

Guía para el Depósito de Datos de Investigación Asociados a Publicaciones

1. Objetivo

Esta guía tiene como objetivo orientar el depósito de conjuntos de datos primarios y crudos recopilados o generados durante el desarrollo de artículos de investigación, tesis, informes y otros productos de investigación, con el propósito de fomentar el acceso abierto y proteger la información generada, garantizando su disponibilidad, rastreabilidad y conexión con las publicaciones del Repositorio Institucional de la Universidad La Salle (en adelante ULASALLE).

2. Alcance

La presente guía es de cumplimiento obligatorio para:

- Investigadores, docentes y estudiantes de ULASALLE.
- Equipos de investigación de ULASALLE.
- Proyectos internos con financiamiento de ULASALLE.
- Proyectos de investigación financiados con fondos concursables del Estado.

3.- Base Normativa

- **Ley N.º 31250**, Ley del Sistema Nacional de Ciencia, Tecnología e Innovación (SINACTI)
- **Ley N.º 30035**, Ley que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto (ALICIA).
- **Ley N.º 29733**, Ley de Protección de Datos Personales y su Reglamento.
- Principios Internacionales **FAIR** (*Findable, Accessible, Interoperable, Reusable*).

4. Creación de la Colección Institucional de Datos Primarios y Crudos

La Dirección de Investigación habilita y mantiene dentro del Repositorio Institucional la “Colección Especializada de Datos Primarios y Crudos de Investigación”, la cual estará interoperable con la Red ALICIA y portales internacionales. Para los datos primarios/crudos se debe definir una estructura propia en el repositorio, tal como sigue:

- **Comunidad:** *Datos Abiertos de Investigación.*
- **Colección Especializada:** *Colección Especializada de Datos Primarios y Crudos de Investigación.*
- **Ítem (Registro):** El conjunto de datos individual derivado de una investigación específica.

5.- Formatos Abiertos e Interoperables

Los datos primarios o crudos deben depositarse de acuerdo con los siguientes formatos abiertos recomendados para el cumplimiento de los principios FAIR:

- Datos y Documentos
 - CSV para hojas de cálculo.
 - OpenDocument (ODT, ODS, ODP) .
 - Markdown o HTML para textos estructurados.
 - PDF para preservación de documentos.
- Imágenes y Gráficos
 - PNG.
 - SVG para gráficos vectoriales.
 - OpenEXR para imágenes HDR.
- Video y Audio
 - MKV.
 - WebM para videos en la web.
 - FLAC u Opus.
- Bases de Datos y Datos Estructurados
 - JSON, YAML o XML.
 - RDF o CSV para datos estructurados.
 - geoJSON o KML para datos geoespaciales.
- Compresión y Archivos
 - TAR, ZIP o 7z.
 - GZIP o XZ para compresión sin pérdida.

6.- Archivo Descriptivo Obligatorio

Todo dataset primario o crudo deberá adjuntar un archivo de texto plano denominado README.txt que sirva como diccionario de datos con la finalidad de presentar:

- La estructura del dataset y descripción de las variables/columnas.
- El software, versión y scripts necesarios para ejecutar/procesar los datos.
- Metodología de recolección de los datos y fechas del experimento o campo.

7.- Configuración del esquema de metadatos

En el repositorio, se debe ingresar o mapear la información técnica del dataset. El esquema estándar de metadatos para datos primarios/crudos utiliza los siguientes campos de Dublin Core (DC) o el esquema ampliado DataCite:

- dc.title: Título descriptivo del dataset.
- dc.type: Valor obligatorio
- dc.type.version: Versión del dato.
- dc.relation.isreferencedby: Enlace Handle o DOI del artículo científico o tesis al cual pertenecen estos datos.

- dc.rights: Tipo de Licencia Open Data .
- dc.coverage.spatial: Coordenadas geográficas o ámbito territorial de recolección de los datos.
- dc.coverage.temporal: Rango de fechas en las que se recolectaron los datos crudos.
- dc.identifier.doi: Identificador Persistente (DOI) asignado directamente al dataset (vía DataCite u otro proveedor).

8. Flujo de carga

El proceso de carga en el repositorio sigue este flujo:

- a) **Autoarchivo / Depósito:** El responsable de la investigación ingresa a la colección de datasets en el repositorio, completa la ficha de metadatos, adjunta el archivo README.txt y sube las tablas o archivos de datos primarios o crudos.

El administrador del repositorio revisa que los archivos estén en formatos abiertos, que no contengan datos sensibles, secretos comerciales, patentes pendientes o información personal identificable, y que los metadatos y el diccionario README.txt estén completos.

- b) **Gestión de Embargo:** Si el proyecto aún requiere proteger la propiedad intelectual antes de una publicación formal, se activa la opción de Embargo, haciendo visibles los metadatos pero restringiendo la descarga del archivo de datos por un periodo de 6 a 12 meses.
- c) **Publicación y Asignación de Handle:** Al aprobar el ítem, se genera un código Handle permanente y, si está integrado con DataCite, le asigna un DOI específico al dataset.
- d) Una vez almacenado correctamente en el repositorio bajo la colección de datasets y con el metadato dc.type = dataset, el repositorio expone esta información vía protocolo OAI-PMH para que sea cosechada automáticamente por el portal ALICIA (CONCYTEC) y por motores de búsqueda globales de datos de investigación como Google Dataset Search o DataCite Commons.

9. Consideraciones sobre confidencialidad y licenciamiento

- No deben incluirse datos personales, médicos o de terceros sin consentimiento informado.
- Los datasets se publican preferentemente bajo licencias Creative Commons Atribución (CC BY 4.0), salvo casos justificados de acceso restringido.
- Los autores conservan la propiedad intelectual de sus datos.

10.- Justificación de Embargo de Datos Primarios/Crudos

De forma excepcional y fundamentada, el responsable de la investigación podrá solicitar la reserva temporal (embargo) de los archivos de datos primarios/crudos, manteniendo visibles únicamente los metadatos generales, bajo las siguientes causales:

- a) **Propiedad Industrial (Patentes):** Cuando los datos primarios sustenten una invención o modelo de utilidad en proceso de registro ante el INDECOPI. Plazo máximo: 24 meses.
- b) **Secretos Comerciales o Contratos de I+D+i:** En proyectos cofinanciados con empresas privadas con cláusula contractual de confidencialidad. Plazo máximo: 36 meses.
- c) **Exclusividad Editorial:** Exigencia explícita de la revista científica (*Journal Data Policy*) previa a la publicación. Plazo máximo: 12 meses.