PLAN DE GESTIÓN DE DATOS UNA

(PGD-UNA)

El Plan de Gestión de Datos e s un instrumento que facilita la planificación de los aspectos relacionados al manejo de los datos durante y después de las investigaciones. Incluye las dimensiones de recolección, utilización, generación, documentación, resguardo, almacenamiento y accesibilidad de los datos. Cada punto debe ser desarrollado con precisión y en la medida de los posible. Las preguntas que se presentan pretenden ser una guía para la reflexión, por lo que no son las únicas posibles y no representan una obligación de respuesta.

EL PGD se debería de diseñar junto a la formulación inicial, pero es un documento flexible y podría modificarse según se ejecute la investigación y emerjan cambios e imprevistos

1. Descripción general de datos del proceso investigativo

Hacer una breve descripción de los datos que generará, reutilizará y creará, incluyendo los reutilizados (propios o de terceras fuentes), anotando para cada caso su contenido, tipo y alcance. Justifique el formato seleccionado y considere las implicaciones de este y del volumen de los datos, en función del almacenamiento que requerirán, el respaldo y el acceso.

* 1. **Preguntas guías sobre la descripción de los datos:** 
     + ¿Qué tipo de datos utilizará y generará? En la UNA se utiliza el vocabulario controlado recomendado por la Confederación de Repositorios de Acceso Abierto (COAR), de los cuales los más utilizados son: datos agregados; datos compilados; datos codificados; datos experimentales; datos genómicos; datos geoespaciales; cuaderno/cuaderno electrónico de laboratorio; datos de medición; prueba datos observacionales; datos registrados (datos grabados); datos de simulación; datos de encuesta; ensayo clínico; datos crudos; datos reutilizados; colecciones físicas (herbario, colecciones de nemátodos); imágenes. Para ampliar detalles, visitar el sitio del Repositorio Institucional <https://docs.google.com/document/d/1ySIOtQ_uNEKJnNOOoxYdlQof5EPSuUQ2liso8oBD1zg/edit>
     + ¿Qué formato y cantidad de datos utilizará y generará? Se busca que se describan los formatos o software a utilizar y por qué. Considere que los formatos deben permitirle a usted y a terceras personas, compartir y acceder a largo plazo a os datos o, por lo que es importante utilizar formatos abiertos, estándares aceptados y utilizados internacionalmente en el área de conocimiento correspondiente. Utilizar formatos estandarizados, intercambiables y abiertos asegura el acceso a largo plazo de los datos. Algunos ejemplos de formatos muy utilizados en la práctica de compartir datos son el texto plano text (.txt), comma-separated values (.csv), geo-referenced TIFF (.tif, .tfw). Para más ejemplos de formatos puede ir a siguiente link <https://repositorioinstitucional.ceu.es/help/formats.jsp>
     + ¿Va a contar con datos reutilizados o reutilizables? ¿propios o de otras fuentes? Por ejemplo, se pueden utilizar datos producto de otras investigaciones propias, o bien datos de bases públicas como el INEC o el Informe del Estado de la Nación o de bases de datos de otros autores, de libre acceso de quienes se tiene el permiso.

1. Recolección y creación de los datos

Describa la estrategia a usar en la generación o colecta de los datos y el o los estándares (nacionales o internacionales) que usará. Considere la forma en que organizará los datos durante la investigación, mencionando por ejemplo la convención de nomenclatura, la organización de las carpetas donde almacenará los datos y cómo administrará el control de las versiones de los archivos. Esto puede implicar procesos como calibración, repetición de muestras o de mediciones, estandarización del registro de los datos, validación de la digitalización de los datos, revisión de los datos por pares o la representación de estos con vocabularios controlados.

Después de recolectada la información detalle la forma en que será procesada. Incluya información sobre software a utilizar, algoritmos, flujos de trabajo científico, entre otros.

* 1. **Peguntas guía sobre la colecta y/o creación de los datos**
     + ¿Qué estándares o metodologías usará para ambos?
     + ¿Cómo estructurará y denominará las carpetas y archivos?
     + ¿Cómo gestionará las versiones?
     + ¿Qué procesos usará para asegurar la calidad de los datos? Ejemplo: Los registros de las entrevistas personales serán almacenados en un sistema de archivo de ficheros independientes ordenados en topología de árbol según la fecha y ubicación de recolecta. Se utilizará replicación de datos en dos servidores independientes y se llevará a cabo una actualización semanal de ambas bases de datos para asegurar la integridad. Se aplicarán los estándares internacionales de la normativa ISO 27000 referente a seguridad de la información (aplicable a empresas públicas o privadas). Esta especifica los requisitos necesarios para establecer, implantar, mantener y mejorar un sistema de gestión de la seguridad de la información (SGSI) según el “Ciclo de Deming”: PDCA - acrónimo de Plan, Do, Check, Act (Planificar, Hacer, Verificar, Actuar).

