Comparación de repositorios y propuesta para la gestión de datos en el área protegida ‘Bosque Pehuen’

*Ricardo Moreno G.*

Investigador Coop-Calahuala

Investigador asoc. ICBTe y Arboretum UACh

# Introducción

Las áreas protegidas son vitales para la protección de la biodiversidad, promover su resiliencia y como patrimonio cultural y ambiental. Sin embargo, el éxito y eficiencia de su conservación depende del manejo eficiente de datos y la información, factor muchas veces desestimado. Así como los ecosistemas enfrentan nuevas y más frecuentes amenazas, la habilidad para colectar, almacenar y aprovechar datos ecológicos de alta calidad es fundamental para tomar decisiones en áreas protegidas basada en evidencia.

Datos confiables sobre poblaciones, salud e integridad del hábitat, y amenazas permite a los manejadores tomar decisiones informadas sobre la priorización de intervenciones. Por ejemplo, el monitoreo de vegetación en parcelas permanentes puede revelar cambios en las especies debido al cambio climático, facilitando pautas y guiando el manejo adaptativo. Compartir los resultados y los aprendizajes en la práctica de la conservación es un factor clave para avanzar en el conocimiento colectivo de una especie y su hábitat y que permite generar redes de colaboración entre múltiples actores. Sin embargo, es necesario considerar que sobre todo en temas de interés colectivo se requiere que los datos sean confiables donde tanto los donantes como el gobierno y la sociedad civil demandan trasparencia en el manejo de la información. Por lo mismo, es importante crear sistemas robustos para prevenir malas prácticas en el manejo de datos ya que pueden afectar años de esfuerzo.

A pesar de la importancia, el manejo de datos en las áreas protegidas es desafiante y un camino difícil de sobre llevar. En muchos casos los datos quedan atrapados en hojas de Excel en múltiples versiones, en correos o archivos locales de algún miembro del equipo administrativo (quienes a veces además migran a otras instituciones). También sucede que las áreas protegidas funcionan normalmente con recursos limitados por lo que carecen de personal capacitado y/o con tiempo suficiente de dedicación para administrar los datos.

Una gestión eficaz de datos en el monitoreo ecológico para la conservación en áreas protegidas implica equilibrar la accesibilidad, la utilidad y la gobernanza. En el Bosque Pehuen, al igual que otras áreas protegidas, los datos son la columna vertebral invisible de la conservación. Gestionarlos con prudencia no es solo una cuestión técnica: es ética, estratégica y urgente. Al adoptar prácticas sistemáticas de gestión de datos, las áreas protegidas pueden transformar las observaciones sin procesar en información útil, garantizando así el desarrollo sostenible de los ecosistemas para las generaciones futuras.

En este documento se realiza una breve revisión de algunas de las distintas alternativas existentes actualmente. En la primera sección de este informe se describe un marco híbrido de gestión de datos para áreas protegidas, basado en el intercambio abierto para conjuntos de datos ecológicos no sensibles; junto con la presentación de herramientas escalables adaptadas a las limitaciones de recursos y que promueven el trabajo colaborativo en redes internacionales. En la segunda sección se esboza un plan preliminar de manejo de datos en el Bosque Pehuen y que puede ir siendo modificado en el tiempo según las necesidades puedan ir cambiando.

# Tipos de repositorios de datos abiertos

En un mundo globalizado y con manejo de datos disponible en la red, existe una variedad de repositorios, desde aquellos gestionados por instituciones académicas o gubernamentales hasta repositorios específicos para determinados tipos de datos (p.ej., genómicos, ambientales). Las tecnologías de la información están permitiendo recopilar grandes cantidades de datos en los llamados *Big data, Open data, Sharing data.* Es en esta plétora de plataformas es fácil confundirse entre las que han sido creadas por instituciones, gubernamentales, universidades, ONG, editoriales, entre muchas otras.

Un esfuerzo por parte de la comunidad científica para establecer unos principios que permitan una gestión eficaz de los datos. Este marco de trabajo, conocido como FAIR (Findable, Accessible, Interoperable, and Reusable), ofrece una serie de recomendaciones para asegurar que los datos sean encontrables, accesibles, interoperables y reutilizables. Este mismo marco se alinea con el principio de los estándares biertos de compartir los aprendizajes. En ecología uno de los repositorios (bibliotecas digitales) más reconocidos son: [Dryad](https://datadryad.org/search), [PANGAEA](https://www.pangaea.de/), [GBIF](https://www.gbif.org/es/), mientras que también existen otras más genéricas como [ZENODO](https://zenodo.org/), DataOne. También han aumentado las plataformas de colaboración de datos ecológica en áreas temáticas, como: Montanas (Geo Mountains), sitios de estudios de largo plazo (LTSER o ILTER), parcelas permanente o semi-permanentes ([sPlot](https://www.idiv.de/research/projects/splot/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc)) con enfoque global, nacional (p.ej., [Suiza](https://wp.unil.ch/ppch/) o [Nueva Zelanda](https://www.scionresearch.com/services/databases-and-collections/permanent-sample-plot-database)) o regional.

