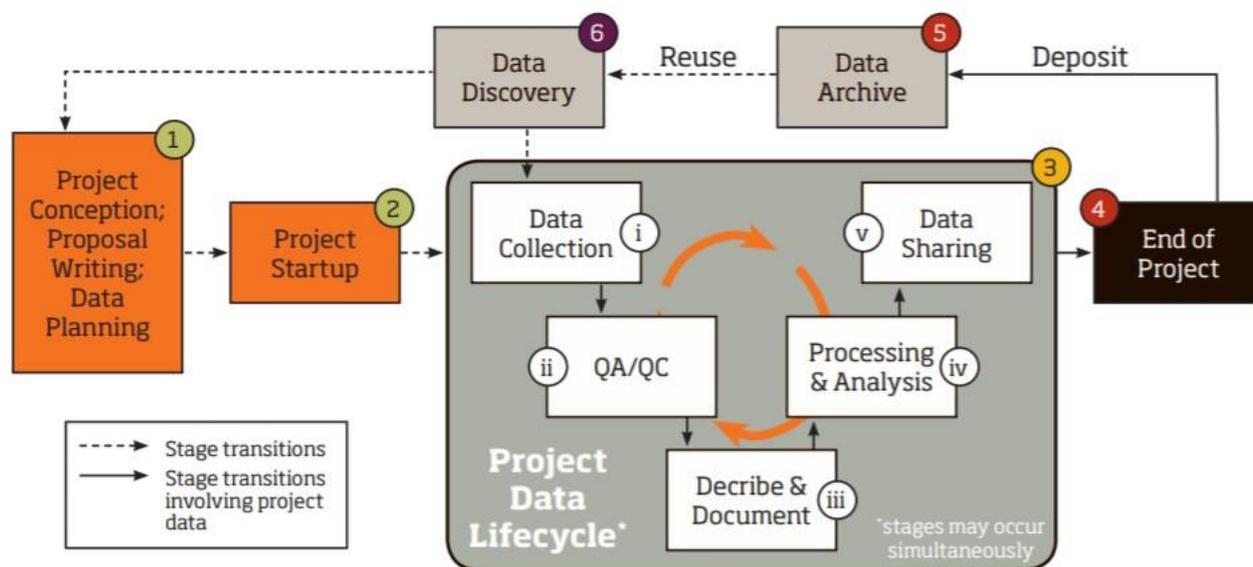


FUNDAMENTOS

01. Ciclo de vida de los datos en proyectos de investigación

Los modelos de *ciclo de vida de datos* en proyectos de investigación ilustran las etapas del manejo de datos y describen como fluyen durante el proyecto de investigación de inicio a fin.

En cada etapa hay que definir la propiedad de los datos, la documentación de los datos, el almacenamiento, la seguridad y el control de calidad de los datos. Todas pueden ocurrir simultáneamente (Figura 1).



Each of the boxes above represents a stage in the research lifecycle for a given project. Transitions between stages involving project data (solid black arrows) are vulnerable to data loss. These phases are often coincident with shifts in responsibility for the data (e.g. from data collector to data analyst, or from graduate student to PI), which makes transitions a critical time for active data management. Data loss may occur in many other ways, for various reasons: data are unusable due to lack of necessary contextual details (lack of sufficient documentation/metadata); failure to move data to the next stage, reaching a dead-end prior to systematic preservation; bit rot or other digital degradation; or data are physically misplaced. The lifecycle model is used to provide context in describing data stewardship activities that should take place over the course of a research project.

Figura 1: Manejo de datos durante el ciclo de vida de los proyectos de investigación
(fuente: <https://www.usgs.gov/datamanagement/>)

Aquí tienes una lista de las etapas o componentes del ciclo de vida de los datos y de la información básica que tendrás que identificar en cada una de ellas (<https://www.usgs.gov/datamanagement/>):

➤ *Elaboración de un Plan de Gestión de Datos (PGD) en proyectos de investigación:*

Un PGD es un documento formal que describe qué vas a hacer con los datos durante el proyecto y después de que éste finalice. La mayoría de los investigadores obtienen sus datos en base a un plan, un cronograma y unos objetivos. El mayor beneficio de tener un plan de gestión de datos es el reconocimiento, su reproducibilidad.

Proporciona la información básica de tu proyecto: título, cualquier número de seguimiento de proyectos, punto de contacto, etc., a continuación, identifica el marco temporal del proyecto y las actividades de colección de datos (ejemplo: fecha de inicio y fin del proyecto, los periodos de recogida de datos). Debes proporcionar también información de contacto para la plantilla de trabajo y los colaboradores involucrados en el proyecto e identificar a la persona responsable en general del manejo de los datos, la adquisición, el procesamiento, control de calidad, documentación y almacenamiento.

Si puedes, proporciona tu presupuesto estimado para las actividades de manejo de datos. Y si procede, identifica un acuerdo para compartir datos que defina los roles y las responsabilidades para la recogida de datos y para compartirlos (estas últimas etapas se definen conforme evoluciona el proyecto).

➤ **Adquisición de los datos:**

Durante esta etapa se debe proporcionar información básica para identificar cada base de datos (título, descripción, fuente, punto de contacto). Describe el **objetivo de cada base de datos** en el contexto del proyecto de investigación. Identifica cualquier restricción inherente al uso de las bases de datos. Identifica el formato de cada base de datos y los requerimientos de almacenamiento de cada base de datos.

➤ **Durante el procesamiento y análisis de los datos:**

Registra las transformaciones de los datos, acciones y cualquier paso en el procesamiento para producir las bases de datos final (destinada al análisis). Describe también las tecnologías y modelos que se usarán durante el procesamiento de los datos.

En el caso de los modelos, programas y código, haz una lista de entradas y salidas de datos y detalles de calibración.

➤ **Almacenamiento de datos:**

Para guardar los datos, documenta quién será el responsable de asegurar la conservación de los datos. Indica también los formatos de los datos y el volumen indicado de las bases de datos finales. También es importante que definas los **repositorios** donde se van a almacenar tus datos y las bases de datos que proporcionarán tus bases de datos finales, los programas, modelos o código.

➤ **Publicación y distribución de datos:**

Proporciona información sobre la identificación preliminar de las **publicaciones que puedan surgir de esa base de datos** (título, descripción, lista de autores, etc.). Describe el formato anticipado de cada publicación (revistas, tipos de datos, modelos, etc.). Indica como planeas mantener y actualizar los registros de

metadatos en formatos estándar FGDC o ISO XML. Describe cómo (herramienta, persona encargada, etc.) se asignará un **número de objeto identificador digital (DOI)** a cada base de datos aprobada para distribuir. Indica también cualquier restricción inherente que pueda ser impuesta en el producto derivado en el uso de salidas de datos con propietario u otros factores.

➤ ***Descripción de los metadatos, manejo de la calidad, recuperación y seguridad de datos:***

Describe las herramientas o procesos que se usarán para crear los **metadatos**, e identifica a la persona responsable de crear los archivos de metadatos. Documenta las responsabilidades y roles del equipo del proyecto y todos los procedimientos operacionales. Identifica la localización de los recursos internos de almacenamiento para proporcionar una réplica y la capacidad de recuperación de datos, y que será empleada para almacenar los datos obtenidos durante el proceso de análisis. Indica el **contacto de la persona** que se encargará de los dispositivos de almacenamiento.

Fuentes:

<https://www.usgs.gov/products/data-and-tools/data-management/data-lifecycle>

<https://www.usgs.gov/products/data-and-tools/data-management/overview-data-management>

<https://guides.library.oregonstate.edu/research-data-services/lifecycle>