

12 - 05 - 26

www.ecodes.org

ECODES Knowledge Platform

Almacén de datos curado y asistentes
de IA para la descarbonización

ÍNDICE

_Toc230695111

INTRODUCCIÓN Y MARCO CONCEPTUAL.....	5
1.1. Contexto.....	5
1.2. Resultados esperados	6
1.3. Enfoque metodológico	7
1.4. Alcance sectorial	8
1.5. Plataforma de referencia: ECODES Knowledge Platform (EKP)	8
1.6. Estructura del informe	9
FILOSOFÍA Y PRINCIPIOS DE ECODES KNOWLEDGE PLATFORM (EKP).....	11
2.1. Visión de la plataforma.....	11
2.2. Principios de diseño	11
2.3. El conocimiento como activo estratégico	13
2.4. Gobernanza distribuida del conocimiento	14
2.5. Filosofía de alineamiento	16
2.6. Relación con el diseño de un almacén de datos curado	17
GOBERNANZA DE DATOS	18
3.1. Sistema de gobernanza: visión general	18
3.2. Modelo de datos: EcodesKPFile	19
3.3. Taxonomía de 5 ejes.....	22
3.4. Curación experta: el rol del especialista.....	24
3.5. Inteligencia artificial para la gobernanza	27
3.6. Gestión de transcripciones.....	28
3.7. Conexión con Amazon S3 (seguridad).....	29
3.8. Permisos de usuarios.....	30
3.9. Panel de control y métricas.....	31
3.10. Configuración de taxonomías	33
3.11. Ejemplos de documentos gobernados	34
INFRAESTRUCTURA TECNOLÓGICA	36
4.1. Stack Técnico	36
4.2. Amazon S3: repositorio de documentos	37

4.3. Qdrant Cloud: base de datos vectorial	38
4.4. Modelo de embeddings	39
4.5. Re-ranking	40
4.7. Búsqueda semántica	41
4.8. Sincronización EKP ↔ Qdrant	42
4.9. Estrategias de tratamiento documental	43
PATRONES, BARRERAS Y OPORTUNIDADES	45
5.1. Análisis cuantitativo del almacén de datos	45
5.2. Distribución por corpus	45
5.3. Distribución por área estratégica	46
5.4. Distribución por tema	47
5.5. Asignación a asistentes de descarbonización	48
5.6. Análisis sectorial: alimentación	49
5.7. Análisis sectorial: energía eléctrica	49
5.8. Análisis sectorial: movilidad	50
5.9. Síntesis: mapa de madurez del almacén	51
5.10. Patrones transversales	51
ARQUITECTURA IA ESPECIALIZADA	52
6.1. Visión general de la arquitectura	52
6.2. Pipeline PRISMA: orquestador multi-corpus	52
6.3. Guardrails: sistema de seguridad multi-Capa	53
6.4. Auditor de alineamiento narrativo (PRISMA Verifier)	54
6.5. Asistentes especializados en descarbonización	56
6.6. Búsqueda web integrada	59
6.8. Servidores MCP: extensión del conocimiento	60
6.9. Auditoría de tokens y estimación de emisiones	62
6.10. Sistema de roles por asistente	65
6.11. Trazabilidad de fuentes y sistema de traza	67
6.12. Relación con la prueba de concepto de asistentes IA R.2.1.3	70
CONCLUSIONES Y RESULTADOS	71
7.1. Cumplimiento de resultados esperados	71
7.2. Métricas del almacén de datos	72
7.3. Líneas futuras de investigación	73
7.4. Conclusión	83

SOBRE EL INFORME

Este informe presenta una línea de investigación y desarrollo orientada a aplicar la inteligencia artificial a los procesos de mitigación, adaptación y seguimiento de la acción climática. Su propósito es explorar cómo los sistemas inteligentes pueden contribuir a identificar oportunidades de reducción de emisiones, mejorar la calidad de la información disponible, automatizar cálculos complejos, apoyar la rendición de cuentas y favorecer cambios de comportamiento en sectores clave. El informe es el resultado del trabajo llevado a cabo en el desarrollo del proyecto de investigación IAMA, y su aplicación en el desarrollo de herramientas concretas en IAMA.2.0. Hemos contado con la colaboración de la Universidad de Zaragoza en la elaboración del sistema y de las pruebas de concepto que presentamos a continuación.

CRÉDITOS

Editado por: ECODES (Fundación Ecología y Desarrollo)

Diseño Gráfico: BAMBAM Studio

Fecha: 12/05/2026

INTRODUCCIÓN Y MARCO CONCEPTUAL

1.1. Contexto

La aceleración de la descarbonización exige nuevas capacidades para transformar grandes volúmenes de información técnica, científica, normativa y operativa en decisiones eficaces, trazables y alineadas con los objetivos climáticos. En este contexto, la inteligencia artificial y la explotación avanzada de datos emergen como herramientas estratégicas para apoyar la transición ecológica, reforzar la acción climática y facilitar que administraciones, empresas, organizaciones y ciudadanía avancen hacia modelos compatibles con el límite de 1,5 °C establecido en el marco del Acuerdo de París.

La descarbonización de la economía y de los modelos de consumo requiere nuevas capacidades para traducir el conocimiento técnico y científico en decisiones concretas, verificables y accesibles. En este contexto, el desarrollo de asistentes de inteligencia artificial orientados a la acción climática representa una oportunidad estratégica para acelerar cambios sistémicos en organizaciones, sectores productivos y ciudadanía.

Este informe presenta una línea de trabajo centrada en el diseño de herramientas inteligentes capaces de acompañar procesos de descarbonización mediante información rigurosa, recomendaciones contextualizadas y apoyo a la toma de decisiones. El objetivo no es sustituir el criterio experto, sino ampliar su alcance, facilitando que datos complejos sobre emisiones, obligaciones legales, hábitos de consumo, movilidad, energía o alimentación puedan convertirse en orientaciones prácticas y adaptadas a distintos perfiles de usuarios

La investigación se articula en torno a tres ámbitos prioritarios por su peso en la huella de carbono y por su capacidad de generar cambios cotidianos y organizativos: la alimentación, el consumo eléctrico y la movilidad. En el ámbito alimentario, se exploran sistemas de recomendación que integren criterios como la reducción del desperdicio, la elección de productos de temporada y el impulso de dietas con mayor presencia vegetal. En materia de consumo eléctrico, el trabajo se orienta a mejorar la comprensión de la demanda distribuida y de las emisiones asociadas al Alcance 2. En movilidad, se abordan las oportunidades vinculadas al cambio modal, la electrificación del transporte y la planificación de trayectorias compatibles con los objetivos climáticos a 2030.

Para garantizar la utilidad y la credibilidad de estos asistentes, el proyecto combina análisis técnico, evaluación crítica de sesgos y desarrollo experimental. Se contempla la revisión de herramientas de inteligencia artificial existentes mediante

baterías de prompts específicos, con el fin de valorar su alineación con criterios climáticos, de salud pública y de reducción de emisiones. Asimismo, se prevé la creación de un almacén de datos curado, basado en fuentes científicas e institucionales, que sirva como base de conocimiento fiable para los prototipos. Estos desarrollos se validan mediante pruebas de concepto que permitan comprobar su capacidad real para acompañar decisiones de descarbonización en contextos cotidianos y profesionales.

En conjunto, esta línea de trabajo busca contribuir a que la inteligencia artificial se convierta en una herramienta al servicio de la transición ecológica, la transparencia y la acción climática efectiva. Su valor reside en conectar conocimiento, tecnología y toma de decisiones, favoreciendo soluciones prácticas que ayuden a organizaciones y ciudadanía a avanzar de manera más informada, justa y coherente hacia la descarbonización.

Este proyecto se ha desarrollado con el apoyo del Ministerio para la Transición Ecológica y el Reto Demográfico a través de la convocatoria para el año 2024 de subvenciones en régimen de concurrencia competitiva “línea a” para el desarrollo de actividades para el interés general consideradas de interés sociales, en el ámbito de la investigación científica y técnica y protección al medio ambiente en materias de competencia estatal.

1.2. Resultados esperados

El OE2.1 contempla dos resultados fundamentales que vertebran este informe:

R.2.1.2. Diseño de un almacén de datos curado

Diseño de un almacén de datos curado con fuentes científicas, institucionales y técnicas, alineado con criterios de descarbonización, identificando patrones, barreras y oportunidades de las medidas de mitigación relacionados con alimentación, consumo de energía eléctrica y movilidad, como base para arquitecturas especializadas de inteligencia artificial.

Este resultado implica la construcción de una infraestructura de conocimiento que no se limite a la mera acumulación de documentos, sino que incorpore:

- **Curación activa:** Clasificación taxonómica rigurosa por ejes temáticos y funcionales.

- **Fuentes heterogéneas:** Integración de corpus normativos (legislación europea, nacional y autonómica), corpus de evidencia científica (informes IPCC, estudios sectoriales, artículos técnicos) y corpus institucional (posicionamientos de ECODES, guías y manuales).
- **Alineamiento con descarbonización:** Toda la arquitectura de metadatos está diseñada para identificar y relacionar documentos con las áreas estratégicas de mitigación climática.
- **Identificación de patrones:** La estructura del almacén permite analizar la distribución del conocimiento por sectores (alimentación, energía, movilidad) y detectar brechas documentales.

R.2.1.3. Prueba de concepto de asistentes IA

Prueba de concepto para el desarrollo de asistentes de inteligencia artificial especializados en la descarbonización con foco en alimentación, consumo de energía eléctrica (A2) y movilidad.

Este resultado materializa el almacén de datos en asistentes conversacionales especializados, capaces de:

- **Responder** preguntas sobre descarbonización fundamentando sus respuestas en el corpus curado.
- **Realizar búsquedas semánticas multi-corpus** (normativo + evidencia + institucional).
- **Citar fuentes** de forma trazable y verificable.
- **Alinear sus respuestas** con los posicionamientos institucionales de ECODES.

1.3. Enfoque metodológico

La aproximación adoptada se fundamenta en tres pilares:

1. **Gobernanza del dato desde el origen:** cada documento que ingresa al sistema es sometido a un proceso de clasificación taxonómica supervisada, con asistencia de inteligencia artificial para la extracción automatizada de metadatos.

2. **Arquitectura RAG (Retrieval Augmented Generation) multi-corpus:** los asistentes no generan respuestas de memoria, sino que recuperan información relevante del almacén vectorial y la sintetizan con citas trazables.

3. **Alineamiento narrativo institucional:** un módulo de auditoría verifica que las respuestas generadas estén alineadas con los posicionamientos oficiales de ECODES sobre cada temática.

1.4. Alcance sectorial

El informe analiza tres sectores prioritarios para la descarbonización:

Sector	Relevancia para la descarbonización	Documentos gobernados
Alimentación	Los sistemas agroalimentarios representan entre el 21-37% de las emisiones globales de GEI (IPCC, 2022). La transformación de la cadena alimentaria es clave para los objetivos del Acuerdo de París.	47 documentos en área específica + 95 asignados al asistente especializado
Consumo de energía eléctrica	La transición energética es el pilar central de la descarbonización. La electrificación de la demanda y el despliegue de renovables son esenciales para alcanzar la neutralidad climática en 2050.	135 documentos en área específica + 158 asignados al asistente especializado
Movilidad	El transporte representa aproximadamente el 25% de las emisiones de CO ₂ en la UE. La movilidad sostenible requiere transformaciones sistémicas en infraestructura, tecnología y comportamiento.	17 documentos en área específica + 75 asignados al asistente especializado

1.5. Plataforma de referencia: ECODES Knowledge Platform (EKP)

Toda la infraestructura descrita en este informe se ha materializado en la **ECODES Knowledge Platform (EKP)**, una plataforma web desarrollada ad hoc que integra:

- Un sistema de **gobernanza documental** con interfaz de administración.
- Un **repositorio** de archivos en la nube (Amazon S3).
- Una **base de datos vectorial** para búsqueda semántica (Qdrant Cloud).
- Un marketplace de **agentes IA** especializados con pipeline RAG multi-corpus.
- Módulos de auditoría, guardrails de seguridad y alineamiento narrativo.

ECODES KNOWLEDGE PLATFORM (EKP)

Plataforma integral para la descarbonización

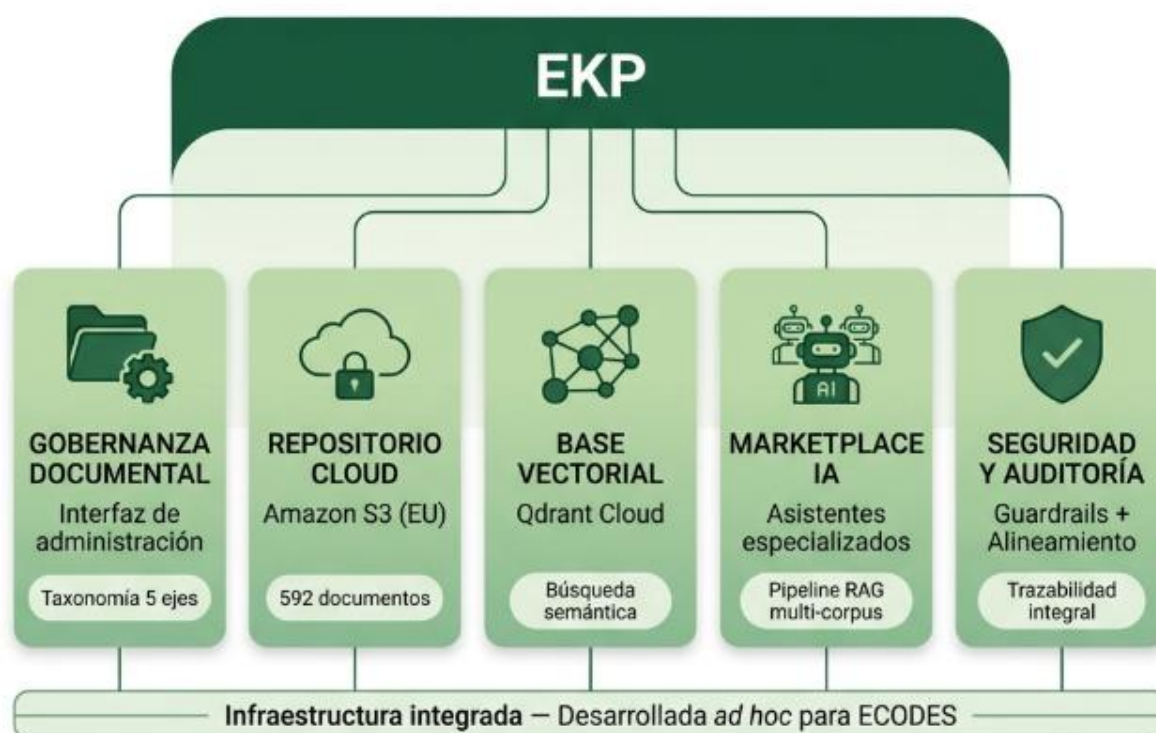


Diagrama de proceso de la plataforma de conocimiento EKP

1.6. Estructura del informe

El presente informe se organiza en siete capítulos:

Capítulo	Contenido
1. Introducción	Contexto, objetivos, enfoque y alcance

2. Filosofía de la EKP	Principios de diseño, visión y arquitectura conceptual
3. Gobernanza de Datos	Taxonomía, metadatos, IA para gobernar, trazabilidad
4. Infraestructura Tecnológica	Stack técnico, S3, Qdrant, embeddings, re-ranking
5. Patrones, Barreras y Oportunidades	Análisis sectorial del almacén de datos
6. Arquitectura IA Especializada	Pipeline PRISMA, guardrails, auditor de alineamiento
7. Conclusiones	Cumplimiento de resultados, métricas y líneas futuras

FILOSOFÍA Y PRINCIPIOS DE ECODES KNOWLEDGE PLATFORM (EKP)

2.1. Visión de la plataforma

ECODES Knowledge Platform (EKP) nace como respuesta a una necesidad estructural: la gestión del conocimiento institucional de ECODES —acumulado durante más de dos décadas de trabajo en sostenibilidad y acción climática— requiere una plataforma que garantice precisión, coherencia y trazabilidad en cada pieza de información que nutre a los asistentes de inteligencia artificial.

