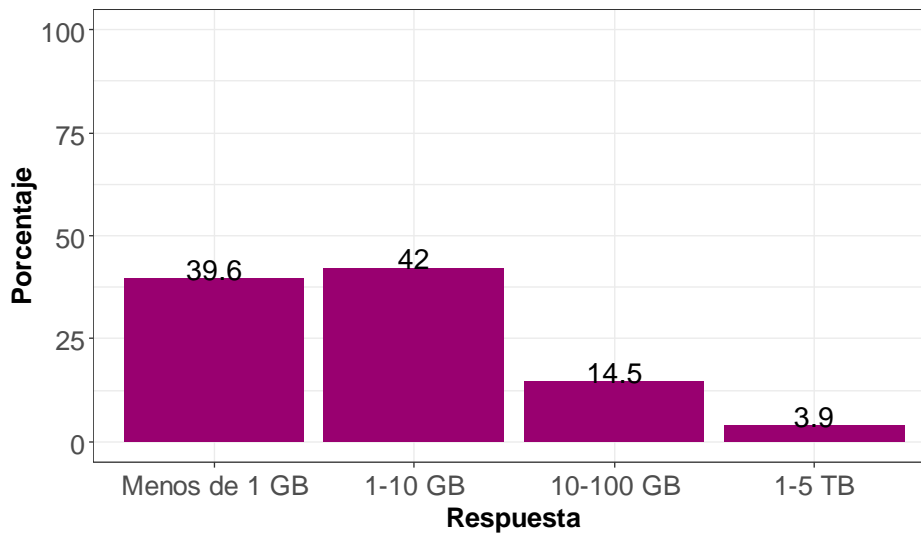
Archivos	Frecuencia
xls/xlsx	116
doc	76
Bases de datos	
XML	24
CSV	34
JSON	7
Datos Geoespaciales	
SHP	4
DBF	4
GeoTIFF	4
netCDF	3
GeoJSON	2
Rinex	3
CAD	
Dxf	6
Sat	3
lgs	2
stp	2
Nube de Puntos PointCloud	
LAZ	1
XYZ	1
PTX	1
Videos	
MPEG	22
AVI	17
MXF	7
MKV	9
Mp4	10
Audio	
WAVE	10
AIFF	7
MP3	24
MXF	7
Otro	
Estadísticas	
ASCII	16
DTA	6
POR	14
SAS	6
SAV	7
Imágenes	

TIFF	39
JPEG	77
JPG	52
PDF	100
PNG	31
GIF	16
BMP	12
SVG	11
Datos tabulares	
CSV	34
TXT	41
Texto	
XML	24
PDF/A	100
HTML	21
ASCII	16
UTF-8	8
Sismología	
SEED	1
Código: (tcl files, py files)	
Jupyter	1
Notebook	1
Archivo web	
WARC	5
Contenedores	
TAR	8
GZIP	4
ZIP	20

3.3 Tamaño esperado de los datos

En el gráfico 2 se ve que hay un similar porcentaje entre archivos de tamaño menor a 1 GB y entre 1 GB y 10 GB, los cuales son los predominantes. Luego un 14,5% utilizan datos entre 10 GB y 100 GB y sólo el 3,9% ocupan entre 1 a 5 TB.

Gráfico 2: Distribución porcentual del tamaño esperado de los datos



4.1 Indicar el origen de datos preexistentes

En esta pregunta de respuesta abierta se contabilizaron por cada registro palabras “claves” para poder categorizar. Se observa que la palabra que más aparece son los datos generados por cada proyecto. Cabe aclarar que en un mismo registro puede contener más de una de estas palabras claves o que no tenga ninguna.

