# Plan de Gestión de Datos – Proyecto: Seguro Indexado 2025

## Información general del proyecto

**Título:** Diseño y desarrollo de un modelo de seguro indexado basado en datos climáticos y agrícolas.  
**Equipo:** Lizeth Daniela Ortiz Perdomo, Diego Dayan Niño Pérez, Miguel Mateo Sandoval Torres, Camilo Andrés Flórez Esquivel  
**Institución:** Universidad de los Andes – MIAD 2025.  
**Duración estimada:** Octubre – Noviembre 2025.  
**Versión del documento:** 1.0  
**Última actualización:** Octubre 2025.

## 1. Descripción del proyecto

Este proyecto propone un seguro agrícola indexado que active pagos de forma automática a los agricultores cuando se cumplen condiciones climáticas objetivas (por ejemplo, sequías o excesos de lluvia), medidas a través de indicadores satelitales y meteorológicos como NDVI, lluvia acumulada y temperatura. La idea es reducir al mínimo el descalce entre el índice y las pérdidas reales, para que los pagos sean oportunos, transparentes y verificables.

En términos prácticos, construiremos y validaremos un índice climático por región, documentaremos todo el flujo de datos y dejaremos un pipeline reproducible para futuras actualizaciones.

**Propósito.** Contribuir a la gestión del riesgo climático en el sector agropecuario, apoyándonos en datos abiertos, buenas prácticas de gobierno de datos y modelos reproducibles.

## 2. Recolección y generación de datos

Datos a recolectar o crear:

- Datos meteorológicos (precipitación, temperatura, humedad relativa) de IDEAM, NASA POWER, CHIRPS y ERA5.  
- Datos satelitales de vegetación (NDVI) de MODIS o Sentinel‑2.  
- Registros históricos de pérdidas agrícolas y pagos de seguros (Finagro, UPRA o aseguradoras), previamente anonimizados.  
- Salidas generadas por el modelo: índices climáticos ajustados por región y reportes de activación.

Obtención de datos:

- Descarga automatizada mediante APIs o scripts en Python.  
- Anonimización de cualquier campo sensible antes del análisis (por ejemplo, identificación del productor o coordenadas exactas).  
- Elaboración de un diccionario de datos con nombre de variable, definición, unidad, fuente y fecha de actualización.

## 3. Documentación y metadatos

Cada conjunto de datos incluirá un archivo README.md y un diccionario de variables. Usaremos metadatos con base en Dublin Core (fuente, fecha de descarga, responsable, frecuencia de actualización y transformaciones aplicadas). El versionamiento de datos y experimentos se controlará con DVC y GitHub (repositorio privado durante el desarrollo).

## 4. Ética y cumplimiento legal

- Los datos públicos se utilizarán respetando sus licencias (p. ej., CC BY 4.0).  
- Cualquier dato privado se manejará bajo acuerdo de confidencialidad y con acceso restringido.  
- Eliminaremos o agregaremos la información que pueda identificar personas o predios (p. ej., coordenadas exactas).  
- Cumpliremos la Ley 1581 de 2012 (Protección de Datos Personales en Colombia) y políticas institucionales.

## 5. Almacenamiento y respaldo

- Almacenamiento principal en AWS S3 con control de accesos (MFA/roles).  
- Copias de seguridad en OneDrive institucional.  
- Backups semanales y versionamiento mediante GitHub + DVC.  
- Accesos limitados al equipo del proyecto y registro de cambios en el repositorio.

## 6. Selección y preservación

- Conservaremos datasets intermedios relevantes, el índice final y la documentación técnica.  
- Eliminaremos datos brutos que contengan información restringida al cierre del proyecto.

## 7. Compartición de datos

- Publicaremos scripts y resultados agregados bajo licencia MIT o CC BY 4.0.  
- Los datos de terceros permanecerán privados o se proveerán en forma agregada.  
- Se disponibilizará un dashboard con resultados regionales y guía de uso.  
- Cualquier solicitud adicional de datos será evaluada por el equipo y el tutor académico.

## 8. Responsabilidades y recursos

Roles del equipo:

- Lizeth Daniela Ortiz Perdomo: Coordinación general, redacción del DMP y control de versiones.  
- Diego Dayan Niño Pérez: Ingesta e integración de datos satelitales y climáticos (APIs/ETL).  
- Miguel Mateo Sandoval Torres: Desarrollo del índice y visualizaciones; validación y experimentos.  
- Camilo Andrés Flórez Esquivel: Calidad de datos, pruebas de reproducibilidad y documentación técnica.  
- Profesor tutor: Acompañamiento metodológico, revisión ética y buenas prácticas de datos.

Recursos:

- Cuentas AWS Educate, DMPTool, GitHub y Google Colab.  
- Stack Python: Pandas, NumPy, Rasterio, xarray, Folium.  
- Almacenamiento: AWS S3 (≈50 GB) y OneDrive institucional.

## 9. Productos esperados

• Dashboard interactivo: visualización de activación de índices y pagos simulados (público; sin datos sensibles).  
• Documento técnico: metodología, resultados y validación (público; sin datos sensibles).  
• Dataset procesado: series de NDVI y precipitación por municipio con el índice calculado (acceso restringido).

## 10. Preservación a largo plazo

- Archivaremos resultados y scripts en GitHub.  
- Política de retención: 3 años después de cerrar el proyecto.  
- Mantendremos README.md, diccionario de datos y registro de versiones para facilitar la reutilización.