La visión fundacional de la EKP se articula en torno a un principio rector:

Desde esta plataforma se gestionan archivos, metadatos y configuraciones que garantizan precisión, coherencia y trazabilidad del conocimiento de ECODES.

Este principio trasciende la mera gestión documental para constituir un sistema de gobierno del conocimiento, donde cada documento es un activo curado con identidad propia, clasificación taxonómica multidimensional y ciclo de vida gestionado.

2.2. Principios de diseño

2.2.1. Trazabilidad integral

Cada documento en la EKP posee un identificador único determinista (Asset ID) con formato semántico:

CORPUS_AREA_TIPO_MMAA_ID

Por ejemplo: EVID_AS_IT_0325_683 identifica un documento del corpus de Evidencia, del área de Alimentación y Sistemas Agroalimentarios, de tipo Informe Técnico, registrado en marzo de 2025, con secuencial 683.

Este Asset ID acompaña al documento a lo largo de toda su vida en el sistema: desde su ingesta en la gobernanza, su almacenamiento en S3, su indexación vectorial en Qdrant, hasta su citación por un asistente IA en una conversación con el usuario final.

2.2.2. Curación humano-IA

La EKP implementa un modelo de curación híbrido:

- **Análisis automatizado con IA:** Al ingestar un documento, un modelo de lenguaje (Qwen2.5-72B-Instruct o similar) analiza su contenido y propone metadatos estructurados (corpus, área estratégica, tipo de contenido, destinatario, ámbito geográfico).
- **Validación estricta:** El sistema valida automáticamente que cada metadato propuesto por la IA pertenezca a las taxonomías cerradas definidas. Valores inventados son descartados antes de presentarse al usuario.
- **Supervisión humana final:** Un gobernante de datos revisa, ajusta y confirma la clasificación antes de persistirla.

Este enfoque asegura que el almacén de datos no sea ni puramente manual (lento y costoso) ni totalmente automatizado (propenso a errores).

2.2.3. Seguridad y control de acceso

La plataforma implementa un modelo de permisos basado en grupos y niveles de acceso granulares:

Rol	Capacidades
Administrador	Acceso total: usuarios, configuración, IA, gobernanza
Gestor de IA	Gestión de asistentes, modelos, system prompts, marketplace
Gobernante de datos	Gobernanza documental, clasificación, taxonomías
Publicador	Publicación de contenidos

El acceso a la gobernanza está restringido a usuarios con el nivel gobierno, mientras que el control de estadísticas requiere nivel administrador o mánager. Esta separación de responsabilidades garantiza que la curación del dato esté en manos de perfiles especializados.

2.2.4. Arquitectura modular

La EKP se estructura en módulos independientes e interconectados:

ECODES Knowledge Platform (EKP)

— Filosofía	→ Principios, arquitectura y gobierno del conocimiento
— Gobernanza	→ Administración de archivos, activos y metadatos
— Control	→ Panel de estadísticas y métricas
— Repositorio S3	→ Gestión de archivos en el bucket S3
— Almacenamiento	→ Buscador semántico, Qdrant Cloud, motores vectoriales
— Configuración	→ Taxonomías, áreas estratégicas, parámetros
— Inspector	→ Diagnóstico de integridad: metadatos faltantes en S3
— Posicionamientos	→ Posicionamientos institucionales y narrativas ECODES
— Ayuda	→ Documentación y guías de uso

Cada módulo tiene su propio conjunto de rutas API, permisos y vistas, permitiendo una evolución independiente sin afectar al resto del sistema.

2.3. El conocimiento como activo estratégico

La EKP trasciende el paradigma de la gestión documental tradicional al tratar cada documento como un activo estratégico con múltiples dimensiones:

Dimensión informacional

Cada documento contiene conocimiento relevante para la descarbonización, codificado en su texto y metadatos.

Dimensión taxonómica

Una clasificación multidimensional de 5 ejes (detallada en el Capítulo 3) permite la navegación, filtrado y recuperación inteligente del conocimiento.

Dimensión posicional

Los documentos de tipo "posicionamiento institucional" codifican la posición oficial de ECODES sobre temáticas específicas, sirviendo como referencia normativa para los asistentes IA.

Dimensión vectorial

Cada documento, una vez procesado, existe también como un conjunto de vectores de 1024 dimensiones en el espacio semántico, permitiendo la recuperación por significado más allá de la coincidencia literal de palabras.

Dimensión operativa

Los documentos tienen estados de ciclo de vida (activo, inactivo, archivado), niveles de acceso (público, privado, interno) y asignaciones a asistentes específicos, controlando dónde y cómo se utiliza cada pieza de conocimiento.

2.4. Gobernanza distribuida del conocimiento

La plataforma implementa una gobernanza del conocimiento que involucra a múltiples actores:

- **Gobernantes de datos:** Ingestan, clasifican y mantienen el corpus documental. Pueden usar la IA como asistente en el proceso de catalogación.
- **Gestores de IA:** Configuran asistentes, definen system prompts, asignan documentos a cada agente y calibran los parámetros de búsqueda semántica.
- **Administradores:** Gestionan taxonomías, permisos, configuración global y monitorizan la salud del sistema a través del panel de control.
- **Usuarios finales:** Interactúan con los asistentes IA especializados, que acceden al almacén curado para fundamentar sus respuestas.

Esta distribución de responsabilidades refleja un modelo de gobierno maduro donde la calidad del dato es responsabilidad compartida, pero con roles claramente definidos.

2.4.1. Comité de Gobernanza de Datos

Para garantizar la coherencia, calidad y evolución sostenida del almacén de datos, se ha constituido un **Comité de Gobernanza de Datos** como órgano transversal encargado de supervisar el ciclo de vida completo del conocimiento dentro de la EKP. Este comité responde a la necesidad de articular competencias diversas — documentales, técnicas, sectoriales y comunicativas— en un único espacio de decisión, evitando que la curación del dato recaiga en un perfil aislado o que las decisiones tecnológicas se tomen sin perspectiva de contenido.

El comité está integrado por un equipo multidisciplinar que reúne los perfiles imprescindibles para una gobernanza integral del conocimiento:

- **Documentalista:** responsable de la catalogación, clasificación taxonómica y control de calidad de los metadatos. Garantiza la coherencia del sistema de gobernanza, supervisa la integridad del modelo de datos (EcodesKPPFile) y

asegura que cada documento cumpla los estándares de descripción, identificación y ciclo de vida definidos.

- **Especialista en descarbonización:** aporta el criterio sectorial para la selección, evaluación y validación de las fuentes documentales. Determina la relevancia de cada documento para las áreas estratégicas de alimentación, energía eléctrica y movilidad, y valida el grado de alineamiento con los posicionamientos institucionales de ECODES.
- **Experta en comunicación:** supervisa la coherencia narrativa del conocimiento curado con la estrategia comunicativa de ECODES. Valida que los posicionamientos institucionales estén correctamente codificados y que las narrativas asociadas a cada subárea temática sean consistentes con el discurso público de la organización.
- **Especialista en IA y desarrollo tecnológico:** responsable de la arquitectura técnica de la plataforma: diseño de la infraestructura (bases de datos, almacenamiento vectorial, modelos de embeddings), configuración de los asistentes de IA, calibración de los parámetros de búsqueda semántica y mantenimiento de los flujos de indexación y sincronización.

COMITÉ DE GOBERNANZA DE DATOS

Funciones principales



Las **funciones principales** del Comité de Gobernanza incluyen:

- Definición y actualización de los criterios de selección de fuentes documentales.

- Supervisión de la calidad del almacén: revisión periódica de metadatos, detección de duplicados, documentos obsoletos o con clasificación incorrecta.
- Validación de las taxonomías y propuestas de ampliación de los ejes de clasificación.
- Evaluación de la cobertura temática del almacén e identificación de brechas documentales prioritarias para cada asistente de descarbonización.
- Coordinación entre la estrategia comunicativa institucional y los posicionamientos codificados en la plataforma, asegurando que los asistentes de IA reflejen fielmente la visión de ECODES.
- Decisiones sobre la incorporación de nuevas fuentes de datos y herramientas externas.

La existencia de este comité multidisciplinar constituye un elemento diferenciador de la EKP frente a sistemas de gestión documental convencionales: el conocimiento no se acumula de forma pasiva, sino que es activamente gobernado por un equipo que combina rigor documental, experiencia sectorial, sensibilidad comunicativa y capacidad tecnológica.

2.5. Filosofía de alineamiento

Un aspecto diferenciador de la EKP es su compromiso con el alineamiento narrativo institucional. Cada subárea temática puede tener asociados:

- **Posicionamiento oficial:** La posición de ECODES sobre ese tema (por ejemplo, su postura sobre la transición energética justa o la alimentación sostenible).
- **Narrativa:** Cómo ECODES comunica ese tema (tono, enfoque, énfasis).

Estos posicionamientos alimentan al **Auditor de Alineamiento** (descrito en el Capítulo 6), que verifica que las respuestas de los asistentes IA sean coherentes con la visión institucional. Esto garantiza que los asistentes no solo sean técnicamente precisos, sino también institucionalmente consistentes.

2.6. Relación con el diseño de un almacén de datos curado

La filosofía de la EKP es el fundamento conceptual del almacén de datos curado exigido por R.2.1.2:

Requisito R.2.1.2	Materialización en la EKP
Fuentes científicas	Corpus de Evidencia (245 documentos)
Fuentes institucionales	Corpus ECODES (110 documentos)
Fuentes técnicas	Corpus Normativo (236 documentos)
Alineado con criterios de descarbonización	Taxonomía de 5 ejes + áreas estratégicas de descarbonización
Identificando patrones	Panel de control con distribución estadística por corpus, área, tema
Barreras y oportunidades	Análisis de brechas documentales por sector
Alimentación, energía, movilidad	Áreas estratégicas específicas + asistentes especializados por sector
Base para arquitecturas de IA	Pipeline PRISMA multi-corpus + marketplace de agentes

GOBERNANZA DE DATOS

3.1. Sistema de gobernanza: visión general

La gobernanza constituye el pilar fundamental del almacén de datos curado. El módulo de Gobierno EKP proporciona una interfaz integral para la gestión de archivos, metadatos y configuraciones que garantizan la precisión, coherencia y trazabilidad del conocimiento.



El sistema de gobernanza implementa un flujo de trabajo completo que abarca desde la ingesta del documento hasta su disponibilidad para los asistentes de IA, pasando por una clasificación taxonómica rigurosa, almacenamiento seguro en la nube y sincronización con la base de datos vectorial.

FLUJO DE GOBERNANZA EKP

Del documento fuente a la respuesta del asistente



3.2. Modelo de datos: EcodesKPFile

El corazón del sistema de gobernanza es el modelo de datos **EcodesKPFile**, que representa cada documento gobernado con más de 40 campos estructurados. Este modelo, implementado sobre PostgreSQL, captura todas las dimensiones relevantes de un activo documental.

☐	ID ▼	Título	Corpus	Área	Tipo	Usos	Estado	Acceso	Acciones
☐	685	Manual de Compostaje	Evidencia	Economía Circular/residuos	Guía práctica	Asistente Hostelería PorElClima	activo	publico	
☐	683	FASE 1: PREPARACIÓN PARA LA HOJA DE RUTA DE DESCARBONIZACIÓN DE LA INDUSTRIA DE ALIMENTACIÓN Y BEBIDAS EN ESPAÑA	Evidencia	Alimentación y Sistemas Agroalimentarios	Informe técnico	Asistente descarbonización alimentación	activo	publico	
☐	682	Tecnologías para la Descarbonización Industrial	Evidencia	Transición Justa	Informe técnico	Asistente descarbonización alimentación	activo	publico	
☐	681	Estado actual y perspectivas de la descarbonización en España	Evidencia	Energía	Informe técnico	Asistente descarbonización alimentación Asistente descarbonización energía eléctrica Asistente descarbonización movilidad Asistente Hostelería PorElClima	activo	publico	

Gobernanza de datos en EKP

3.2.1. Identificación y metadatos básicos

Campo	Tipo	Descripción
`id`	Integer (PK)	Identificador secuencial interno
`asset_id`	String (único)	Identificador determinista del activo ('CORPUS_AREA_TIPO_MMAA_ID')
`titulo`	String	Título oficial y completo del documento
`corpus`	String	Corpus al que pertenece (ECODES, Normativo, Evidencia, Contraste)
`tipo_activo`	String	Tipo de contenido según taxonomía cerrada
`area_tematica`	String	Área estratégica de descarbonización

3.2.2. Contenido y descripción

Campo	Tipo	Descripción
`descripcion_corta`	Text	Resumen del documento (máximo 280 caracteres)
`resumen_largo`	Text	Transcripción o resumen detallado del documento
`autor`	String	Autor/autores del documento
`fecha_publicacion`	Date	Fecha de publicación original
`idioma`	String	Idioma del documento (defecto: "es")

3.2.3. Clasificación taxonómica extendida

Campo	Tipo	Descripción
`subareas_tematicas`	JSONB (array)	Subáreas temáticas asociadas (Eje 2)
`publico_objetivo`	JSONB	Audiencia/destinatario del documento
`ambito_geografico`	String	Ámbito geográfico (Eje 6)
`pais_iso`	JSONB (array)	Países relacionados (códigos ISO)
`comunidad_autonoma`	String	Comunidad autónoma (España)

3.2.4. Gobernanza y ciclo de vida

Campo	Tipo	Valores	Descripción
`estado`	String	`activo`, `inactivo`, `archivado`	Estado del ciclo de vida
`acceso`	String	`privado`, `interno`, `publico`	Nivel de acceso
`responsable`	String	—	Persona responsable del documento
`notas`	Text	—	Notas de gobernanza

3.2.5. Posicionamiento institucional

Campo	Tipo	Descripción
`es_posicionamiento_ecodes`	Boolean	Indica si es un posicionamiento oficial
`tipo_posicionamiento`	String	Criterio, Recomendación, etc.
`grado_alineamiento_ecodes`	String	Alto, Medio, Bajo
`base_argumental`	Text	Argumentación de respaldo

3.2.6. Integración con asistentes IA

Campo	Tipo	Descripción
`usos_archivo`	JSONB (array)	Asistentes IA a los que está asignado el documento

3.2.7. Auditoría

Campo	Tipo	Descripción
`created_at`	DateTime	Fecha de creación del registro
`updated_at`	DateTime	Fecha de última modificación
`created_by`	FK → User	Usuario que creó el registro
`modified_by`	FK → User	Usuario que realizó la última modificación

3.3. Taxonomía de 5 ejes

La clasificación taxonómica es el mecanismo fundamental de organización del almacén de datos. Se estructura en 5 ejes cerrados más un eje abierto:

Eje 1. Área estratégica

Define el dominio temático principal del documento en relación con la descarbonización.