A continuación, se presenta una tabla comparativa entre las ventajas y desventajas de distintos tipos de repositorios, con algunas menciones claves a determinados tipos de repositorios.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Ventajas** | **Desventajas** | **Alternativas** |
| **Datos totalmente abiertos** (Repositorio público) | - Maximiza la colaboración: Los investigadores, otras áreas protegidas y los responsables políticos pueden usar los datos libremente.  - Transparencia: Genera confianza con las partes interesadas y el público.  - Preservación a largo plazo: Los repositorios (p. ej., **GBIF, Dryad, Zenodo**) garantizan la persistencia de los datos.  - Citas: Aumenta la visibilidad y el impacto del trabajo de la fundación. | - Riesgo de uso indebido: Los datos podrían malinterpretarse o usarse sin el contexto adecuado.  - Exposición de datos sensibles: La ubicación de especies en peligro de extinción o ecosistemas frágiles podría ser explotada.  - Pérdida de control: La fundación podría no realizar un seguimiento del uso de los datos. | - Períodos de embargo: Publicar los datos con un retraso (p. ej., 1 o 2 años) para que la fundación priorice su uso.  - Acceso por niveles: Metadatos abiertos, pero restringir los datos sensibles (p. ej., coordenadas de especies). |
| **Acceso controlado** (usuarios registrados o permisos) | - Intercambio dirigido: Los datos solo están disponibles para investigadores o instituciones autorizados.  - Gobernanza flexible: La fundación puede establecer condiciones (p. ej., memorandos de entendimiento, acuerdos de uso de datos).  - Protección de la información sensible: Las coordenadas o los datos de especies raras pueden ser censurados o difuminados. | - Carga administrativa: Requiere la gestión de solicitudes y permisos.  -Frena la colaboración: Las barreras pueden desalentar un uso más amplio. | - Redes de confianza: Compartir a través de plataformas como [DataONE](https://www.dataone.org/) o [**DEIMS-SDR de la red ILTER](https://deims.org**), que aplican estándares.  - Licencias por niveles: Utilizar licencias Creative Commons (p. ej., CC-BY para atribución, CC-BY-NC para uso no comercial). |
| **Datos gestionados por la fundación** (solo para uso interno) | - Control total: Garantiza que los datos se alineen con las prioridades de la fundación.  - Velocidad en la toma de decisiones: Sin retrasos por escrutinio externo ni disputas. | - Impacto limitado: Reduce las oportunidades de aprendizaje interinstitucional. Riesgo de duplicación: Otros podrían replicar esfuerzos sin saberlo. | - Informes anuales: Compartir los hallazgos agregados públicamente, manteniendo los datos sin procesar internamente.  - Asociaciones: Colaborar selectivamente con organizaciones alineadas. |

## Enfoque híbrido: Lo mejor de ambos mundos para las áreas protegidas

- Metadatos públicos (diseño del estudio, métodos, estadísticas resumidas) + datos sin procesar restringidos (basados ​​en solicitudes).

- Embargo dinámico: Los datos se abren después de los hitos del proyecto (p. ej., decisiones de gestión tomadas).

Decisiones claves:

1. Clasificar la sensibilidad de los datos: Identificar qué datos pueden ser abiertos (p. ej., tendencias de la vegetación) y cuáles restringidos (p. ej., presencia de especies raras y amenazadas).

2. Usar formatos estándar: Garantizar la compatibilidad con sistemas globales (p. ej., Darwin Core para la biodiversidad).

3. Documentación: Incluir metadatos, licencias y directrices de uso para evitar el uso indebido.

4. Involucrar a las partes interesadas: Consultar con la fundación, los investigadores y los gestores locales sobre sus necesidades.