1. Documentación y metadatos

Describa el tipo de documentación que se asociará a los datos para ayudar a los usuarios secundarios a entenderlos y reutilizarlos (e.g. minutas, cuadernos de laboratorio, procedimientos, normativa, entre muchas otros). Al menos debe incluir los detalles básicos que le permitirán a las personas encontrar los datos; la identificación de las personas que los crearon o contribuyeron a hacerlo; el título, la fecha de creación y las condiciones para su acceso.   
La documentación podría incluir detalles de la metodología usada, información sobre análisis y procedimientos, la definición de variables, el vocabulario, las unidades de medida, los supuestos Los metadatos de archivos como word, pdf, excel, se pueden generar en la misma aplicación. Este tipo de metadatos puede aplicar de acuerdo al área del conocimiento. Por ejemplo, un metadato automáticamente generado es el nombre de la persona que genera el archivo. Para más información de metadatos <http://rd-alliance.github.io/metadata-directory/standards/>

* 1. **Preguntas guía para la documentación y creación de metadatos:** 
     + ¿Qué información es necesaria para que los datos puedan ser leídos e interpretados en el futuro?

* + - Describa la forma en que reportará los metadatos y cuál estándar utilizará y por qué Puede solicitar apoyo de la persona bibliotecóloga de su Facultad, Centro o Sede.

1. Cumplimientos éticos y legales

Los aspectos éticos afectan la forma como se almacenan los datos, su acceso, uso y el tiempo que se conservarán, por lo que es necesario reportarlo. El manejo de estos aspectos pude incluir: anonimato de los datos, comités éticos institucionales, acuerdos de consentimiento informado, soberanía de datos indígenas, entre otros. Si su investigación involucra personas, además debe asegurarse de obtener su consentimiento para el compartir y reutilizar los datos. Por ejemplo: *El tratamiento de los datos será totalmente anónimo y no será incluida ninguna información de carácter ideológico, orientación sexual, racial o religioso.*

Respecto a los derechos de autor(a), la universidad recomienda utilizar en su mayoría la licencia Creative Commons Reconocimiento-Nocomercial-CompartirIgual (CC BY NC SA) pero hay excepciones que pueden utilizar otro tipo de licenciamiento o bien tienen Copyrigth. Tome en cuenta que en proyectos elaborados por personas de diferentes instituciones (por ejemplo, proyectos FS, colaboradores internacionales), deben tomarse acuerdos mutuos sobre autorías y derechos. Se deben considerar todos los aspectos legales relacionados con el tema que rigen en cada institución participante. Aquí también debe informar sobre los permisos de reutilización de los datos de otros autores y cualquier otra restricción relacionada con el compartir los datos. En este apartado debe quedar expreso a quién le pertenecen los datos y si restricciones para la reutilización de datos generados por terceros o relacionados a los consentimientos informados. Considere si la socialización de los datos se pospondrá o restringirá para su publicación o patentamiento.

Para más información en casos particulares puede contactar a las asesoras de la VI y de la OTVE.

* 1. **Preguntas guía sobre el tema ético y legal** 
     + ¿Ha considera los aspectos éticos de la iniciativa y del acceso a los datos generados en la misma?
     + ¿Cómo ha previsto identificar y tratar los aspectos legales?

1. Almacenamiento, respaldo, seguridad y acceso personal

Describa dónde se almacenarán y respaldarán los datos durante el curso de la ejecución de la investigación. Esto puede variar si está haciendo trabajo de campo o trabajando en varios sitios, así que explique cada procedimiento. Identifique quién será responsable de la copia de seguridad y con qué frecuencia se realizará, la ubicación de estas, así como la cantidad de copias que manejará. Almacenar los datos de forma exclusiva en portátiles, espacios compartidos o unidades de respaldo externo es una práctica de alto riesgo, por lo que es recomendable el uso de opciones de almacenamiento robustas, y administradas con respaldo automático como las provistas por la UNA (Repositorios, servidores, otros). De igual manera, es preferible contar con sistemas automáticos de respaldo más allá de los sistemas manuales (discos duros extraíbles, dispositivos de almacenamiento USB, otros similares). Si escoge un proveedor de respaldo, debe asegurarse que no existan conflictos con las políticas de financiantes e institucionales, por ejemplo en el caso de datos sensibles.

* 1. **Preguntas guía sobre el tema de almacenamiento y respaldo**
     + ¿Tiene suficiente capacidad de almacenamiento?
     + ¿Cómo se respaldarán los datos?
     + ¿Quién será responsable de hacer los respaldos y la recuperación de los datos?
     + En caso de considerar las opciones de respaldo institucionales ¿qué hará en caso de que accidentalmente se pierdan?

1. Acceso y seguridad

Una consideración para tomar en cuenta son los potenciales riesgos ante la desaparición involuntaria de los datos o bien el robo de estos. Si los datos son confidenciales, debe delimitar las medidas de seguridad que implementará, así como los estándares internacionales que utilizará, según sea el caso, por ejemplo: ISO 27001.

* 1. **Preguntas guía sobre el tema de acceso y seguridad**
     + ¿Cuáles son los riesgos relacionados con la seguridad de los datos y cómo se manejarán esos riesgos?
     + ¿Cómo controlará el acceso a los datos para mantener su seguridad?
     + ¿Cómo conseguirá que las personas colaboradoras tengan acceso a los datos de forma segura?
     + Si se generan o colectan datos en campo ¿cómo garantizará su transferencia segura a su sistema principal de seguridad?