Cada área posee:

Área	Documentos	Relevancia para descarbonización
Clima y Energía	309	Mitigación, adaptación, transición energética
Transición Energética	135	Renovables, eficiencia, electrificación
Alimentación y Sistemas Agroalimentarios	47	Dietas sostenibles, producción, distribución
Economía Circular/residuos	34	Reducción de materiales, reciclaje
Movilidad sostenible	17	Transporte limpio, electromovilidad
Agua	13	Gestión del recurso, huella hídrica
Transformación Digital	12	IA, digitalización para sostenibilidad
Inclusión Social	10	Transición justa, desigualdad
Biodiversidad	8	Ecosistemas, naturaleza
General	7	Multisectorial

Eje 2. Subárea temática

Desglose más granular dentro de cada área estratégica. Cada subárea puede tener asociados:

- **Posicionamiento:** La posición oficial de ECODES sobre esa temática.
- **Narrativa:** El enfoque comunicativo institucional.

Estas subáreas alimentan el sistema de alineamiento narrativo (Capítulo 6).

Eje 3. Tipo de contenido

Clasifica la naturaleza formal del documento:

Tipo	Descripción
Informe técnico	Informes de investigación o análisis sectoriales
Marco conceptual	Documentos teóricos de referencia
Guía práctica	Manuales y guías de implementación
Artículo de investigación	Publicaciones científicas o académicas
Normativa	Legislación, directivas, reglamentos
Posicionamiento institucional	Posiciones oficiales de ECODES
Estrategia/Plan	Planes estratégicos, hojas de ruta
Datos/Estadísticas	Datasets, indicadores, series temporales

Eje 4. Destinatario

Define el público objetivo del documento:

Destinatario	Descripción
Interno ECODES	Documentos de uso interno
Empresas	Sector empresarial
Administraciones públicas	Gobierno, ayuntamientos
Sociedad civil	Ciudadanía, ONGs
Universidades/investigación	Comunidad académica
Medios de comunicación	Prensa, divulgación

Eje 5. Ámbito geográfico

Define la cobertura territorial del documento:

Ámbito	Ejemplos
Internacional / Global	Informes IPCC, acuerdos internacionales
Unión Europea	Directivas, estrategias EU
España	Legislación nacional, PNIEC
Comunidad Autónoma	Leyes autonómicas, planes regionales
Ámbito local / municipal	Ordenanzas, planes municipales

Eje 6 (abierto). Usos del archivo

Este eje permite la asignación flexible de documentos a asistentes IA y otros usos. A diferencia de los ejes cerrados, los usos se definen dinámicamente y permiten:

- Asignar un documento a uno o varios asistentes IA.
- Definir usos restringidos solo para administradores.
- Crear nuevos usos sin modificar la taxonomía base.

Ejemplos de usos configurados:

- Asistente descarbonización alimentación
- Asistente descarbonización energía eléctrica
- Asistente descarbonización movilidad

3.4. Curación experta: el rol del especialista

3.4.1. El factor humano como garantía de calidad

El almacén de datos curado no es el resultado de un proceso puramente tecnológico. Detrás de cada documento gobernado hay un trabajo de selección, evaluación y curación realizado por especialistas de la plantilla de ECODES con conocimiento sectorial profundo en sus respectivas áreas de descarbonización.

Este enfoque diferencia a la EKP de los sistemas de recopilación automatizada de documentos (web scraping, crawlers), donde la calidad del corpus depende exclusivamente de algoritmos. En la EKP, la tecnología (IA para análisis de metadatos, búsqueda semántica, embeddings) es un amplificador del criterio humano, no un sustituto.

3.4.2. Perfil de los curadores

Los especialistas que participan en la curación del almacén de datos reúnen las siguientes características:

- **Expertise sectorial:** Conocimiento técnico en descarbonización, transición energética, alimentación sostenible, movilidad o economía circular adquirido a lo largo de años de trabajo en ECODES.
- **Criterio institucional:** Comprensión profunda de los posicionamientos de ECODES, lo que permite evaluar si un documento es relevante no solo

técnicamente sino también desde la perspectiva estratégica de la organización.

- **Capacidad de evaluación de fuentes:** habilidad para distinguir entre fuentes rigurosas (publicaciones científicas, organismos oficiales) y fuentes de menor fiabilidad.

3.4.3. Criterios de selección de contenidos

La selección de documentos para el almacén se rige por criterios explícitos:

Criterio	Descripción	Ejemplo
Relevancia temática	El documento aborda directamente una temática de descarbonización	Informe IPCC sobre mitigación en sistemas alimentarios
Rigor científico	Procedencia de fuentes reconocidas (instituciones, universidades, organismos internacionales)	Directiva europea de eficiencia energética
Vigencia	El documento está actualizado y sus datos/recomendaciones siguen siendo aplicables	PNIEC actualizado 2024, no la versión de 2019
Alineamiento institucional	El contenido es coherente o relevante para los posicionamientos de ECODES	Estudio sobre transición justa y pobreza energética
Utilidad para asistentes	El documento tiene potencial para ser útil en respuestas de los asistentes IA	Guía práctica sobre comunidades energéticas
Complementariedad	El documento aporta perspectivas no cubiertas por el corpus existente	Informe sobre movilidad sostenible en ciudades medianas

3.4.4. Proceso completo de curación

El ciclo de vida de la curación sigue un flujo de cinco etapas:

1. **Identificación:** el especialista detecta un documento relevante a través de su actividad profesional (seguimiento de publicaciones, asistencia a eventos, redes institucionales, monitorización legislativa).

2. **Evaluación:** se valora el documento contra los criterios de selección. No todo documento identificado se incorpora al almacén; solo aquellos que superan el umbral de calidad y relevancia.
3. **Análisis IA:** el documento se somete al análisis automatizado (Qwen2.5-72B) que propone una clasificación taxonómica completa.
4. **Validación:** el especialista revisa los metadatos propuestos por la IA, corrigiendo errores y añadiendo matices que solo el conocimiento humano puede aportar (por ejemplo, el grado de alineamiento con ECODES o la asignación a asistentes específicos).
5. **Publicación:** el documento se almacena en S3, se indexa en Qdrant y queda disponible para los asistentes IA.



3.4.5. Impacto en la calidad del almacén

Este modelo de curación híbrido (humano + IA) tiene un impacto directo en la calidad del almacén:

- **Precisión de metadatos:** la supervisión humana corrige errores de la IA (estimación: ~15-20% de los metadatos propuestos requieren ajuste).
- **Coherencia temática:** los especialistas garantizan que el almacén mantiene un foco coherente en descarbonización, evitando la acumulación de documentos tangenciales.
- **Calidad de las transcripciones:** cuando la extracción automática de PDF es deficiente, los especialistas pueden proporcionar transcripciones manuales o revisadas que mejoran sustancialmente la calidad de los embeddings.

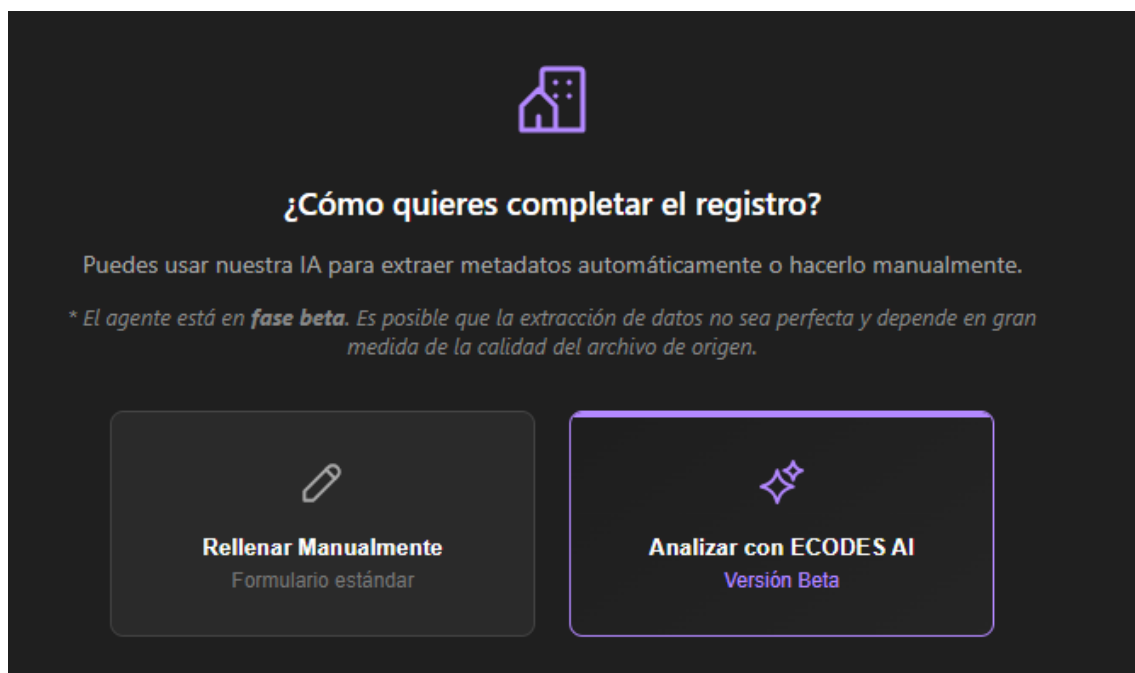
- **Contextualización institucional:** solo un conocedor de ECODES puede determinar si un documento constituye un posicionamiento institucional y cuál es su grado de alineamiento.

3.5. Inteligencia artificial para la gobernanza

3.5.1. Análisis automatizado de documentos

El sistema incorpora un módulo de análisis automatizado que utiliza modelos de lenguaje para extraer metadatos estructurados de cada documento ingresado. Este proceso soporta tres orígenes de contenido:

1. Transcripción directa: Texto pegado por el usuario.
2. Archivo subido: PDF o documento de texto.
3. Archivo desde S3: Descarga automática desde el repositorio cloud.



3.5.2. Modelo de análisis

El análisis utiliza el modelo tipo *Qwen2.5-72B-Instruct* (o equivalente), un modelo de lenguaje de código abierto seleccionado por:

- Gratuidad: sin coste por token (cumpliendo la licencia y asumiendo los costes de ejecución).
- Capacidad multilingüe: excelente rendimiento en español.

- Calidad de clasificación: alta precisión en la extracción de metadatos estructurados.

3.5.3. Validación

Más allá de las instrucciones al modelo, el sistema implementa una validación server-side estricta que descarta metadatos inventados. Esta doble validación asegura que ningún metadato ficticio contamina el almacén de datos, manteniendo la integridad taxonómica.

3.5.4. Carga dinámica de taxonomías

Las listas de valores permitidos se cargan dinámicamente desde la base de datos en cada análisis. Esto permite que la evolución de las taxonomías (nuevas áreas, nuevos tipos de contenido) se refleje inmediatamente en los análisis de IA sin necesidad de modificar el código.

Análisis IA en proceso

✓ **Análisis completado. Revisa los datos antes de aplicar.**

CLASIFICACIÓN SUGERIDA		DATOS CLAVE DETECTADOS
Área:	Transformación Digital	#Gestión de riesgos y resiliencia
Corpus:	corpus normativo	#Planificación y gobernanza cambio climático
Audiencia:	Administración General del Estado	
Geo:	España	
Tipo:	Normas	

► Ver JSON crudo

Cancelar **Aplicar datos**

Ejemplo de carga dinámica de taxonomías

3.6. Gestión de transcripciones

Las transcripciones constituyen una fuente fundamental de contenido para el almacén vectorial. El sistema prioriza las fuentes de texto en el siguiente orden:

- **Transcripción:** almacenada directamente en la base de datos PostgreSQL. Puede ser:
 - Texto transcrito manualmente.
 - Resultado de un proceso de transcripción automática
 - Resumen detallado elaborado por el usuario.
- **PDF original:** Si no hay transcripción disponible, el sistema descarga el PDF desde Amazon S3 y extrae el texto usando la biblioteca PdfReader.

La calidad de la transcripción afecta directamente la calidad de los embeddings y, por tanto, la precisión de la búsqueda semántica.

Por ello, el pipeline de re-indexación (Capítulo 4) incluye un paso de limpieza de texto que corrige artefactos típicos de la extracción PDF.

3.7. Conexión con Amazon S3 (seguridad)

3.7.1. Arquitectura del repositorio

El almacenamiento físico de archivos se delega a Amazon S3 (Simple Storage Service), un servicio de almacenamiento de objetos en la nube con alta disponibilidad y durabilidad:

Parámetro	Valor
Servicio	Amazon S3
Región	`eu-central-1` (Frankfurt, UE)
Autenticación	AWS IAM (Access Key + Secret Key)

3.7.2. Modelo de seguridad

La seguridad del repositorio S3 se implementa en múltiples capas:

- Autenticación IAM.
- ACLs de objetos.
- URLs prefiradas.

- Cifrado en tránsito y reposo.
- Papelera
- Metadatos S3

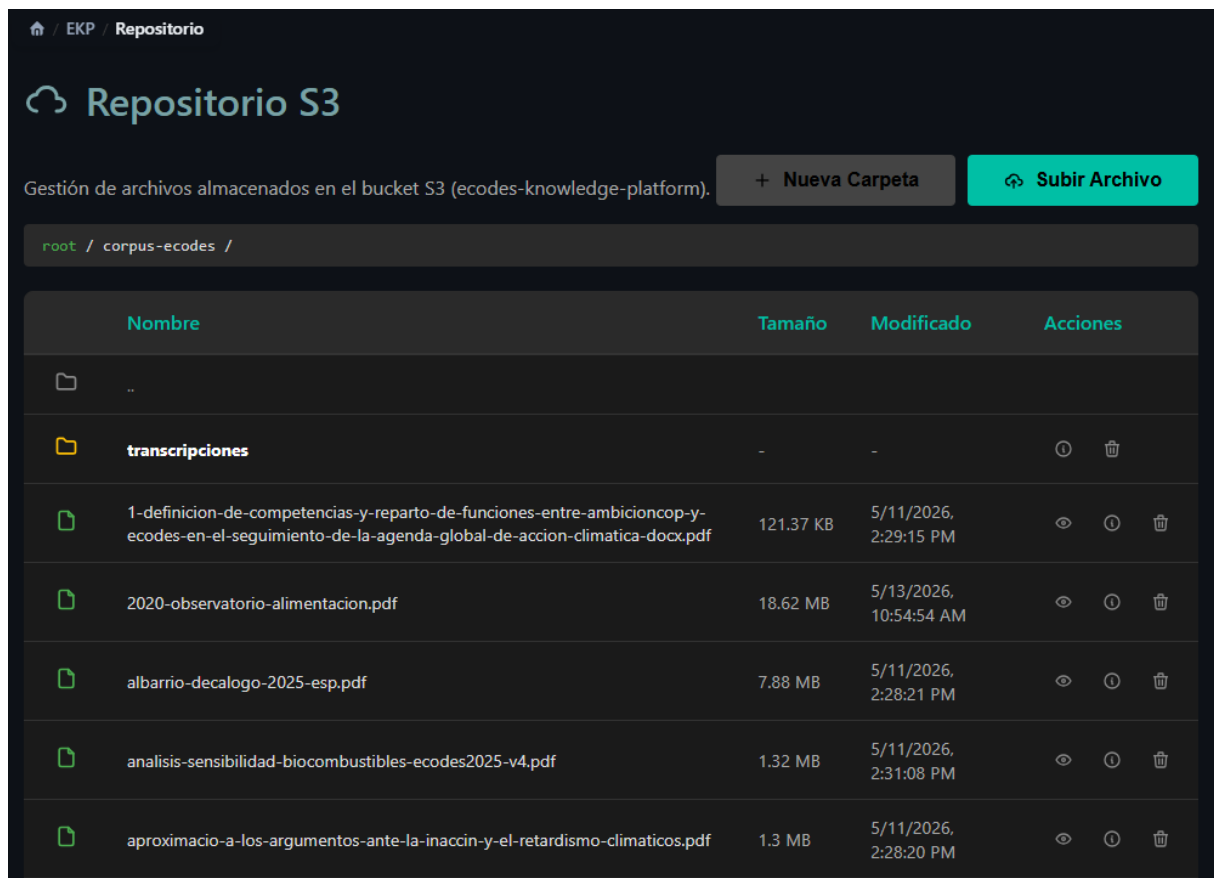


Imagen de la gestión del repositorio en S3 de EKP

3.7.3. Localización de datos (RGPD)

La elección de la región eu-central-1 (Frankfurt) garantiza que todos los datos permanecen dentro de la Unión Europea, cumpliendo con las exigencias del Reglamento General de Protección de Datos (RGPD).