Para un enfoque híbrido (que combina datos abiertos y de acceso controlado), la plataforma ideal debería:

* Admitir acceso diferenciado (datos públicos vs. restringidos).
* Integrarse con estándares de datos ecológicos (p. ej., Darwin Core, ISO 19115).
* Ofrecer preservación a largo plazo y asignación de DOI.
* Permitir la personalización de la marca y los logotipos de la fundación.

### Para datos de biodiversidad global (FAIR + Acceso híbrido)

Plataforma: GBIF (Infraestructura de Información sobre Biodiversidad Global)

Ventajas:

* Gratuita, ampliamente utilizada y compatible con Darwin Core.
* Permite ocultar datos sensibles (p. ej., generalizar coordenadas).
* DOI para citas; se integra con publicaciones académicas.

Desventajas:

* Menor flexibilidad para datos no relacionados con la biodiversidad (p. ej., encuestas climáticas o sociales).
* Totalmente pública; sin permisos de usuario.
* Solución alternativa: Publicar solo datos no sensibles en GBIF; almacenar los datos restringidos en otro lugar.

Alternativa: DataONE, o una red de repositorios (p. ej., KNB, Dryad, Pangaea) que admite diversos datos ecológicos. Algunos nodos miembros como Pangaea permiten acceso restringido.

### Para Colaboración Académica (Seguro + Citable)

Plataforma: Zenodo, Dryad, Pangaea

Ventajas:

* Gratuito, asigna DOI y se integra con GitHub/publicaciones.
* Zenodo permite embargos (p. ej., privado durante 2 años, luego público).

Desventajas:

* Control menos granular (no permite restringir subconjuntos de archivos).

## Opciones para Datos ecológicos en ambientes de montaña + bosques + largo plazo

Para su proyecto de área protegida de bosque de montaña (con datos de parcelas permanentes, composición de especies a lo largo del tiempo y registros de flora y fauna), aquí presentamos una comparación detallada de ILTER, GEO Mountains y SPlot, junto con sus ventajas y desventajas para su caso de uso.

### ILTER (Red Internacional de Investigación Ecológica a Largo Plazo)

Ideal para datos ecológicos estandarizados a largo plazo de parcelas permanentes. Algunos nodos ILTER permiten el intercambio de datos restringido.

Ventajas:

* Diseñado para parcelas permanentes: Admite datos de series temporales (p. ej., demografía arbórea, métricas del suelo).
* Integración en red local: Muchos sitios ILTER (p. ej., LTSER-Chile, US LTER, European eLTER) ofrecen marcos regionales para el intercambio de datos.
* Rigor de los metadatos: Requiere documentación detallada en EML (Lenguaje de Metadatos Ecológicos).

Desventajas:

* Burocrático: La adhesión puede requerir el registro formal del sitio o el cumplimiento de los protocolos ILTER.
* Menor visibilidad global: Los datos suelen estar aislados en redes nacionales/regionales, a menos que se publiquen en portales como DEIMS.
* Se recomienda utilizar si la fundación planea un monitoreo a largo plazo y desea alinearse con la comunidad global de LTER.

### GEO Mountains (Grupo de Observaciones de la Tierra)

Ideal para conjuntos de datos específicos de montaña + integración con datos de teledetección/clima. Abierta pero flexible: Fomenta los datos FAIR, pero no exige una apertura total.

Ventajas:

* Enfoque en la montaña: Adaptado a los ecosistemas alpinos (p. ej., biodiversidad, interacciones climáticas).
* Interoperabilidad: Enlaces a conjuntos de datos globales (p. ej., AppEEARS de la NASA, Copernicus).
* Relevancia política: Diseñado para fundamentar decisiones de conservación (se alinea con los objetivos de la fundación).

Desventajas:

* Menos especializada para datos a nivel de especie: Mejor para datos a mayor escala (paisaje/clima).
* Plataforma más reciente: Menos ejemplos consolidados de conjuntos de datos de flora/fauna.
* Es recomendable usarlos si los datos se combinarán con datos climáticos, de altitud o satelitales para realizar comparaciones entre sitios. No es excluyente de ILTER

### SPlot (Base de Datos Global de Parcelas de Vegetación)

Ideal para datos de composición de la vegetación (p. ej., estudios de comunidades vegetales). Se recomiendo cuando los datos de flora son una prioridad y puede compartirlos abiertamente.

Ventajas:

* El estándar de oro para datos de plantas: Utilizado por proyectos como sPlotOpen.
* Integración de rasgos: Vincula los registros de especies con rasgos funcionales (p. ej., área foliar, densidad de la madera).
* Incentivos de citación: Alta visibilidad en revistas de ecología.