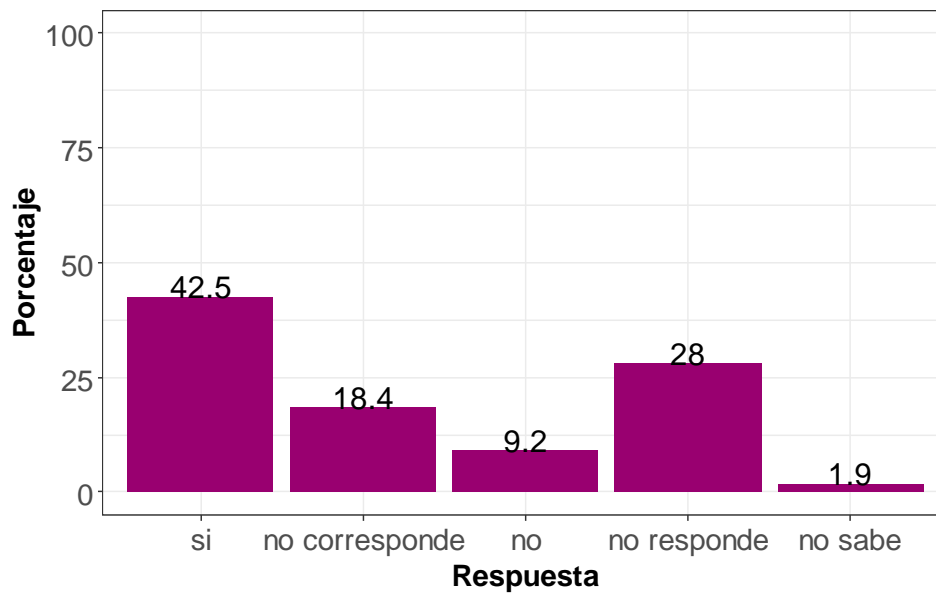
Palabras claves	Frecuencia
Propios	53
Terceros	12
Publicados	45
Bibliografía	17
Repositorio	26
No corresponde/no aplica/ no se utilizarán datos	17

4.2 Indicar si estos datos están en acceso abierto

En esta pregunta de respuesta abierta se contabilizaron por cada registro palabras “claves” para poder categorizar. Se observa que la mayor frecuencia ocurre en proyectos que tienen accesos abiertos.

Respuestas	Frecuencia
Sí	88
No	19
No corresponde/no aplica	38
No responde	58
No sabe	4
Total	207

Gráfico 3: Distribución de si los datos están en acceso abierto



5 Utilidad de datos

Al ser una pregunta de respuesta abierta y con gran amplitud, no fue posible resumirla.

6.1.1 Almacenamiento de los datos

En esta pregunta de respuesta abierta se contabilizaron por cada registro palabras “claves” para poder categorizar. Se visualiza que la fuente de almacenamiento más elegida es la nube. Cabe aclarar que en un mismo registro puede contener más de una de estas palabras claves o que no tenga ninguna.

Palabras claves	Frecuencia
Memoria extraíble	67
Discos internos	64
Discos externos	30
Almacenamiento en la nube	161
Servidores	36

6.1.2 Copias de seguridad

En esta pregunta de respuesta abierta se contabilizaron por cada registro palabras “claves” para poder categorizar. Se visualiza que casi en la totalidad de los proyectos realizan copias de seguridad.

Respuestas	Frecuencia
Sí	197
No	2
No corresponde/no aplica	4
No Responde	4
Total	207

6.1.3 Compartir datos con el equipo de investigación

En esta pregunta de respuesta abierta se contabilizaron por cada registro palabras “claves” para poder categorizar. Se visualiza que casi en la totalidad comparten datos con el equipo de investigación y aproximadamente la mitad en Google Drive. Cabe aclarar que en un mismo registro puede contener más de una de estas palabras claves o que no tenga ninguna.

Palabras claves	Frecuencia
Nube*	34
Google Drive	108
Dropbox	5
Servidor compartido	5
Spotify	1
Carpetas **	24
Repositorio	3
No corresponde/no aplica	6
No responde	6

Ref.: * Se contabilizó el número de veces que figure nube pero que no contenga google drive.

** Se contabilizó la cantidad de veces que aparece carpetas pero que no contenga google drive ni nube.

Consideraciones

Tras un análisis detallado se tomó la decisión de revisar y actualizar el formulario del Plan de Gestión de Datos de la UNR en colaboración con la Unidad de Gestión de Acceso Abierto. El objetivo principal de esta modificación fue optimizar la utilización del formulario para mejorar un posterior análisis de los datos recopilados.

Para lograrlo se implementó un enfoque centrado en la categorización de las respuestas, transformando las preguntas abiertas en preguntas cerradas. Esta estrategia busca minimizar las interpretaciones subjetivas por parte de los investigadores asegurando una mayor uniformidad y coherencia en los datos recogidos.