3.8. Permisos de usuarios

3.8.1. Modelo de permisos

El sistema implementa un modelo de acceso basado en Grupos y Niveles de Acceso. Cada usuario pertenece a uno o varios Grupos (por ejemplo, administradores, gestores de IA, gobernantes de datos), y cada Grupo tiene

asociado un conjunto de Niveles de Acceso que definen, de forma granular, las operaciones que sus miembros pueden realizar dentro de la plataforma.

Los Niveles de Acceso funcionan como permisos atómicos: cada uno representa una capacidad concreta del sistema (gestionar usuarios, acceder a la gobernanza, configurar asistentes de IA, visualizar estadísticas, etc.) y se identifica mediante un nombre descriptivo y un código interno normalizado. Este diseño permite componer perfiles de acceso flexibles sin necesidad de crear roles monolíticos: un grupo puede combinar exactamente los niveles que necesita, y un usuario que pertenezca a varios grupos acumula las capacidades de todos ellos.

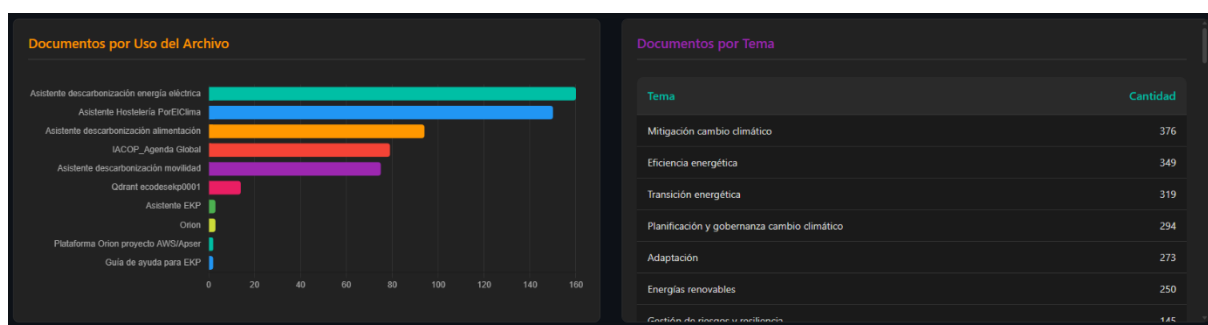
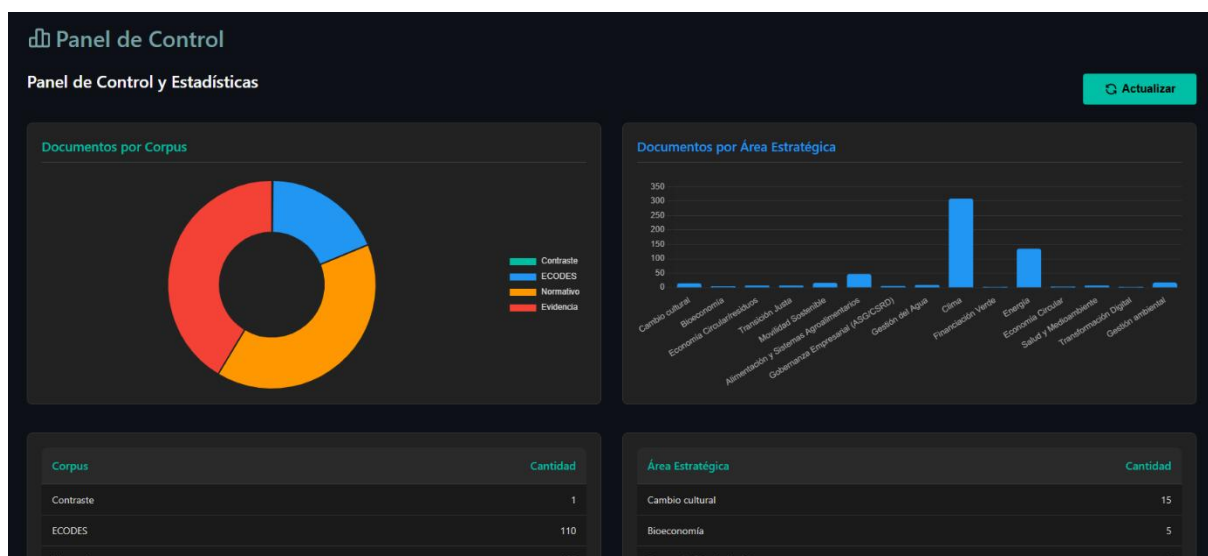
La relación entre Grupos y Niveles de Acceso se gestiona a través de una tabla de asociación muchos-a-muchos, lo que permite modificar los permisos de un grupo completo de usuarios con un único cambio de configuración, sin necesidad de actualizar cada cuenta individualmente. La interfaz de administración proporciona un panel CRUD completo para crear, editar y eliminar tanto Grupos como Niveles de Acceso, así como para asignar o revocar niveles a cada grupo de forma visual e inmediata.

3.9. Panel de control y métricas

3.9.1. Dashboard de control

El panel de control, accesible para administradores y gestores de IA, proporciona una visión en tiempo real del estado del almacén de datos mediante:

- Gráfico de distribución por corpus (Donut chart): Visualiza la proporción de documentos en cada corpus (Evidencia, Normativo, ECODES, Contraste).
- Gráfico de distribución por área estratégica (Bar chart): Muestra la cantidad de documentos por cada área de descarbonización.
- Gráfico de distribución por uso (Horizontal bar): Presenta la asignación de documentos a asistentes IA y otros usos.
- Tabla de distribución por tema: Desglose detallado de documentos por subárea temática.



Imágenes de la interfaz de control de EKP

3.9.2. Datos cuantitativos del almacén (mayo 2026)

Distribución por Corpus:

Corpus	Documentos	Porcentaje
Evidencia	245	41,4%
Normativo	236	39,9%
ECODES	110	18,6%
Contraste	1	0,2%
Total	592	100%

Distribución por Área Estratégica (Top 10):

Área	Documentos
Clima y Energía	309
Transición Energética	135
Alimentación y Sistemas Agroalimentarios	47
Economía Circular/residuos	34
Movilidad sostenible	17
Agua	13
Transformación Digital	12
Inclusión Social	10
Biodiversidad	8
General	7

Distribución por Uso/Asistente (asignaciones a asistentes de descarbonización):

Uso	Documentos asignados
Asistente descarbonización energía eléctrica	158
Asistente descarbonización alimentación	95
Asistente descarbonización movilidad	75

3.10. Configuración de taxonomías

La configuración de las taxonomías se realiza desde una interfaz CRUD completa para cada eje:

- **Parámetros generales:** Versión de taxonomía (actualmente v1.0.0), fecha de última actualización.
- **Taxonomía cerrada (Eje 1 y 2):** Gestión de áreas estratégicas y subáreas temáticas.
- **Taxonomía cerrada (Eje 3):** Gestión de tipos de contenido.
- **Actores/Destinatarios (Eje 5):** Gestión de audiencias objetivo.
- **Taxonomía abierta (Usos):** Gestión de usos del archivo.

3.11. Ejemplos de documentos gobernados

Para ilustrar la riqueza del almacén de datos, se presentan ejemplos reales de documentos gobernados en las tres áreas de descarbonización objetivo:

Alimentación

Asset ID	Título	Corpus	Tipo
EVID_AS_IT_0325_683	Hoja de Ruta de descarbonización de la Industria de alimentación y Bebidas en España	Evidencia	Informe técnico
NORM_AS_NM_0425_698	Estrategia de la UE - De la Granja a la Mesa: sistema alimentario justo, saludable y respetuoso con el medio ambiente	Normativo	Normativa
ECOD_AS_PI_0325_715	Posicionamiento de ECODES sobre alimentación sostenible y huella de carbono	ECODES	Posicionamiento institucional

Energía eléctrica

Documentos que cubren desde directivas europeas de eficiencia energética hasta planes nacionales de energía y clima, pasando por informes técnicos sobre renovables y electrificación.

Asset ID	Título	Corpus	Tipo
NORM_CE_NM_0425_701	Plan Nacional Integrado de Energía y Clima (PNIEC) 2021-2030 - Actualización 2024	Normativo	Normativa
EVID_CE_IT_0325_689	Informe de situación de las energías renovables en España 2024	Evidencia	Informe técnico
ECOD_TE_PI_0325_712	Posicionamiento de ECODES sobre la transición energética justa	ECODES	Posicionamiento institucional

Movilidad

Documentos sobre transporte limpio, electromovilidad, planificación urbana sostenible y regulación de emisiones vehiculares.

Asset ID	Título	Corpus	Tipo
NORM_MS_NM_0425_723	Estrategia de Movilidad Segura, Sostenible y Conectada 2030	Normativo	Normativa
EVID_MS_IT_0325_718	Informe sobre electromovilidad y puntos de recarga en España	Evidencia	Informe técnico
ECOD_MS_PI_0425_735	Posicionamiento de ECODES sobre movilidad sostenible y calidad del aire	ECODES	Posicionamiento institucional

INFRAESTRUCTURA TECNOLÓGICA

4.1. Stack Técnico

El stack técnico —o pila tecnológica— es el conjunto de tecnologías, servicios y herramientas que, combinados, constituyen la infraestructura sobre la que opera la ECODES Knowledge Platform. La selección de cada componente responde a criterios de rendimiento, escalabilidad, coste y alineamiento con los requisitos funcionales del almacén de datos curado y los asistentes de IA.

La arquitectura de la EKP se apoya en un ecosistema de componentes especializados que cubren el ciclo completo del dato: desde su almacenamiento seguro y su representación vectorial para búsqueda semántica, hasta la generación de respuestas fundamentadas por los asistentes de inteligencia artificial. La siguiente tabla resume los componentes principales, la tecnología seleccionada para cada función y su papel dentro del sistema.

Componente	Tecnología	Función
Backend	FastAPI (Python 3.11+)	API REST asíncrona
Base de datos	PostgreSQL + SQLAlchemy	Modelo de datos, taxonomías, usuarios
Almacenamiento	Amazon S3 (eu-central-1)	Repositorio de documentos
Base vectorial	Qdrant Cloud	Embeddings y búsqueda semántica
Embeddings	Cohere Embed Multilingual v3 (Bedrock)	Vectorización de documentos
Re-ranking	Cohere Rerank 3.5 (Bedrock)	Reordenamiento por relevancia
Análisis documental	Qwen2.5-72B-Instruct (HuggingFace)	Extracción de metadatos
Agentes IA	OpenAI GPT-4o / GPT-5.2	Asistentes conversacionales
Extracción PDF	pypdf (PdfReader)	Texto de PDFs
Chunking	LangChain Text Splitters	División semántica
Búsqueda web	Tavily API	Información en tiempo real

4.2. Amazon S3: repositorio de documentos

4.2.1. Arquitectura

El almacenamiento físico de los documentos del corpus requiere un repositorio que garantice durabilidad, disponibilidad, seguridad y cumplimiento normativo europeo. Tras evaluar las alternativas disponibles en el mercado, se seleccionó Amazon S3 (Simple Storage Service) como servicio de almacenamiento de objetos en la nube.

Alternativas evaluadas:

- **Almacenamiento local (servidor propio):** Descartado por los costes de mantenimiento, la dificultad de garantizar alta disponibilidad y la ausencia de redundancia geográfica. Un fallo de hardware comprometería la integridad del corpus completo.
- **Google Cloud Storage:** Alternativa competitiva en prestaciones y precios, pero la integración con el resto del ecosistema AWS (Bedrock para embeddings y re-ranking, IAM para gestión de identidades) habría añadido complejidad innecesaria al requerir autenticación cruzada entre proveedores.
- **Azure Blob Storage:** Opción sólida, especialmente si la infraestructura de IA se apoyase en modelos de OpenAI vía Azure. Sin embargo, la elección de AWS Bedrock como proveedor de modelos de embeddings y re-ranking hacía más coherente mantener todo el ecosistema cloud dentro de AWS.
- **MinIO (autoalojado compatible con S3):** Ofrece compatibilidad total con la API de S3, pero traslada la responsabilidad del mantenimiento, las copias de seguridad y la seguridad física al equipo de ECODES, algo inviable con los recursos disponibles.

Razones de la elección de Amazon S3:

- **Durabilidad:** 99,999999999% (11 nueves) de durabilidad de los objetos almacenados, lo que hace prácticamente imposible la pérdida de un documento.
- **Localización en la UE:** La región eu-central-1 (Frankfurt) garantiza que todos los datos permanecen dentro de la Unión Europea, cumpliendo con las exigencias del Reglamento General de Protección de Datos (RGPD).
- **Coherencia con el ecosistema:** Al utilizar AWS Bedrock para los modelos de embeddings (Cohere Embed v3) y re-ranking (Cohere Rerank 3.5), mantener el almacenamiento en el mismo proveedor simplifica la gestión de credenciales, la facturación y la latencia entre servicios.

- **Seguridad nativa:** Cifrado en reposo (SSE-S3), cifrado en tránsito (TLS), gestión de acceso granular mediante IAM y generación de URLs prefiradas con expiración temporal para descargas controladas.
- **Coste predecible:** Modelo de pago por uso sin compromisos mínimos, con un coste muy reducido para el volumen actual del corpus (~600 documentos).

4.3. Qdrant Cloud: base de datos vectorial

4.3.1. Selección

Una base de datos vectorial es el componente que almacena los embeddings generados a partir del corpus documental y permite buscar por similitud semántica en milisegundos. La elección de este componente es determinante para el rendimiento de todo el sistema RAG, ya que afecta a la velocidad de las búsquedas, la capacidad de filtrado y la escalabilidad del almacén.

Alternativas evaluadas:

- **Pinecone:** Servicio gestionado líder en el mercado, con excelente rendimiento y facilidad de uso. Sin embargo, su modelo de precios basado en número de vectores almacenados resulta elevado a medida que el corpus crece, y su plan gratuito es muy limitado. Además, no ofrece despliegue en la UE, lo que planteaba dudas respecto al cumplimiento del RGPD.
- **Weaviate:** Base de datos vectorial de código abierto con capacidades de búsqueda híbrida nativas. Aunque técnicamente sólida, requiere autoalojamiento para un control total, lo que incrementa la carga de mantenimiento. Su versión cloud estaba en fase temprana en el momento de la evaluación.
- **ChromaDB:** Opción ligera y sencilla, ideal para prototipos y proyectos pequeños. Descartada por carecer de filtrado avanzado por metadatos, soporte limitado para producción y ausencia de un servicio cloud gestionado maduro.
- **pgvector (extensión PostgreSQL):** Permite almacenar vectores directamente en la base de datos relacional existente. Atractivo por su simplicidad, pero su rendimiento en búsquedas a escala y sus capacidades de filtrado no alcanzan el nivel de una base de datos vectorial especializada.