1. Selección de datos para su reutilización y preservación

Recuerde que no es lo mismo almacenar que resguardar, pues lo segundo implica una selección en base a criterios por usted definidos, del valor a largo plazo de los datos, que le permitan determinar cuáles pueden reutilizarse por usted o por terceros, los que deben de conservarse y aquellos que deben pasar por periodos de embargo.

Considere la forma en que los datos deben de reutilizarse. Por ejemplo, para validar los hallazgos de la investigación o para fines docentes. Decida cuáles datos mantener y por cuánto tiempo. Esta decisión puede basarse en cualquier obligación de retener datos, en la estimación de su valor de reutilización, en la viabilidad económica de mantenerlos y en cualquier otro esfuerzo requerido para preparar los datos para compartirlos y preservarlos, como por ejemplo un eventual cambio de formato.

Ver la guía del Data Curation Centre: [How to appraise and select research data for curation](http://www.dcc.ac.uk/resources/how-guides/appraise-select-data) o Mejores prácticas DataONE: [Identifying data with long-term value](https://www.dataone.org/best-practices/identify-data-long-term-value). Para casos específicos consultar con una asesora académica de la VI o bien con la OTVE.

* 1. **Preguntas guía para el tema de selección y preservación** 
     + ¿Cuáles datos tienen valor a largo plazo? ¿Cuáles deberían de retenerse, compartirse o conservarse? ¿Qué criterios usará para decidir esto?

* + - ¿Cuáles datos deben ser conservados o destruidos, de acuerdo con regulaciones contractuales y legales de la universidad?
    - ¿Cuánto tiempo los datos serán conservados y preservados?

* + - ¿Cuál es el plan para el almacenamiento de los datos a largo plazo? Describa el tipo de almacenamiento y curaduría por parte de las personas investigadoras de las bases de datos con valor de largo plazo.
    - ¿Dónde o en cuál repositorio? En el caso de la UNA en el Repositorio Institucional (https://repositorio.una.ac.cr/), inclusive para el caso de la mayoría de las iniciativas de cooperación internacional.

1. Compartición de los datos de investigación

Resuelva dónde, cómo y para quién deben estar disponibles los datos en el largo plazo. Los métodos usados para compartir los datos dependerán de factores como el tipo, tamaño, complejidad y la discrecionalidad en el manejo de los datos. Considere las formas en que terceras personas reconocerán los datos para su reutilización.

* 1. **Preguntas guía sobre el tema de cómo compartir** 
     + ¿Cómo se enterarán los potenciales usuarios de sus conjuntos de datos?

* + - ¿Con quién compartirá los datos y bajo qué condiciones?
    - ¿Compartirá los datos a través del repositorio, atendiendo solicitudes directas u otro mecanismo?
    - ¿Cuándo podrá a disposición los datos?

1. Restricciones al compartir los datos

Determine si los datos tienen restricciones debido a aspectos de confidencialidad, consentimiento o sensibilidad de los datos. Considere si un acuerdo de confidencialidad brindaría suficiente protección para los datos. Recuerde que el compartir datos debe garantizar el cumplimiento de los principios [FAIR](https://biblioguias.cepal.org/c.php?g=495473&p=8022713#:~:text=Estos%20cuatro%20principios%20son%20llamados,actores%20se%20pueden%20ver%20beneficiados).

* 1. **Preguntas guía sobre el tema de cómo compartir**
     + ¿Qué acciones implementará para evitar o minimizar las restricciones?

* + - ¿Por cuánto tiempo necesitará un uso exclusivo de los datos y por qué?
    - ¿Se necesitarán acuerdos de algún tipo para compartir los datos?

1. Responsabilidades y recursos

Resuelva las responsabilidades de las personas involucradas sobre el manejo de los datos y del plan de gestión de datos. Considere cuidadosamente cualquier recurso necesario para ejecutar el plan, por ejemplo software, hardware, conocimientos técnicos, etc. Cuando se necesiten recursos específicos, estos deben describirse y justificarse.

* 1. **Preguntas guía sobre el tema de las responsabilidades y los recursos:**
     + ¿Quién(es) será(n) responsable(s) del manejo de los datos y cuáles serán sus responsabilidades en este manejo?
     + ¿Quién es la persona responsable de la implementación del plan de gestión de datos, y de garantizar su escrutinio y revisión?
     + ¿Cómo estarán distribuidas las responsabilidades entre las diversas entidades participantes? (En el caso de proyectos con otras instituciones)

* + - ¿La propiedad de los datos y las responsabilidades para la gestión de los datos de investigación serán parte de algún convenio?
    - ¿Qué recursos requiere para ejecutar el plan de gestión de datos?
    - ¿Se requiere tener adicionalmente el apoyo de especialistas, por ejemplo, para dar entrenamiento o para administrar datos científicos?
    - ¿Se requiere hardware o software adicional al existente en la institución?

1. Diagrama

   Descripción generada automáticamenteEjemplos de modelos de planes de gestión de datos