Desventajas:

* Solo plantas: No está diseñado para datos de fauna ni abióticos.
* Totalmente abierto: Requiere publicación pública; no hay opción de acceso restringido.

## Recomendaciones para un enfoque híbrido

Combine plataformas:

* Usar ILTER para datos de parcelas permanentes (series temporales, acceso controlado).
* Comparta datos de vegetación a través de SPlot (si el acceso abierto es aceptable).
* Envíe datos resumidos a GEO Mountains para obtener información sobre políticas y otras regiones.

**Sugerencia fina**l: Comience con ILTER si su fundación valora la participación local a largo plazo. Combínelo con GEO Mountains para una mayor visibilidad. Evite SPlot si necesita o existen datos de fauna o acceso restringido.

# Comparación con otras plataformas usadas en conservación y reconocidas en los estándares abiertos

## SMART

La plataforma SMART (Herramienta de Monitoreo y Reporte Espacial) es una potente herramienta para el monitoreo de campo en áreas de conservación, pero su propósito es diferente al de ILTER, GEO Mountains o SPlot. A continuación, se presenta una comparación y un análisis detallados para determinar si SMART se ajusta a sus necesidades:

SMART es una herramienta de recolección de datos de campo y gestión de patrullas para la lucha contra la caza furtiva, el monitoreo de la fauna silvestre y la aplicación de la ley en áreas protegidas. Es utilizada por guardabosques y personal de campo para registrar observaciones (p. ej., avistamientos de animales, invasión humana, actividades ilegales).

Características principales:

✔ Aplicación móvil para la recolección de datos sin conexión (Android/iOS).

✔ Paneles de control en tiempo real para visualizar las actividades de patrullaje y las amenazas.

✔ Integración con GPS/sensores (p. ej., cámaras trampa, collares satelitales).

✔ Análisis centrado en la conservación (p. ej., puntos críticos de caza furtiva, cobertura de patrullaje).

Comparación de SMART con ILTER/GEO Mountains/SPlot

Aspecto SMART ILTER / GEO Mountains / SPlot

Propósito: Patrullas de campo, fuerzas del orden, investigación ecológica a largo plazo

Tipo de datos: Interacciones entre humanos y vida silvestre, registros de patrullaje, composición de especies, clima, rasgos

Acceso: Restringido (administradores/guardabosques), Abierto/híbrido (investigadores/público)

Estándares: Personalizado (formatos específicos de SMART): Darwin Core, EML, ISO 19115

Análisis: Mapeo de amenazas, métricas de cumplimiento, tendencias de biodiversidad, salud del ecosistema

Ventajas y desventajas

Ventajas de usar SMART

Eficiencia de campo:

Ideal si su fundación gestiona patrullas de guardabosques o necesita rastrear amenazas (p. ej., tala, caza furtiva).

Reemplaza las encuestas en papel por flujos de trabajo digitales.

Alertas en tiempo real:

Detecta anomalías (p. ej., ausencia repentina de una especie en un área de patrullaje).

Potencial de integración:

Permite incorporar observaciones incidentales de especies en GBIF o ILTER (si está configurado).

Desventajas (Por qué no es una alternativa directa):

No está diseñado para la investigación ecológica:

Carece de compatibilidad con datos de parcelas permanentes, rasgos de especies o análisis de series temporales.

Interoperabilidad científica limitada:

Los datos no están formateados para plataformas como GBIF o SPlot.

Ecosistema cerrado:

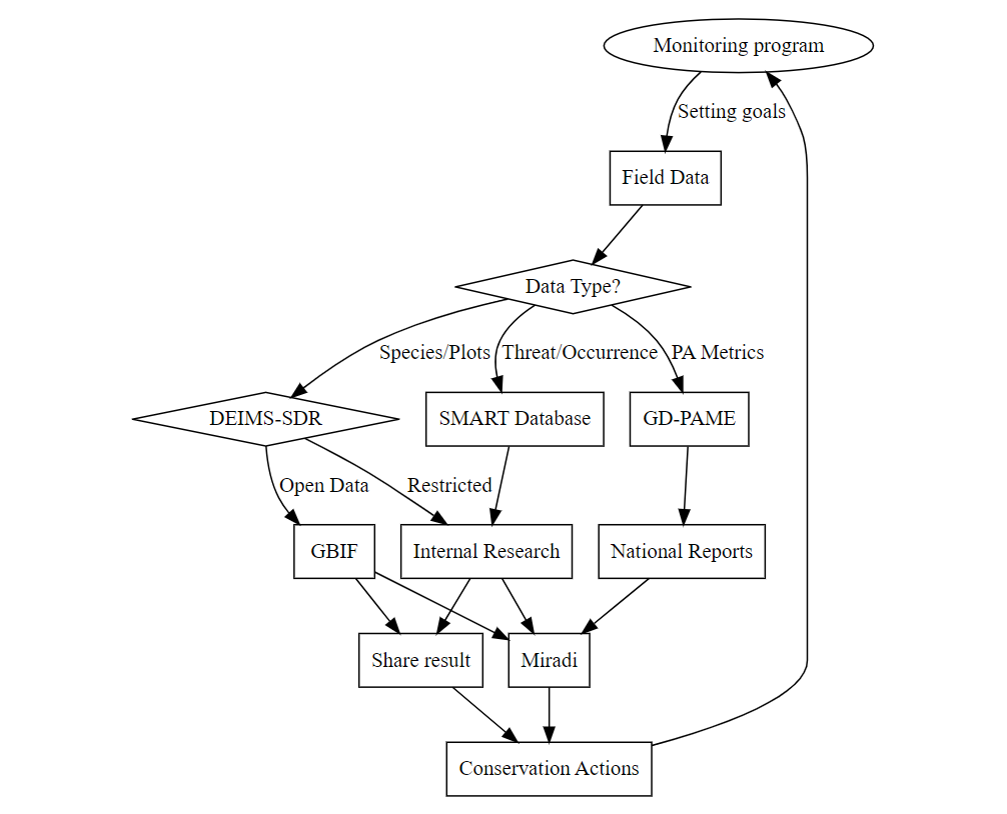
Menor flexibilidad para la colaboración más allá del personal de conservación.

No obstante, es posible usar SMART junto con ILTER/GEO Mountains el equipo realiza patrullajes regulares y necesita registrar amenazas/avistamientos. Prioriza la lucha contra la caza furtiva o la protección del hábitat.

## MIRADI

Miradi es una herramienta de software especializada diseñada para la gestión de proyectos de conservación, pero **no es una plataforma principal de gestión o almacenamiento de datos** los propósitos reconocidos de Miradi, son:

* Planificación y monitoreo de la conservación
  + Ayuda a los equipos a diseñar modelos de teoría del cambio, marcos lógicos y evaluaciones de amenazas (p. ej., estándares de la Asociación de Medidas de Conservación de la UICN).
  + Realiza un seguimiento de los indicadores del proyecto (p. ej., "% de aumento en avistamientos de tigres"), pero no almacena conjuntos de datos sin procesar.
* Manejo limitado de datos
  + Puede adjuntar datos resumidos (p. ej., archivos de Excel) a los proyectos, pero carece de: Ingesta masiva de datos.
  + API para vincular con repositorios como GBIF o CKAN.
  + Control de versiones o estándares de metadatos (p. ej., EML, Darwin Core).
* Visualización e informes
  + Genera paneles de control para donantes y partes interesadas (p. ej., progreso en la reducción de amenazas).



# Plan de manejo de datos

A continuación, se muestra un resumen de un Plan de gestión de datos estructurado (DMP) adaptado a un Proyecto de Monitoreo Ecológico de Bosques de Montaña, alineado con principios FAIR y diseñado para compartir datos bajo un formato híbrido.

|  |  |
| --- | --- |
| Presentación metadata |  |
| Título del proyecto | Monitoreo ecológico en [Nombre del área protegida] |
| Organización principal | [Su base] |
| Socios | [Listas de colaboradores, por ejemplo, investigadores, ONG] |
| Cuerpo de financiación | [Nombre del financiador] |

## Tipos de datos y recopilación

|  |  |  |  |
| --- | --- | --- | --- |
| Tipo de datos | Método de colección | Frecuencia | Sensibilidad |
| Ocurrencias de especies | Parcelas permanentes, transeptos, trampas de cámara, etc. | [Trimestralmente] | Bajo-Alto (especie rara) |
| Composición de vegetación | Muestreo en transeptos o parcelas | [ 1-4 años ] | Bajo |
| Datos de patrulla/amenaza | Aplicación o cuaderno de campo (registros de guardabosques) | [Semanal ] | Alto (datos de ubicación) |
| Clima/Suelo | Sensores, análisis de laboratorio | [Continuo ] | Bajo |

## Almacenamiento de datos y copia de seguridad

### A corto plazo (datos del campo al servidor)

- Smart/Kobo Toolbox: Sincronización con almacenamiento en la nube cifrado (AWS S3/NextCloud).