Se seleccionó Qdrant Cloud por la combinación de los siguientes factores:

- **Filtrado avanzado por payload:** Permite aplicar filtros complejos sobre los metadatos de cada vector (corpus, área temática, estado, acceso, usos)

directamente en la consulta, lo cual es imprescindible para que los asistentes solo accedan a documentos autorizados.

- **API Python excelente:** Cliente oficial bien documentado, con soporte asíncrono y tipado completo, lo que aceleró significativamente el desarrollo.
- **Plan gratuito generoso:** Suficiente para el volumen actual del corpus (~600 documentos, ~15.000 vectores), lo que permite operar sin coste de infraestructura vectorial en esta fase.
- **Algoritmo HNSW optimizado:** Implementación de alta calidad del algoritmo Hierarchical Navigable Small World, que garantiza búsquedas rápidas y precisas incluso a escala.
- **Servicio gestionado:** Sin necesidad de aprovisionar servidores, gestionar actualizaciones ni configurar réplicas. Qdrant Cloud se encarga del mantenimiento completo de la infraestructura.

4.4. Modelo de embeddings

Cohere Embed Multilingual v3 (AWS Bedrock)

Un modelo de embeddings es el componente que transforma el texto de los documentos —y las consultas de los usuarios— en representaciones numéricas (vectores) que capturan su significado semántico. La calidad de este modelo determina directamente la precisión de la búsqueda: un buen embedding permite que la consulta "¿cómo reducir el consumo eléctrico en el hogar?" recupere documentos relevantes, aunque ninguno contenga esas palabras exactas.

Tras una primera fase en la que se utilizó un modelo de código abierto alojado en HuggingFace (que presentaba problemas de cold starts y latencia variable), se migró a **Cohere Embed Multilingual v3** desplegado a través de **AWS Bedrock**. Las razones de esta elección fueron:

- **Rendimiento multilingüe nativo:** A diferencia de modelos entrenados predominantemente en inglés, Cohere Embed v3 fue diseñado desde su origen para más de 100 idiomas, con un rendimiento especialmente sólido en español, lo que resulta crítico para un corpus íntegramente en castellano.
- **Disponibilidad instantánea:** Al ejecutarse como servicio gestionado en AWS Bedrock, el modelo está siempre activo y no requiere tiempo de arranque (cold start), un problema recurrente con la infraestructura anterior en HuggingFace donde las primeras consultas podían tardar 15-30 segundos.
- **Coherencia con el ecosistema AWS:** Al compartir proveedor con el almacenamiento S3 y el modelo de re-ranking (Cohere Rerank 3.5), se simplifica la gestión de credenciales IAM y se minimiza la latencia entre servicios.

- **Coste competitivo:** El modelo ofrece un equilibrio óptimo entre calidad y precio, con capacidad de procesamiento por lotes de hasta 96 textos simultáneos, lo que reduce significativamente el tiempo y coste de la indexación masiva del corpus.

4.5. Re-ranking

Cohere Rerank 3.5 (AWS Bedrock)

La búsqueda semántica por similitud vectorial es eficaz para recuperar documentos temáticamente relacionados con una consulta, pero no siempre ordena los resultados de forma óptima: un documento que menciona tangencialmente un tema puede obtener una puntuación de similitud similar a otro que lo aborda en profundidad. El re-ranking resuelve esta limitación aplicando una segunda etapa de evaluación, más precisa y costosa computacionalmente, sobre un conjunto reducido de candidatos preseleccionados.

La EKP implementa una estrategia de **recuperación en dos etapas** (two-stage retrieval): en la primera, Qdrant devuelve un conjunto amplio de candidatos ($N \times 2$ del número final deseado) mediante similitud coseno; en la segunda, Cohere Rerank 3.5 evalúa la relevancia de cada candidato respecto a la consulta original y reordena los resultados con una precisión significativamente mayor.

Se seleccionó **Cohere Rerank 3.5** desplegado en **AWS Bedrock** por las siguientes razones:

- **Precisión cross-lingual:** El modelo evalúa la relevancia semántica profunda entre consulta y documento, no solo la proximidad vectorial, lo que mejora sustancialmente la calidad de los resultados, especialmente en consultas complejas o ambiguas.
- **Capacidad de procesamiento:** Soporta hasta 1.000 documentos por solicitud con pasajes de hasta 2.000 caracteres, lo que permite re-ranear corpus amplios en una sola llamada.
- **Coherencia de ecosistema:** Al igual que el modelo de embeddings, su despliegue en AWS Bedrock permite compartir credenciales IAM, simplificar la facturación y minimizar la latencia con Qdrant y S3.
- **Filtrado inteligente post-rerank:** El pipeline descarta automáticamente resultados con puntuación inferior a 0,001 (irrelevantes), elimina duplicados y aplica un mecanismo de backfill que completa los huecos con los siguientes mejores candidatos, garantizando que el asistente siempre recibe el número de documentos solicitado.

4.7. Búsqueda semántica

La búsqueda semántica es la capacidad central que permite a la EKP recuperar información no por coincidencia literal de palabras clave, sino por proximidad de significado.

La plataforma ofrece tres modos de búsqueda, cada uno con un equilibrio distinto entre velocidad y precisión:

Modos de búsqueda

- **Cosine similarity (rápido):** La consulta se vectoriza y se compara directamente contra los embeddings almacenados en Qdrant mediante similitud coseno. Es el modo más rápido (~50-100ms) y adecuado para consultas sencillas o exploratorias donde la velocidad es prioritaria.
- **Two-stage rerank (preciso):** La búsqueda inicial en Qdrant devuelve un conjunto amplio de candidatos ($N \times 2$) que son reevaluados por Cohere Rerank 3.5 para obtener un ordenamiento más preciso. Este modo añade ~200-400ms de latencia, pero mejora significativamente la relevancia de los primeros resultados, especialmente en consultas complejas o ambiguas.
- **Híbrida (Semantic + Keyword con RRF fusion):** Combina la búsqueda semántica vectorial con una búsqueda por palabras clave tokenizada, fusionando ambos rankings mediante el algoritmo Reciprocal Rank Fusion (RRF). Este modo es especialmente útil cuando la consulta contiene términos técnicos específicos (nombres de normativas, siglas, códigos) que la búsqueda vectorial podría no capturar con exactitud.

Buscador semántico integrado en la EKP

Además de su uso interno por los asistentes de IA, la EKP pone a disposición de los administradores y gestores un buscador semántico con interfaz visual propia, accesible desde el módulo de Almacenamiento de la plataforma.

Esta herramienta permite a los miembros del equipo explorar directamente el contenido del almacén vectorial: formular consultas en lenguaje natural, visualizar los fragmentos recuperados con su puntuación de relevancia, identificar de qué documento y corpus procede cada resultado, y verificar la calidad de la indexación sin necesidad de interactuar con los asistentes de IA.



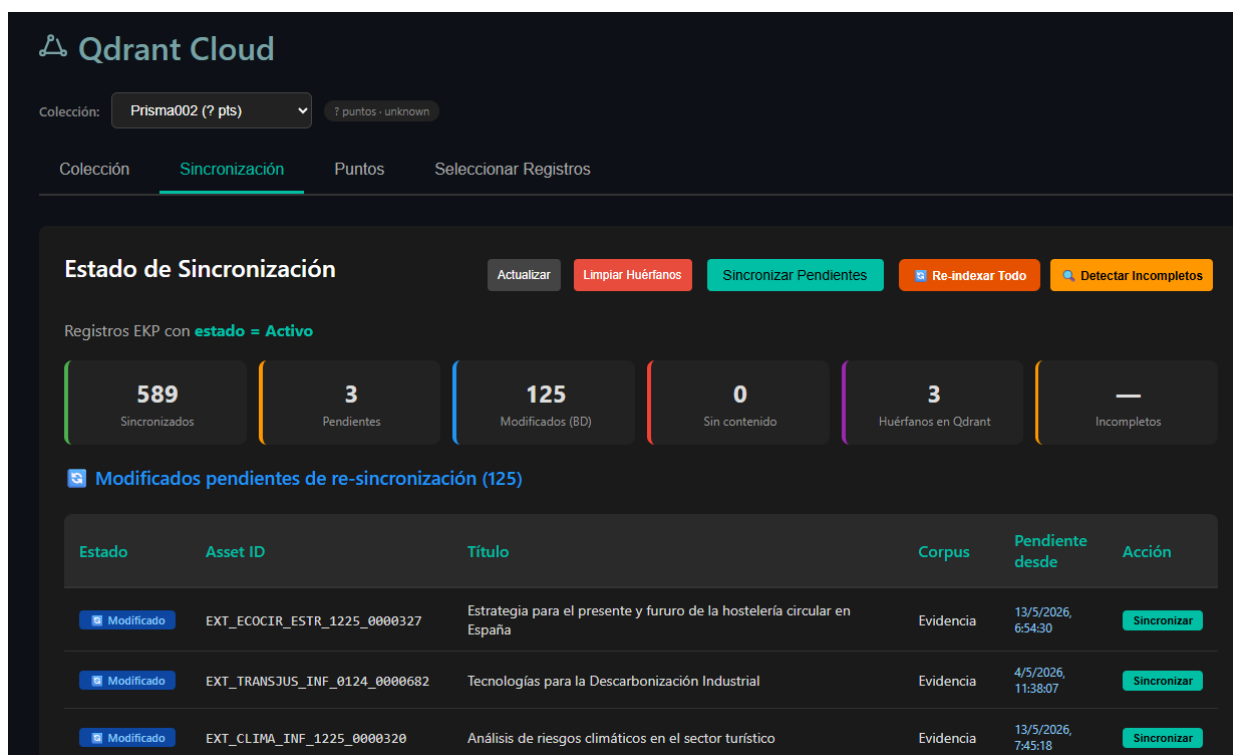
El buscador constituye así una herramienta de diagnóstico y exploración esencial para el Comité de Gobernanza, ya que permite detectar brechas de cobertura, evaluar la calidad de los embeddings y validar que las búsquedas devuelven resultados coherentes antes de que lleguen al usuario final.

4.8. Sincronización EKP ↔ Qdrant

La plataforma gestiona dos representaciones complementarias de cada documento: los metadatos estructurados en PostgreSQL (título, corpus, taxonomía, estado, acceso) y los embeddings vectoriales en Qdrant (fragmentos de texto convertidos en vectores para búsqueda semántica). Mantener ambas representaciones consistentes es un requisito crítico: si un documento se modifica, desactiva o elimina en la gobernanza, su representación vectorial debe reflejar ese cambio de forma inmediata; de lo contrario, los asistentes de IA podrían recuperar información obsoleta, restringida o inexistente.

Para garantizar esta coherencia, la EKP implementa un sistema de sincronización que clasifica cada documento según su estado de integridad entre ambos sistemas y ofrece cuatro modos de operación adaptados a distintas necesidades:

- **Sincronización automática**
- **Indexación manual de pendientes**
- **Re-indexación completa**
- **Limpieza de huérfanos**



Gestión de la sincronización de EKP con Qdrant

4.9. Estrategias de tratamiento documental

La calidad de las respuestas de los asistentes de IA depende directamente de la calidad del contenido indexado en la base vectorial. A lo largo del desarrollo de la EKP se han identificado y adoptado diversas estrategias para maximizar la calidad de los embeddings generados.

4.9.1. Análisis comparativo de estrategias

No todos los documentos del almacén se pueden tratar de la misma forma. Un informe técnico con estructura limpia produce buenos resultados con la extracción automática de texto, pero un documento escaneado, un vídeo de un webinar o un informe de 200 páginas con tablas y gráficos requieren estrategias distintas para que los embeddings generados capturen fielmente su contenido.

A lo largo del desarrollo de la EKP se han probado y documentado cinco estrategias de tratamiento documental, cada una con un perfil diferente de calidad, coste y escalabilidad. La siguiente tabla resume sus características para que el Comité de Gobernanza pueda decidir qué estrategia aplicar en función de la naturaleza y la importancia de cada documento.

Estrategia	Calidad RAG	Coste	Escalabilidad	Cuándo usar
Extracción PDF directa (pypdf)	Media	Bajo	Alta	Documentos con buena estructura PDF
Transcripción manual revisada	Alta	Alto	Baja	Documentos clave, posicionamientos
División previa en capítulos	Alta	Medio	Media	Documentos largos con TOC
Speech-to-text + revisión	Media-Alta	Medio	Media	Vídeos, podcasts, webinars
Resumen largo curado	Alta	Alto	Baja	Documentos complejos multisección

4.9.2. Transcripciones manuales: la estrategia de mayor impacto

La experiencia ha demostrado que las transcripciones manuales o revisadas por especialistas producen embeddings significativamente más precisos que la extracción PDF automática:

- **Eliminación de ruido:** los artefactos PDF (encabezados repetidos, notas al pie descontextualizadas, tablas rotas) desaparecen por completo.
- **Coherencia narrativa:** el texto fluye de forma natural, manteniendo la estructura lógica del documento.
- **Enriquecimiento contextual:** el especialista puede añadir contexto que no está explícito en el PDF (por ejemplo, el marco legislativo en el que se inscribe un informe).

Por ello, el sistema prioriza siempre la transcripción sobre la extracción automática de PDF, y los especialistas de ECODES son alentados a proporcionar transcripciones para los documentos más relevantes.

4.9.3. División en capítulos para documentos largos

Los documentos extensos (>50 páginas) con tabla de contenidos presentan un reto para el chunking automático: los chunks generados pueden mezclar contenido de diferentes secciones, diluyendo la especificidad temática.

Estrategia recomendada:

1. Identificar la estructura de capítulos del documento.
2. Dividir el documento en secciones temáticas antes del chunking.
3. Aplicar chunking independiente a cada sección.
4. Mantener el título del capítulo como contexto en cada chunk.

PATRONES, BARRERAS Y OPORTUNIDADES

5.1. Análisis cuantitativo del almacén de datos

El almacén de datos curado contiene, a fecha de mayo de 2026, 592 documentos gobernados distribuidos en tres corpus complementarios y múltiples áreas estratégicas de descarbonización. Este capítulo analiza la distribución del conocimiento, identifica patrones temáticos, señala barreras documentales y propone oportunidades de mejora.

5.2. Distribución por corpus

Corpus	Documentos	%	Función
Evidencia	245	41,4%	Informes científicos, estudios sectoriales, datos técnicos
Normativo	236	39,9%	Legislación europea, nacional y autonómica
ECODES	110	18,6%	Posicionamientos, guías, materiales institucionales
Contraste	1	0,2%	Fuentes de contraste/debate
Total	592	100%	—

Patrón identificado: equilibrio evidencia-normativo

El almacén presenta un equilibrio notable entre el corpus de Evidencia (41,4%) y el Normativo (39,9%), lo que indica una curación intencionada para que los asistentes de IA puedan fundamentar sus respuestas tanto en evidencia científica como en el marco legal aplicable.

El corpus ECODES (18,6%) aporta la perspectiva institucional y los posicionamientos oficiales.

Oportunidad: fortalecimiento del debate

Incorporar documentos de debate, análisis críticos y posiciones alternativas sobre políticas de descarbonización permitiría a los asistentes ofrecer respuestas más matizadas y equilibradas.

