

Condiciones de reutilización — borrador

Proyecto: JEM-Full — archivo del Jurado de Enjuiciamiento de Magistrados preparado para IA **Fase:** 1 del plan de [JEM.md](#) **Estado:** borrador — ninguna licencia ha sido adoptada todavía

Usa la convención de etiquetado de [marco-legal.md](#) §0: [Dato], [Inferencia], [Propuesta], [Verificar].

JEM.md §3.3 señala que la propuesta original pide "licencia o condiciones de reutilización" como indicador FAIR pero nunca declara bajo qué licencia se publica el propio dataset derivado. Este documento cierra esa omisión como borrador para decisión del equipo.

1. Principio rector

[Propuesta] **Nadie puede licenciar lo que no le pertenece.**

El proyecto no adquiere derechos sobre los documentos del JEM por el hecho de descargarlos, hashearlos o procesarlos. Lo que sí produce —y por tanto lo único que puede licenciar— son sus propias obras derivadas: el índice, los metadatos normalizados, las estructuras extraídas, el código y las mediciones.

2. Los cuatro objetos y sus regímenes

[Propuesta]

2.1 Documentos originales del JEM — no relicenciados

Binarios PDF, DOCX y DOC tal como los publica el órgano.

- Conservan íntegramente su régimen legal de origen.
- El proyecto **no los republica como copia auténtica**, no altera su contenido y no reclama derecho alguno sobre ellos.
- La fuente autorizada sigue siendo el JEM. Cualquier uso con efecto jurídico debe remitirse al original del órgano, no a la copia del proyecto.
- [Verificar] Si el "Reglamento del Expediente Electrónico" del JEM contiene cláusulas expresas de reutilización o de autenticidad de copias, esta sección debe reescribirse conforme a ellas. Ver [marco-legal.md](#) §4.

2.2 Metadatos y estructuras derivadas — CC BY 4.0 propuesta

Inventario, manifiesto, tipos documentales, números y fechas de resolución, vínculos causa↔documento, entidades normalizadas de funcionarios públicos, votos estructurados, métricas de calidad e índice FAIR.

[Propuesta] **Creative Commons Atribución 4.0 Internacional (CC BY 4.0).**

Fundamento: es la licencia recomendada por la mayoría de las políticas de datos abiertos de investigación, es compatible con los principios FAIR —que exigen una licencia clara y legible por máquina— y no impone

restricciones que dificulten la reutilización académica. [Inferencia] Una licencia más restrictiva (NC, ND) contradiría el propio argumento del proyecto sobre preparación para la reutilización.

2.3 Texto extraído por OCR — decisión pendiente

[Propuesta] **No publicar hasta cerrar la Fase 6 y decidir sobre seudonimización.**

El texto OCR es un híbrido incómodo: reproduce el contenido del original, pero con errores introducidos por el proyecto. Publicarlo sin su tasa de error medida equivale a difundir una versión degradada de un documento público sin advertirlo.

[Propuesta] Si se publica, debe acompañarse obligatoriamente de:

- la tasa de error medida (CER/WER global y por categoría crítica);
- la versión del motor de extracción y sus parámetros;
- una advertencia visible de que **no es copia auténtica**;
- el enlace al documento original en el sitio del JEM, cuando exista.

2.4 Código y scripts — licencia pendiente de decisión

[Propuesta] Licencia permisiva, **MIT** o **Apache-2.0**. Apache-2.0 añade una concesión expresa de patentes y un requisito de constancia de cambios; MIT es más breve. [Inferencia] Para un proyecto académico sin componente patentable, MIT basta; si se prevé colaboración institucional, Apache-2.0 da más certidumbre.

Decisión abierta. Debe resolverse antes de la publicación del repositorio.

3. Texto de atribución propuesto

[Propuesta] Para quien reutilice los derivados bajo CC BY 4.0:

Datos derivados del proyecto JEM-Full (<autores>, <año>), a partir de documentos públicos del Jurado de Enjuiciamiento de Magistrados de la República del Paraguay. Corte del corpus: <identificador>, SHA-256 <hash de manifest.json>. Licencia CC BY 4.0. Los documentos originales no forman parte de esta licencia y conservan su régimen legal de origen.

[Inferencia] Incluir el identificador del corte y el hash en la atribución no es formalismo: es lo que hace que una cita de estos datos sea verificable. Es el mismo argumento que el artículo defiende, aplicado al propio artículo.

4. Advertencias obligatorias

[Propuesta] Acompañan a toda publicación del dataset derivado:

1. **No es copia auténtica.** Los datos derivan de un procesamiento automatizado con tasa de error conocida y distinta de cero. Carecen de valor probatorio.
2. **La fuente autorizada es el JEM.** Ante discrepancia entre este dataset y el documento del órgano, prevalece el documento del órgano.

3. **Corte temporal.** Los datos corresponden a un corte fechado y no reflejan actuaciones posteriores. El identificador del corte es parte del dato.
 4. **Incertidumbre explícita.** Los estados de determinación (`confirmado`, `probable`, `ambiguo`, `no_determinable`, `no_aplicable`) forman parte del dato. Colapsarlos a un valor binario o a cero **desvirtúa el dataset** y queda fuera del uso previsto.
 5. **Canal de corrección.** Toda persona mencionada puede solicitar rectificación o supresión en los datos derivados. Ver [marco-legal.md](#) §5.
-

5. Lista de control antes de publicar

- Decidir la licencia del código (MIT / Apache-2.0).
 - Confirmar CC BY 4.0 para los derivados, o adoptar alternativa razonada.
 - Revisar el Reglamento del Expediente Electrónico del JEM por cláusulas de reutilización.
 - Resolver si el texto OCR se publica, y bajo qué condiciones.
 - Definir la política de seudonimización para denunciante y terceros (ver [alcance-etico.md](#) §4.2).
 - Redactar la ficha del dataset (*datasheet for datasets*) exigida por [JEM.md](#) §4.4.
 - Establecer el canal de contacto y el compromiso de plazo de respuesta.
 - Elegir repositorio de publicación con DOI persistente (Zenodo u otro) — requisito de la dimensión *Findable* del índice FAIR.
 - Revisión del borrador por profesional del derecho paraguayo.
-

6. Lo que este borrador no resuelve

[Inferencia] Quedan tres tensiones sin resolver, que conviene nombrar en lugar de disimular:

1. **Apertura vs. protección.** CC BY 4.0 permite que un tercero reutilice los datos para exactamente aquello que [alcance-etico.md](#) §8 prohíbe al proyecto —perfilar personas, por ejemplo—. Una licencia abierta no puede impedirlo. [Propuesta] Asumirlo explícitamente y compensarlo en el diseño del dato: no publicar lo que no debería reutilizarse, en vez de publicarlo con una prohibición inaplicable.
 2. **Derecho de supresión vs. inmutabilidad del corte.** Si un titular ejerce su derecho de supresión, un corte hashado y publicado ya no puede modificarse sin romper su hash. [Propuesta] Versionar los cortes y declarar las supresiones como cambio de versión, manteniendo el registro de que hubo una supresión sin revelar su contenido.
 3. **Licencia del texto OCR.** Es un derivado de un documento público cuyo régimen no está claro. [Verificar] Requiere criterio jurídico profesional.
-

Documentos relacionados

- [marco-legal.md](#) — Leyes 5282/2014 y 7593/2025
- [alcance-etico.md](#) — alcance de recolección y publicación
- [../.. /MANIFEST.md](#) — corte vigente e identificador citable