- Dispositivos de campo: tabletas con tarjetas SD cifradas; Datos cargados; libretas y formularios de campo traspasados a archivos Excel, Texto, o CSV.

### a largo plazo (archivo)

|  |  |  |
| --- | --- | --- |
| Protocolo de respaldo | Tipo de datos | Plataforma de almacenamiento |
| Copias de seguridad incrementales diarias + AWS Glacier | Datos de campo sin procesar | Portal CKAN (auto-hospedado) |
| Automatizado a través de API | Conjuntos de datos procesados ​​ | [GBIF, Geo Mountains] |
| Protegido con contraseña, requerido 2FA | Datos confidenciales | [Nodo Ilter (DEIMS-SDR) ] |

## Intercambio de datos y acceso

### Modelo de acceso escalonado

|  |  |  |  |
| --- | --- | --- | --- |
| Nivel | Ejemplos de datos | Plataforma | Política de acceso |
| Público | Tendencias de especies agregadas | GBIF, Geo Mountains | CC-By 4.0 u otro mas restrictivo |
| Restringido | Coordenadas de especies raras | Ckan/**llter** | Basado en solicitudes |
| Interno | Registros de patrulla, mapas de amenazas | SMART [almacenamiento de datos cualquiera] | Guarda parques, administradores e investigadores asociados |

PERÍODOS DE EMBARGO

- Datos públicos: Lanzado después de 1 año (para priorizar el uso de la base).

- Datos confidenciales: restricción indefinida; Versiones generalizadas compartidas (por ejemplo, cuadrícula de 10 km).

## Metadatos y documentación

-Estándar: lenguaje de metadatos ecológicos (EML) para Ilter; ISO 19115 para datos espaciales.

-Campos mínimos requeridos:

- Título del proyecto, contacto PI

- Coordenadas geográficas (con notas de precisión)

- Métodos (por ejemplo, "Parcelas permanentes de 25 m², escala de cobertura de blanquecería Braun")

- Licencia (CC-By para datos públicos)

- Herramientas:

- SQL, Monga Data base, Access u otro sistema de manejo de bases de datos relacionales (idealmente de acceso libre)

- QGIS para metadatos espaciales.

## Procesamiento de datos y control de calidad

### Flujo de trabajo

1. Validación de campo: Los valores atípicos de los Guarda Parques (por ejemplo, puntos GPS imposibles) en Smart.

2. Verificación automatizada de datos:

- R/python scripts para detectar duplicados, errores taxonómicos (a través de [GBIF Name Parser] (https://www.gbif.org/tools/species-lookup)).

- Filtros de sensibilidad: Ubicaciones de desenfoque de las especies de la lista roja de la UICN.

3. Control de versión:

- Datos sin procesar (inmutables, archivados).

- Datos procesados ​​(versiones etiquetadas).

## Preservación y longevidad

- Formatos:

- Tabular: CSV (UTF-8, Darwin Core).

- Espacial: Geopackage (no archivos de forma).

- Repositorios:

- GBIF, Zenodo, Dryad, Pangaea (con doi) para conjuntos de datos finales.

- Copias locales: Regla 3-2-1 (3 copias, 2 tipos de medios, 1 fuera del sitio).

## Roles y responsabilidades

|  |  |  |
| --- | --- | --- |
| Contacto | Rol | Tareas |
| [Nombre/correo electrónico] | Dirección programa de monitoreo | Supervisión y coordinación general |
| [Nombre/correo electrónico] | Administrador de datos | Asegura el cumplimiento de los metadatos, las copias de seguridad |
| [Nombre/correo electrónico] | Coordinador de campo | Valida los datos Smart u otra fuente |
| [Nombre/correo electrónico] | Analista de datos | Analiza información local para evaluar lineamientos de las estrategias de conservación |
| [Enlace de plantilla de mou] | Investigadores externos | Análisis de datos provenientes de estudios, Solicitudes de datos restringidos |

## Presupuesto

Muchas de las alternativas planteadas y recomendadas hasta ahora son de libre acceso. Aunque existen para algunas funciones a lo largo del proceso alternativas pagadas, lo más recomendable es mantener personal capacitado para que cumplan los distintos roles que se han identificado y que se puedan necesitar. El tiempo en estados iniciales puede ser relativamente bajo ya que se espera que el volumen de información a manejar es poco y del mismo modo, el flujo de solicitudes y entregas también debería serlo.