5.3. Distribución por área estratégica

#	Área Estratégica	Documentos	% del total
1	Clima y Energía	309	52,2%
2	Transición Energética	135	22,8%
3	Alimentación y Sistema. Agroalimentarios	47	7,9%
4	Economía Circular/residuos	34	5,7%
5	Movilidad sostenible	17	2,9%
6	Agua	13	2,2%
7	Transformación Digital	12	2,0%
8	Inclusión Social	10	1,7%
9	Biodiversidad	8	1,4%
10	General	7	1,2%

Patrón: concentración en clima y energía

Más de la mitad del almacén (52,2%) se concentra en el área de Clima y Energía, seguida por Transición Energética (22,8%). Esto refleja la prioridad histórica de ECODES en la acción climática y la centralidad de la energía en las estrategias de descarbonización.

Barrera: desproporción en movilidad

Con solo 17 documentos específicos (2,9% del total), el área de Movilidad sostenible es la más infrarrepresentada de las tres áreas objetivo del OE2.1. Esta brecha documental limita la capacidad del asistente de movilidad para ofrecer respuestas completas y matizadas.

Oportunidad: expansión de movilidad y alimentación

La incorporación de documentos adicionales sobre movilidad (regulación de emisiones vehiculares, planes de movilidad urbana, infraestructura de carga eléctrica) y alimentación (cadenas de suministro, impacto ambiental de dietas, políticas agroalimentarias) fortalecería significativamente la base de conocimiento de estos asistentes.

5.4. Distribución por tema

Las temáticas revelan la granularidad del conocimiento curado:

Tema	Documentos
Mitigación del cambio climático	376
Eficiencia energética	349
Transición energética	319
Adaptación al cambio climático	245
Huella de carbono	201
Emisiones de gases de efecto invernadero	194
Descarbonización	187
Energías renovables	164
Transición justa	113
Políticas climáticas	106
Electrificación	97
Comunidades energéticas	72
Alimentación sostenible	60
Movilidad eléctrica	45
Pobreza energética	38

***Nota**:* Un documento puede pertenecer a múltiples temáticas, por lo que las cifras no suman 592.

Patrón: dominancia de mitigación y eficiencia

Las tres subáreas dominantes (Mitigación: 376, Eficiencia energética: 349, Transición energética: 319) confirman que el almacén está sólidamente orientado hacia la descarbonización. La presencia de subáreas como Transición justa (113) y Pobreza energética (38) indica una perspectiva de justicia social en la transición.

5.5. Asignación a asistentes de descarbonización

El sistema de usos del archivo permite asignar documentos a asistentes IA específicos. La distribución actual es:

Asistente	Documentos asignados	Ratio sobre total
Descarbonización Energía Eléctrica	158	26,7%
Descarbonización Alimentación	95	16,0%
Descarbonización Movilidad	75	12,7%

Patrón: cross-corpus por asistente

Cada asistente accede a documentos de múltiples corpus. Por ejemplo, el asistente de energía eléctrica (158 docs.) incluye documentos normativos (directivas EU, leyes nacionales), de evidencia (informes técnicos sobre renovables) y de ECODES (posicionamientos sobre transición energética). Este cross-corpus es fundamental para respuestas completas.

Patrón: enriquecimiento más allá del área específica

El asistente de movilidad tiene 75 documentos asignados, pero solo 17 están en el área específica de "Movilidad sostenible". Los 58 documentos adicionales provienen de áreas relacionadas (Clima, Energía, Economía Circular), evidenciando una curación que reconoce la naturaleza transversal de la movilidad sostenible.

5.6. Análisis sectorial: alimentación

Fortalezas del corpus

- Cobertura de la cadena completa: producción, distribución, consumo, residuos.
- Documentos institucionales propios (posicionamientos ECODES sobre alimentación sostenible).
- Informes técnicos europeos sobre impacto ambiental de sistemas agroalimentarios.

Brechas identificadas

- Escasa documentación sobre dietas regionales y su huella de carbono.
- Pocas guías prácticas para empresas del sector hostelero.
- Ausencia de análisis comparativos de políticas agroalimentarias internacionales.

Oportunidades

- Incorporar datos del IPCC Working Group III sobre sistemas alimentarios.
- Añadir normativa europea sobre cadenas de suministro sostenibles.
- Desarrollar fichas técnicas por tipo de alimento y su impacto en descarbonización.

5.7. Análisis sectorial: energía eléctrica

Fortalezas del corpus

- Corpus más robusto con 158 documentos asignados al asistente especializado.
- Fuerte representación normativa (directivas de eficiencia energética, PNIEC).
- Buena cobertura de subáreas: renovables (164), electrificación (97), comunidades energéticas (72), pobreza energética (38).
- Presencia de posicionamientos institucionales sobre transición energética justa.

Brechas identificadas

- Poca documentación sobre almacenamiento energético y redes inteligentes.
- Documentación sobre autoconsumo limitada a nivel normativo.
- Falta de estudios de caso de implementación exitosa a nivel municipal.

Oportunidades

- Incorporar informes de la Agencia Internacional de Energía (IEA).
- Añadir documentos sobre las nuevas tecnologías de almacenamiento (baterías, hidrógeno verde).
- Enriquecer con estudios de caso de comunidades energéticas existentes.

5.8. Análisis sectorial: movilidad

Fortalezas del corpus

- 75 documentos asignados al asistente, con visión transversal.
- Inclusión de normativa sobre emisiones vehiculares y zonas de bajas emisiones.
- Documentos sobre electromovilidad y movilidad activa.

Brechas identificadas

- Área más infrarrepresentada con solo 17 documentos específicos (2,9%).
- Escasa documentación sobre:
 - Logística de última milla sostenible.
 - Impacto del teletrabajo en emisiones de transporte.
 - Planes de movilidad urbana sostenible (PMUS) a nivel municipal.
 - Transporte público eléctrico.
 - Infraestructura de carga para vehículos eléctricos.

Oportunidades

- Prioridad alta: Expandir el corpus de movilidad como área estratégica.
- Incorporar documentos de la Estrategia de Movilidad Sostenible de España.
- Añadir informes sobre la transformación del transporte pesado.
- Incluir casos de éxito de ciudades con movilidad sostenible avanzada.

5.9. Síntesis: mapa de madurez del almacén

Sector	Docs. específicos	Docs. asistente	Madurez	Prioridad de expansión
Energía	135 + 309	158	● Alta	Baja (mantener)
Alimentación	47	95	● Media	Media
Movilidad	17	75	● Baja	Alta

5.10. Patrones transversales

Patrón 1: complementariedad multi-corpus

Los tres asistentes acceden a documentos de los tres corpus principales, lo que permite respuestas que integran evidencia científica, marco normativo y perspectiva institucional de ECODES.

Patrón 2: gobernanza como filtro de calidad

El proceso de gobernanza con clasificación taxonómica de 5 ejes asegura que cada documento está correctamente categorizado y es recuperable por el sistema de búsqueda semántica. Esto es especialmente relevante para documentos transversales que afectan a múltiples sectores.

Patrón 3: efecto multiplicador del cross-tagging

La capacidad de asignar un documento a múltiples usos (asistentes) y subáreas temáticas multiplica la utilidad del corpus. Un documento sobre "huella de carbono de la cadena alimentaria" puede servir simultáneamente a los asistentes de alimentación, energía y movilidad.

ARQUITECTURA IA ESPECIALIZADA

6.1. Visión general de la arquitectura

La prueba de concepto de asistentes IA especializados en descarbonización (R.2.1.3) se materializa en una arquitectura multi-capa que integra: recuperación semántica multi-corpus, guardrails de seguridad, alineamiento narrativo institucional y citación trazable.



6.2. Pipeline PRISMA: orquestador multi-corpus

6.2.1. Concepto

El pipeline PRISMA (acrónimo de Procesamiento de Recuperación Inteligente con Síntesis Multi-corpus Alineada) es el motor central de los asistentes de IA. Su función es orquestar la recuperación de información desde múltiples corpus de forma transparente, generando respuestas fundamentadas y citadas.

6.2.2. Fases del pipeline

Fase 1: Query Rewrite

El pipeline reformula la consulta del usuario para optimizar la búsqueda semántica:

Fase 2: autodetección Geográfica

El sistema analiza la consulta para detectar referencias geográficas y aplicar filtros automáticos

Fase 3: retrieval Multi-corpus

La búsqueda se ejecuta en paralelo contra múltiples corpus, cada uno con sus propios filtros.

Fase 4: re-ranking global

Todos los resultados de los diferentes corpus se re-rankeea para obtener una clasificación global de relevancia.

Fase 5: generación con citación

El LLM genera la respuesta final con instrucciones estrictas de citación:

6.2.3. Filtros por Agente

Cada asistente tiene filtros de seguridad que restringen qué documentos puede consultar:

Filtro	Descripción	Ejemplo
`acceso`	Nivel de acceso permitido	`["publico", "interno"]` y privado
`estado`	Solo documentos activos	`["activo"]`
`usos_archivo`	Solo documentos asignados	`["Asistente desc. alimentación"]`

Estos filtros se aplican como condiciones obligatorias en cada consulta a Qdrant, garantizando que el asistente solo accede a documentos autorizados.

6.3. Guardrails: sistema de seguridad multi-Capa

Los asistentes de IA de la EKP incorporan un sistema de protección en tres capas que opera antes de que la consulta llegue al pipeline PRISMA. La primera capa

(Filtro de Contenido) bloquea mensajes ofensivos, intentos de inyección de prompts y patrones de manipulación mediante reglas deterministas, sin consumir tokens ni añadir latencia.

La segunda capa (**Evaluación de Relevancia**) verifica que la consulta pertenezca al ámbito temático del asistente utilizando la similitud coseno entre el embedding de la consulta y un conjunto de anclas temáticas predefinidas; al reutilizar el vector que ya se genera para la búsqueda RAG, esta evaluación no requiere llamadas LLM adicionales y es completamente determinista.

La tercera capa (**Auto-pass**) detecta saludos, despedidas y preguntas sobre el propio asistente ("¿quién eres?", "¿qué puedes hacer?") y las deja pasar directamente sin activar la búsqueda en el corpus. Este diseño multicapa garantiza que solo las consultas legítimas y relevantes consuman recursos del sistema, minimizando costes y protegiendo la integridad del almacén de conocimiento.

La mayoría de las plataformas de IA utilizan un modelo de lenguaje adicional para determinar si una consulta es pertinente, lo que supone un coste extra en tokens, latencia y complejidad.

La EKP adopta un enfoque distinto: aprovecha el embedding que ya se genera para la búsqueda semántica y lo compara contra un conjunto de anclas temáticas calibradas para cada asistente. Si la similitud coseno supera un umbral definido empíricamente (calibrado con consultas reales), la consulta se considera relevante; en caso contrario, se rechaza con un mensaje explicativo. Este diseño ofrece ventajas claras frente a los guardrails basados en LLM.

6.4. Auditor de alineamiento narrativo (PRISMA Verifier)

El Auditor de Alineamiento es un módulo que verifica y corrige las respuestas de los asistentes IA para que reflejen fielmente los posicionamientos oficiales y narrativas institucionales de ECODES.

AUDITOR DE ALINEAMIENTO NARRATIVO

Verificación de coherencia institucional



El auditor consulta dos fuentes de referencia: los posicionamientos asociados a cada temática (por ejemplo, la postura de ECODES sobre transición energética justa o alimentación sostenible) y los documentos de posicionamiento oficial almacenados en el corpus ECODES.

La imagen muestra una captura de pantalla de la interfaz de usuario del auditor de posicionamiento. En la parte superior, se indica 'Auditoría de posicionamiento' con un indicador de estado 'ALTO' y '3 cambios'. Debajo, se listan los 'CAMBIOS REALIZADOS' con tres categorías: 'Gap cubierto', 'Desviación corregida' y 'Énfasis ajustado', cada una con una descripción de los cambios y un enlace a 'Posicionamiento de ECODES sobre combustibles de origen biológico o biogénico'. En la sección 'POSICIONAMIENTOS APLICADOS', se muestran dos documentos consultados: 'Posicionamiento de ECODES sobre combustibles de origen biológico o biogénico' y 'La eficiencia energética debe primar en la transición ecológica'. Al final, se menciona 'Subáreas evaluadas: nutrición, hábitos alimentarios, sostenibilidad alimentaria'.

El proceso se ejecuta en tres fases.

Primero, el sistema identifica las subáreas temáticas relevantes en la respuesta generada y recopila los posicionamientos institucionales correspondientes.

A continuación, un modelo de lenguaje compara la respuesta con dichos posicionamientos y produce un diagnóstico estructurado que incluye: el grado de alineamiento (alto, medio o bajo), una versión corregida de la respuesta si es necesario, la lista detallada de cambios realizados con su justificación, y los posicionamientos que se han consultado.

Las correcciones aplicadas pueden ser de tres tipos: incorporación de un posicionamiento de ECODES que faltaba en la respuesta original, corrección de una desviación respecto al enfoque institucional, o ajuste del énfasis o el tono para que resulte coherente con la narrativa de la organización.

El auditor identifica las subáreas temáticas presentes en el texto, recupera los posicionamientos institucionales correspondientes y genera un informe estructurado que indica el grado de alineamiento, las desviaciones detectadas y las correcciones sugeridas. Esta capacidad convierte al auditor en una herramienta de control de calidad narrativa que trasciende el ámbito de los asistentes de IA: cualquier contenido que ECODES produzca o reciba puede ser contrastado con su marco de posicionamientos, asegurando coherencia institucional en todas las comunicaciones de la organización.

Además de la auditoría automática que se ejecuta tras cada respuesta del asistente, la EKP permite realizar auditorías manuales de textos o documentos externos contra los posicionamientos institucionales de ECODES. Esta funcionalidad permite al equipo pegar un texto cualquiera —un borrador de comunicado, un artículo de prensa, un documento de un tercero, una propuesta de proyecto— y someterlo al mismo proceso de verificación de alineamiento que se aplica a las respuestas de los asistentes.

6.5. Asistentes especializados en descarbonización

La prueba de concepto del OE2.1 se materializa en tres asistentes conversacionales, cada uno especializado en un sector prioritario para la descarbonización. Los tres comparten la misma arquitectura técnica —pipeline PRISMA, guardrails, auditor de alineamiento— pero se diferencian en el subconjunto de documentos al que acceden, las subáreas temáticas que cubren y las capacidades específicas que ofrecen al usuario.

Cada asistente opera sobre una selección curada de documentos procedentes de los tres corpus (ECODES, Normativo y Evidencia), lo que le permite fundamentar sus respuestas combinando la perspectiva institucional de ECODES, el marco legislativo aplicable y la evidencia científica disponible. Esta configuración multi-corpus es la que permite que un asistente, ante una misma consulta, pueda citar simultáneamente una directiva europea, un informe técnico y el posicionamiento oficial de ECODES sobre la materia.

Las siguientes fichas resumen la configuración de cada asistente, incluyendo el volumen de documentos asignados, los corpus de origen, las subáreas temáticas cubiertas y sus capacidades principales.

6.5.1. Asistente de descarbonización: energía eléctrica

Parámetro	Valor
Documentos asignados	158
Corpus	ECODES + Normativo + Evidencia
Subáreas clave	Transición energética, Eficiencia energética, Renovables, Electrificación, Comunidades energéticas, Pobreza energética
Capacidades	Consulta normativa EU/ES, análisis de evidencia científica, posicionamiento ECODES sobre transición justa

6.5.2. Asistente de descarbonización: alimentación

Parámetro	Valor
Documentos asignados	95
Corpus	ECODES + Normativo + Evidencia
Subáreas clave	Alimentación sostenible, Huella de carbono, Economía circular, Sistemas agroalimentarios
Capacidades	Impacto ambiental de sistemas alimentarios, guías de alimentación sostenible, normativa sanitaria y ambiental

6.5.3. Asistente de descarbonización: movilidad

Parámetro	Valor
Documentos asignados	75
Corpus	ECODES + Normativo + Evidencia
Subáreas clave	Movilidad eléctrica, Transporte sostenible, Calidad del aire, Planificación urbana
Capacidades	Regulación de emisiones, electromovilidad, planificación de movilidad urbana, transporte público

6.5.4. Modo de conocimiento: solo corpus ECODES

Este modo es especialmente relevante para consultas donde la trazabilidad de la fuente es imprescindible: informes oficiales, respuestas a terceros o análisis que requieren que cada afirmación esté respaldada por un documento identificable del almacén.

