



Anexo 4. Contenido de Propuesta

Componente	Descripción esperada	Preguntas para la autoevaluación del investigador
Título del proyecto	Debe ser claro, específico y reflejar el propósito del trabajo con relación a datos e IA.	• ¿El título refleja el enfoque del proyecto hacia la gestión y publicación de datos? • ¿Es entendible para un público técnico y general?
Resumen ejecutivo	Síntesis del propósito del conjunto de datos, origen de los datos, licenciamiento, metodología y resultados esperados.	• ¿Resume claramente el propósito, fuente de datos y resultados esperados? • ¿Hace evidente la alineación con IA y principios FAIR?
Problema de investigación	Descripción clara de la naturaleza de los datos y como su curaduría, enriquecimiento o transformación contribuye al avance de la ciencia.	¿Se plantea cómo el conjunto de datos propuesto contribuirá a la comprensión o solución de una o varias problemáticas relevantes a nivel local, regional o nacional? ¿Describe claramente la propuesta la fuente del conjunto de datos existente, su estructura actual (variables, registros, formato), nivel de completitud y calidad, y cuenta con el soporte o autorización para su uso y publicación? ¿Se justifica la cantidad mínima de registros de acuerdo con la técnica prospectiva (Shallow vs Deep Learning)?
Estado del arte	Explora antecedentes, estudios previos, datasets con características similares o brechas existentes.	• ¿Incluye revisiones previas sobre datos disponibles o necesidades? • ¿Identifica claramente la novedad del valor agregado del dataset y de su publicación?
Pregunta de investigación	Debe derivarse del problema y guiar el propósito del dataset o su aplicación futura.	• ¿La pregunta es clara y viable de abordar con los datos propuestos? • ¿Permite formular objetivos medibles relacionados con ciencia abierta?



Objetivo general	Propósito concreto y medible, alineado con la publicación y documentación del dataset.	• ¿Describe lo que se espera lograr con la estructuración o documentación del dataset? • ¿Incluye el enfoque de calidad, apertura o potencial uso con IA?
Conocimiento que generará el proyecto	Describe qué aportes nuevos se obtendrán del procesamiento o documentación del dataset.	• ¿Se explica cómo este proyecto contribuye a la ciencia de datos abierta? • ¿Genera insumos útiles para entrenamientos futuros en IA?
Justificación del proyecto	Razones que respaldan el desarrollo del proyecto, impacto social, científico o institucional.	• ¿Se justifica la necesidad de curar/publicar los datos? • ¿Se describe el valor que aporta el dataset para investigación o toma de decisiones? ¿La propuesta explica de manera clara y argumentada por qué este conjunto de datos es relevante para el uso futuro en aplicaciones de IA? ¿Se identifica cómo el dataset puede contribuir a los ODS, o a los focos priorizados; ¿Salud Planetaria o Seguridad Alimentaria?
Consideraciones éticas	Describe los principios éticos y legales del uso de datos.	• ¿Existe autorización para el uso de los datos si proviene de terceros? • ¿Se consideran los principios éticos y legales para el manejo y gestión de los datos? • ¿Se consideran todos los aspectos éticos y legales relacionados a la Gestión de la información recolectada?
Metodología	Describir cómo se proyecta curar, enriquecer, anotar, y publicar los datos. En caso del dato estar en un formato privativo o binario especificar el proceso, método y herramienta para convertir el dato en dato abierto.	• ¿Se indican y se justifican brevemente los métodos para mejorar la calidad de los datos? • ¿Se mencionan técnicas específicas de anotación, limpieza o estructuración? • ¿Se menciona como se convertirán los datos de formato privativo o binario a abierto?



Productos resultados de la investigación	Especificar entregables como dataset documentado, informes técnicos, publicaciones.	• ¿Se proyecta un conjunto de datos completo, con metadatos y diccionario? • ¿Se incluyen posibles artículos u otros resultados académicos asociados?
Impacto, transferencia e innovación del proyecto	Debe evidenciar cómo los datos pueden ser reutilizados y aplicados por terceros.	• ¿Se identifica el valor del dataset para otros grupos, sectores o instituciones? • ¿Se contempla su integración a ecosistemas de ciencia abierta o IA?