Alternativamente, el usuario puede activar el uso del conocimiento del modelo como complemento, lo que permite al asistente enriquecer la respuesta con contexto general cuando el corpus no ofrece cobertura suficiente. Esta opción se presenta de forma transparente en la interfaz del chat, permitiendo al usuario decidir en cada momento el nivel de control que desea ejercer sobre las fuentes de información.

La plataforma permite configurar cada conversación para que el asistente utilice exclusivamente el conocimiento almacenado en el corpus curado de ECODES, desactivando por completo el conocimiento interno del modelo de lenguaje. En este modo, si la información solicitada no se encuentra en los documentos del almacén, el asistente lo indica explícitamente en lugar de recurrir a su conocimiento general, que podría estar desactualizado, contener imprecisiones o no reflejar la perspectiva institucional de ECODES.

6.5.5. Personalización de la búsqueda por el usuario

Los asistentes de la EKP no limitan al usuario a formular preguntas y esperar una respuesta genérica sobre todo el corpus. La plataforma ofrece dos mecanismos de personalización que permiten al usuario acotar, dirigir y refinar la búsqueda antes de que el pipeline la ejecute.

Filtrado por metadatos. El usuario puede aplicar filtros avanzados sobre los metadatos del almacén para restringir el ámbito de la búsqueda. Estos filtros se aplican directamente sobre los índices de payload de Qdrant, lo que permite acotar la búsqueda vectorial sin penalización de rendimiento. Un técnico municipal de Aragón, por ejemplo, podría filtrar por su comunidad autónoma para obtener respuestas centradas en la normativa y los recursos de su territorio.

Selección manual de documentos fuente. El usuario puede abrir un explorador del almacén documental y seleccionar manualmente los documentos que desea que el asistente utilice como fuente para su respuesta. Cuando el usuario confirma su selección, el asistente limita su búsqueda semántica exclusivamente a los documentos seleccionados, garantizando que la respuesta se fundamenta en las fuentes elegidas por el usuario y no en el corpus general.

La combinación de ambos mecanismos otorga al usuario un control granular sobre el proceso de búsqueda, convirtiendo al asistente en una herramienta que no solo responde, sino que responde sobre lo que el usuario le pide que busque.

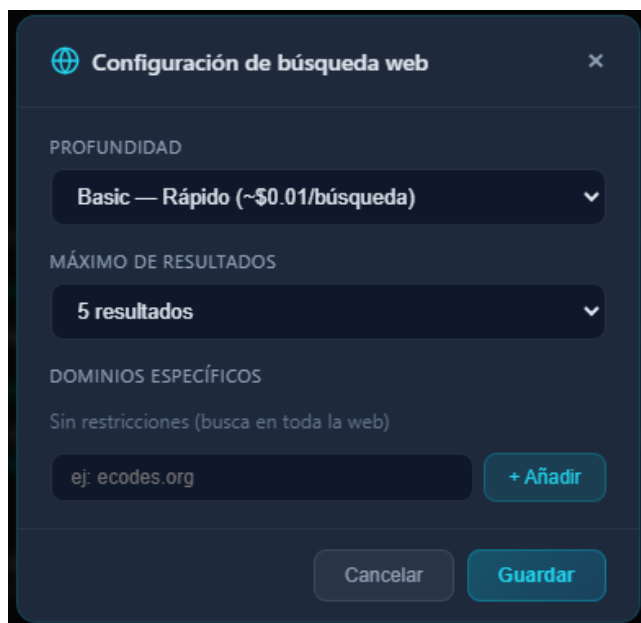
6.6. Búsqueda web integrada

El almacén de datos curado es, por definición, un corpus estático: contiene documentos seleccionados, clasificados y validados en un momento determinado. Sin embargo, la descarbonización es un ámbito en constante evolución donde la normativa se actualiza, se publican nuevos informes y surgen datos que el corpus aún no recoge.

Para cubrir esta brecha temporal, los asistentes de la EKP incorporan la capacidad de realizar búsquedas web en tiempo real como complemento al conocimiento curado.

Se seleccionó **Tavily Search API** como servicio de búsqueda web tras evaluar las alternativas disponibles:

- **Google Custom Search API:** Ofrece resultados de alta calidad, pero su modelo de precios por consulta es elevado para un uso frecuente y los resultados incluyen abundante contenido publicitario que requiere filtrado adicional.
- **Bing Web Search API (Azure):** Alternativa competitiva, pero habría añadido un proveedor cloud adicional (Azure) al ecosistema ya basado en AWS, complicando la gestión de credenciales y facturación.
- **SerpAPI / Serper:** Servicios de scraping de resultados de búsqueda que, si bien funcionan, operan en una zona gris legal respecto a los términos de servicio de los buscadores.
- **Tavily Search API:** Diseñada específicamente para aplicaciones de IA y sistemas RAG. Devuelve resultados limpios, estructurados y optimizados para ser consumidos por modelos de lenguaje, sin publicidad ni ruido. Su coste es competitivo y su integración es directa.



Configuración de opciones de búsqueda web

6.8. Servidores MCP: extensión del conocimiento

6.8.1. Concepto: Model Context Protocol

El protocolo MCP (Model Context Protocol) permite a los asistentes de IA acceder a fuentes de información externas mediante servidores estandarizados. La EKP implementa un sistema de servidores MCP que extiende las capacidades de los asistentes más allá del corpus curado, permitiendo consultas en tiempo real a bases de datos externas, APIs públicas y servicios web.

6.8.2. Servidores MCP Implementados

La plataforma cuenta con cuatro servidores MCP operativos:

Servidor	Función
BOE	Consulta del Boletín Oficial del Estado: búsqueda de normativa, descarga de sumarios, filtrado por sección y departamento
Gobernanza	Consulta de documentos gobernados, metadatos y taxonomías desde agentes
PorEIClima	Acceso a datos del programa PorEIClima de ECODES
Hostelería	Información sobre hostelería sostenible y buenas prácticas

Ejemplo: Servidor MCP del BOE

El servidor del Boletín Oficial del Estado permite a los asistentes buscar normativa publicada recientemente. Soporta:

- Búsqueda por palabras clave en los últimos N días.
- Consulta por fecha específica con filtrado por sección.
- Listado de departamentos que publicaron un día concreto.
- Cache en base de datos local para evitar consultas repetidas a la API.

6.8.3. Oportunidades de expansión MCP para descarbonización

Se han identificado seis oportunidades de expansión mediante nuevos servidores MCP que fortalecerían significativamente las capacidades de los asistentes especializados:

Servidor MCP propuesto	Fuente de datos	Asistente beneficiado	Utilidad
Recetas sostenibles	Open Food Facts API / BD interna	Alimentación	Huella de carbono por receta, alternativas vegetales, impacto ambiental por ingrediente
Consumos energéticos	REE (Red Eléctrica de España), IDAE	Energía eléctrica	Generación renovable en tiempo real, mix eléctrico actual, factor de emisión horario
Movilidad urbana	EMT, DGT, APIs de transporte	Movilidad	Comparativa de emisiones entre modos de transporte, rutas sostenibles, estado del tráfico
Precios de energía	OMIE, CNMC	Energía eléctrica	Precio horario de electricidad, tarifas reguladas, coste de autoconsumo
Indicadores climáticos	AEMET, Copernicus	Todos	Datos meteorológicos, anomalías térmicas, calidad del aire por zona
Normativa EU	EUR-Lex API	Todos	Directivas y reglamentos europeos actualizados sobre descarbonización

Estas expansiones permitirían a los asistentes complementar el conocimiento estático del corpus curado con datos dinámicos y en tiempo real, enriqueciendo significativamente las respuestas.

6.9. Auditoría de tokens y estimación de emisiones

6.9.1. Trazabilidad del consumo de tokens

Cada interacción con un modelo de lenguaje tiene un coste medible en tokens: unidades de texto que el modelo procesa como entrada (la consulta del usuario, el contexto recuperado, el system prompt) y genera como salida (la respuesta). Conocer con precisión cuántos tokens consume cada mensaje es fundamental para una organización como ECODES, que necesita controlar los costes operativos de sus asistentes de IA y, al mismo tiempo, estimar la huella ambiental asociada a su uso.

MÉTRICAS			
TOKENS INPUT ~1210	TOKENS OUTPUT ~1648	TIEMPO 30.46s	TTFT 5286ms
MODELO GPT-5.4	PROVIDER openai	TEMPERATURA 0.7	MAX TOKENS 2048
RAG COLECCIÓN Prisma002	RAG MÉTODO tool_call_search	RAG RESULTADOS 8	

La EKP registra esta información a nivel de cada mensaje individual dentro de cada conversación, almacenando no solo los contadores de tokens sino también los precios unitarios del modelo utilizado y los metadatos completos de la llamada al proveedor. La siguiente tabla detalla los campos que se registran para cada interacción.

6.9.2. Catálogo de modelos con precios

No todas las tareas que realiza un asistente de IA requieren la misma potencia de cálculo. Reformular una consulta de búsqueda, generar el contexto de un chunk o clasificar un mensaje como saludo son operaciones que un modelo ligero resuelve con la misma eficacia que uno de gran escala, pero consumiendo una fracción de los tokens y, por tanto, una fracción de la energía. Utilizar un modelo de 200.000 millones de parámetros para tareas auxiliares que uno de 8.000 millones puede resolver es un desperdicio de recursos económicos y energéticos que una organización comprometida con la descarbonización no puede permitirse.

Por este motivo, la EKP mantiene un catálogo centralizado de modelos de lenguaje que recoge las especificaciones técnicas y los precios de cada uno, permitiendo al equipo tomar decisiones informadas sobre qué modelo asignar a cada función del sistema. La siguiente tabla resume los campos que se registran para cada modelo.

Este catálogo permite, por ejemplo, constatar que asignar GPT-4o-mini a tareas auxiliares (reescritura de consultas, contextual retrieval, auditoría de alineamiento) en lugar del modelo principal reduce el coste por token entre un 80% y un 95%, con un impacto proporcional en el consumo energético asociado. La arquitectura de la EKP aplica este principio de forma sistemática: los modelos más potentes se reservan exclusivamente para la generación de la respuesta final al usuario, mientras que el resto de las operaciones del pipeline se delegan en modelos más eficientes.

6.9.3. Estimación de emisiones de CO₂

La trazabilidad completa del consumo de tokens que implementa la EKP no solo permite controlar costes económicos, sino que abre la puerta a una dimensión poco habitual en las plataformas de IA: la estimación de la huella de carbono generada por cada interacción con los asistentes.

Cada llamada a un modelo de lenguaje implica un consumo energético real en los centros de datos del proveedor: las GPUs procesan millones de operaciones matemáticas para generar cada respuesta, y esa electricidad lleva asociada una huella de carbono que varía en función de la región del datacenter, su eficiencia energética (PUE) y el mix eléctrico del país donde se ubica. Aunque los proveedores de servicios cloud no publican factores de emisión exactos por token, es posible realizar estimaciones razonables a partir de datos públicos sobre consumo energético de GPUs, eficiencia de los centros de datos y factores de emisión de las redes eléctricas de cada región. La siguiente tabla recoge las estimaciones utilizadas como referencia.

Es importante señalar que estos factores son **aproximaciones** basadas en la información disponible actualmente, no mediciones directas. Los valores reales pueden variar significativamente en función de la carga del datacenter, la optimización del hardware, el uso de energías renovables por parte del proveedor y la arquitectura específica del modelo utilizado.



Ejemplo de estimación de emisiones de una conversación

No obstante, el valor de este ejercicio no reside en la precisión absoluta de las cifras, sino en la **infraestructura de medición** que lo hace posible. Al registrar cada token consumido, cada modelo utilizado y cada proveedor invocado, la EKP sienta las bases para que, a medida que los proveedores publiquen factores de emisión más precisos o surjan estándares de medición (como los que impulsan iniciativas como la AI Carbon Footprint de la Unión Europea), ECODES pueda calcular con rigor creciente la huella de carbono de sus asistentes de IA y tomar decisiones informadas para reducirla: seleccionando modelos más eficientes, priorizando proveedores con mayor porcentaje de renovables o ajustando la arquitectura del pipeline para minimizar las llamadas innecesarias.

Para una organización cuya misión es la descarbonización, medir — aunque sea de forma aproximada— el impacto ambiental de sus propias herramientas de IA no es un ejercicio menor: es una cuestión de coherencia.

6.9.4. Optimización del consumo

Cada token consumido por un asistente de IA representa un coste económico y un impacto ambiental. Si bien individualmente son magnitudes pequeñas, a escala — centenares de conversaciones diarias, miles de documentos indexados— se convierten en un factor operativo relevante.

La arquitectura de la EKP ha sido diseñada con la eficiencia como criterio transversal, incorporando múltiples estrategias que, combinadas, reducen drásticamente el consumo de tokens sin sacrificar la calidad de las respuestas.

Guardrails sin consumo de tokens. La mayoría de las plataformas de IA utilizan un modelo de lenguaje para evaluar si una consulta es segura o relevante, lo que supone entre 500 y 1.000 tokens por cada evaluación. La EKP sustituye este enfoque por guardrails basados en embeddings y reglas deterministas que operan a coste cero: ni un solo token adicional, ni una sola milésima de segundo de latencia.

Contextual Retrieval adaptativo. Para enriquecer cada fragmento de documento con contexto antes de indexarlo, los sistemas convencionales invocan un LLM por cada chunk (~100 tokens por fragmento). La EKP utiliza un prefijo estático calculado una única vez por documento, eliminando por completo las llamadas al modelo de lenguaje en la fase de indexación.

Caché de embeddings en memoria. Un sistema LRU (Least Recently Used) de 256 entradas almacena los embeddings de las consultas más frecuentes, evitando recalcularlos en cada interacción. Las preguntas habituales —que en la práctica representan un porcentaje significativo del tráfico— se resuelven sin ninguna llamada al modelo de embeddings.

Selección de modelo por tarea. No todas las operaciones del pipeline necesitan el mismo modelo. La EKP asigna modelos ligeros y eficientes (como GPT-4o-mini) a las tareas auxiliares —reescritura de consultas, generación de contexto, evaluación de alineamiento— y reserva los modelos de mayor capacidad exclusivamente para la generación de la respuesta final al usuario. Esta estrategia reduce el coste de las operaciones internas entre un 80% y un 95%.

Configuración granular por asistente. La plataforma permite ajustar parámetros de consumo a nivel de cada asistente: activar o desactivar el uso del conocimiento curado (para consultas que no requieren RAG), limitar la longitud de la respuesta, seleccionar el modelo de lenguaje más adecuado para cada caso de uso, o forzar el modo de búsqueda web cuando el corpus no es necesario. Estas opciones de configuración permiten al equipo de gestión calibrar el equilibrio entre profundidad de la respuesta y coste operativo según las necesidades de cada asistente.

El resultado combinado de estas estrategias es que la EKP puede operar con costes operativos significativamente inferiores a los de una arquitectura convencional, donde cada capa de seguridad, cada indexación y cada tarea auxiliar consumiría tokens de modelos de gran escala.

6.10. Sistema de roles por asistente

6.10.1. Concepto

El sistema de roles permite modificar el comportamiento de un asistente de IA sin cambiar su configuración base. Cada rol inyecta instrucciones adicionales en el system prompt que orientan el tono, enfoque y nivel de detalle de las respuestas.

6.10.2. Implementación

Los roles se gestionan como entidades independientes, y cada agente del marketplace puede restringir qué roles están disponibles, y cada sesión de chat tiene un rol activo asociado.

6.10.3. Roles propuestos para asistentes de descarbonización

Se proponen nueve roles especializados, tres por cada asistente de descarbonización:

Asistente de energía eléctrica

Rol	Categoría	Enfoque
Analista de política energética	Análisis	Foco en regulación, directivas EU, PNIEC, legislación nacional. Respuestas con referencias normativas precisas.
Asesor de comunidades energéticas	Técnico	Orientación práctica para crear y gestionar comunidades energéticas ciudadanas. Lenguaje accesible.
Técnico de eficiencia energética	Técnico	Auditorías energéticas, certificación de edificios, rehabilitación. Datos cuantitativos y estándares.

Asistente de alimentación

Rol	Categoría	Enfoque
Nutricionista sostenible	Técnico	Dietas de bajo impacto ambiental, menús escolares sostenibles, alternativas vegetales con huella reducida.
Asesor de cadena de suministro	Estrategia	Producción sostenible, logística de proximidad, packaging biodegradable, reducción de desperdicio.
Comunicador ambiental	Comunicación	Narrativas para sensibilizar sobre alimentación y clima. Lenguaje divulgativo, datos impactantes, call-to-action.

Asistente de movilidad

Rol	Categoría	Enfoque
Planificador de movilidad urbana	Estrategia	Planes de movilidad urbana sostenible (PMUS), zonas de bajas emisiones, infraestructura ciclista.
Asesor de electromovilidad	Técnico	Puntos de carga, flotas eléctricas, incentivos fiscales, autonomía y tecnologías de baterías.
Analista de transporte de mercancías	Análisis	Logística de última milla sostenible, hidrógeno verde para transporte pesado, intermodalidad.

El conjunto de estos mecanismos de trazabilidad convierte a la EKP en una plataforma donde ninguna respuesta es opaca: cada afirmación tiene su fuente, cada fuente tiene su puntuación de relevancia, cada paso del pipeline tiene su registro temporal y cada conversación tiene su coste y su huella. Esta transparencia no es solo una garantía técnica, sino un requisito ético para una organización que promueve la rendición de cuentas y la responsabilidad ambiental.

Cada conversación se cierra con un resumen cuantitativo que incluye los tokens de entrada y salida consumidos, el total de tokens utilizados, el coste económico de la conversación en euros, el número de intercambios realizados, la estimación de emisiones de CO₂ equivalente en miligramos y su equivalencia en kilómetros recorridos en coche. Este resumen materializa el compromiso de la EKP con la transparencia operativa: cada interacción con los asistentes tiene un coste medible y un impacto ambiental estimable, y el usuario —o el administrador— puede conocerlo.

6.11. Trazabilidad de fuentes y sistema de traza

Una de las funcionalidades más relevantes de la EKP es su sistema de trazabilidad completa, que permite reconstruir con total transparencia el camino que sigue cada consulta desde que el usuario la formula hasta que recibe la respuesta. Esta trazabilidad opera en dos niveles complementarios: la citación de fuentes en la respuesta y el registro detallado de cada paso del pipeline en el sistema de traza.

6.11.1. Citación de fuentes en la respuesta

Cada respuesta generada por los asistentes incluye al final un bloque de fuentes documentales que identifica todos los documentos del corpus que han sido consultados y citados. El usuario puede ver, para cada fuente, el título del documento, el corpus de procedencia (ECODES, Normativo o Evidencia) y el número de referencia con el que aparece citada en el cuerpo de la respuesta. Esta citación no es decorativa: cada afirmación del asistente que se fundamenta en el corpus lleva asociada su referencia numérica [1], [2], [3], lo que permite al usuario verificar de forma inmediata el origen de cada dato o recomendación. Este sistema de citación reproduce la lógica de las referencias bibliográficas en publicaciones científicas, adaptada a un entorno conversacional.

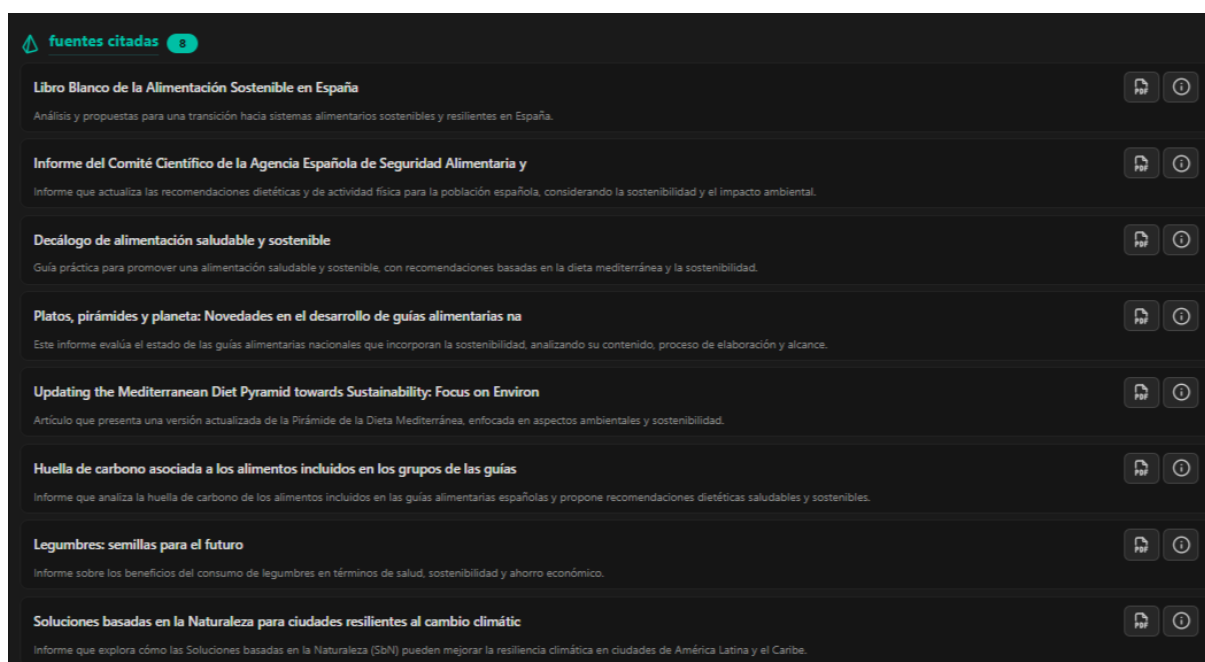
6.11.2. Sistema de traza del pipeline

Más allá de las citas visibles para el usuario, la EKP registra una traza completa de cada interacción que documenta todas las operaciones internas del pipeline. Esta traza constituye una auditoría técnica exhaustiva que permite al equipo de gestión entender exactamente qué ha ocurrido en cada paso del proceso. El sistema registra los siguientes eventos con marca de tiempo:

- Recepción de la consulta: Mensaje del usuario con identificador de sesión.
- Filtro de seguridad: Resultado de la evaluación de contenido (aprobado/bloqueado) y tiempo de procesamiento.
- Evaluación de contexto (Context Manager): Modelo utilizado, tokens consumidos, coste de la evaluación, nivel de confianza, área temática detectada y subáreas identificadas. Incluye la decisión tomada (contexto suficiente, clarificación necesaria o respuesta directa) y la query reformulada para optimizar la búsqueda.
- Configuración de herramientas: Preparación de las herramientas de búsqueda con los corpus disponibles (ECODES, Normativo, Evidencia).
- Construcción del prompt: Número de mensajes, tokens estimados del sistema, turnos de historial y porcentaje de uso de la ventana de contexto del modelo.
- Llamada al LLM: Modelo utilizado, tiempo hasta el primer token (TTFT) y decisión del modelo sobre qué herramienta invocar.
- Búsqueda en el corpus: Query utilizada, corpus consultado, área temática aplicada como filtro, número de documentos encontrados y tamaño del contexto recuperado.
- Generación de la respuesta: Número de chunks de streaming, tokens de salida, tiempo total de generación.
- Auditoría de alineamiento: Subáreas evaluadas, grado de alineamiento (alto/medio/bajo), número de cambios aplicados, posicionamientos consultados y detalle de cada corrección realizada.

6.11.3. Panel de fuentes RAG

Al finalizar cada conversación, el sistema presenta un panel resumen de las fuentes RAG que detalla para cada documento recuperado: su título, corpus de origen, puntuación de similitud coseno (relevancia semántica), puntuación de re-ranking (relevancia refinada) y si ha sido efectivamente citado en la respuesta. Esta información permite al equipo de gestión evaluar la calidad de la recuperación: verificar que los documentos más relevantes están siendo priorizados, detectar si el re-ranking está mejorando el orden de los resultados, e identificar documentos que se recuperan, pero no se citan (potencialmente irrelevantes o redundantes).



6.11.4. Resumen de consumo y huella por conversación

El conjunto de estos mecanismos de trazabilidad convierte a la EKP en una plataforma donde ninguna respuesta es opaca: cada afirmación tiene su fuente, cada fuente tiene su puntuación de relevancia, cada paso del pipeline tiene su registro temporal y cada conversación tiene su coste y su huella. Esta transparencia no es solo una garantía técnica, sino un requisito ético para una organización que promueve la rendición de cuentas y la responsabilidad ambiental.

Cada conversación se cierra con un resumen cuantitativo que incluye los tokens de entrada y salida consumidos, el total de tokens utilizados, el coste económico de la conversación en euros, el número de intercambios realizados, la estimación de emisiones de CO2 equivalente en miligramos y su equivalencia en kilómetros recorridos en coche. Este resumen materializa el compromiso de la EKP con la transparencia operativa: cada interacción con los asistentes tiene un coste medible y un impacto ambiental estimable, y el usuario o el administrador puede conocerlo.

El Anexo A de este informe recoge tres demostraciones reales de trazabilidad integral, una por cada asistente de descarbonización (energía eléctrica, alimentación y movilidad), mostrando el ciclo completo desde la consulta del usuario hasta la respuesta con citas, incluyendo la traza del pipeline, el panel de fuentes RAG con sus puntuaciones de relevancia, la auditoría de alineamiento y el resumen de consumo con estimación de huella de carbono.

6.12. Relación con la prueba de concepto de asistentes IA R.2.1.3

La arquitectura descrita constituye la prueba de concepto exigida por R.2.1.3:

Requisito R.2.1.3	Implementación
Asistentes IA especializados	3 asistentes (energía, alimentación, movilidad)
Foco en descarbonización	Taxonomía de 5 ejes + filtros por área estratégica
Alimentación	95 documentos asignados, subáreas específicas
Consumo de energía eléctrica	158 documentos, subáreas de eficiencia y renovables
Movilidad	75 documentos, subáreas de movilidad eléctrica

CONCLUSIONES Y RESULTADOS

7.1. Cumplimiento de resultados esperados

7.1.1. R.2.1.2 — Almacén de datos curado

El presente informe documenta la construcción y operación de un almacén de datos curado que satisface plenamente los requisitos del resultado R.2.1.2:

Requisito	Estado	Evidencia
Fuentes científicas	✓ Cumplido	Corpus de Evidencia: 245 documentos (informes IPCC, estudios sectoriales, artículos técnicos)
Fuentes institucionales	✓ Cumplido	Corpus ECODES: 110 documentos (posicionamientos, guías, materiales institucionales)
Fuentes técnicas	✓ Cumplido	Corpus Normativo: 236 documentos (legislación EU, nacional, autonómica)
Alineado con criterios de descarbonización	✓ Cumplido	Taxonomía de 5 ejes con áreas estratégicas de descarbonización: Clima y Energía, Transición Energética, Alimentación, Movilidad
Identificando patrones	✓ Cumplido	Panel de control con distribución estadística. 376 documentos etiquetados con "Mitigación del cambio climático", 349 con "Eficiencia energética"
Barreras	✓ Identificadas	Infrarrepresentación de movilidad (17 docs. específicos), corpus de Contraste con 1 documento
Oportunidades	✓ Identificadas	Expansión del corpus de movilidad, fortalecimiento del debate (contraste), enriquecimiento con IEA/IPCC
Alimentación	✓ Cubierto	47 documentos en área específica + 95 asignados al asistente
Energía eléctrica	✓ Cubierto	135 documentos en área específica + 158 asignados al asistente

Movilidad	☒ Cubierto	17 documentos en área específica + 75 asignados al asistente
Base para arquitecturas de IA	☒ Cumplido	Pipeline PRISMA multi-corpus, búsqueda vectorial con re-ranking, embedding service

7.1.2. R.2.1.3 — Prueba de concepto de asistentes IA

Requisito	Estado	Evidencia
Asistentes IA especializados	☒ Desarrollados	3 asistentes operativos (energía, alimentación, movilidad)
Descarbonización	☒ Foco central	Taxonomía, filtros y posicionamientos centrados en mitigación climática
Alimentación	☒ Asistente operativo	95 documentos, búsqueda multi-corpus, citación trazable
Energía eléctrica	☒ Asistente operativo	158 documentos, subáreas de eficiencia, renovables, electrificación
Movilidad	☒ Asistente operativo	75 documentos, subáreas de movilidad eléctrica y sostenible

7.2. Métricas del almacén de datos

Métricas cuantitativas (mayo 2026)

Métrica	Valor
Total documentos gobernados	592
Corpus diferenciados	4 (Evidencia, Normativo, ECODES, Contraste)
Áreas estratégicas	10
Subáreas temáticas	15+
Ejes taxonómicos	5 cerrados + 1 abierto
Campos por documento	40+
Vectores en Qdrant	~3.500 puntos (chunks)
Dimensionalidad vectorial	1024
Modelos de IA integrados	4 (Cohere Embed, Cohere Rerank, Qwen2.5-72B, GPT-4o/5.2)
Asistentes especializados	3 (energía, alimentación, movilidad)

Métricas de infraestructura

Componente	Especificación
Almacenamiento S3	Bucket eu-central-1, cifrado SSE, URLs prefirmadas
Base vectorial	Qdrant Cloud, HNSW (m=16, ef=200), cosine distance
Embeddings	Cohere Embed Multilingual v3, batch 96, sin cold starts
Re-ranking	Cohere Rerank 3.5, two-stage retrieval
Chunking	1500 chars, overlap 300, Contextual Retrieval
Guardrails	Multi-capa (profanity, injection, embedding similarity)
Alineamiento	Auditor PRISMA v2 con corrección activa

7.3. Líneas futuras de investigación

7.3.1. Expansión del corpus

- Incorporación sistemática de informes IEA, IPCC Working Group III, y Agencia Europea de Medio Ambiente.
- Desarrollo del corpus de movilidad con documentos sobre logística sostenible, PMUS y transporte público eléctrico.
- Creación de un corpus de Contraste robusto con perspectivas críticas.

7.3.2. Aprendizaje por feedback: ciclo de mejora continua

La calidad de un asistente de IA no se establece en el momento de su puesta en marcha, sino que debe evolucionar con el uso. La EKP incorpora un mecanismo de retroalimentación a nivel de mensaje que permite a los usuarios valorar cada respuesta del asistente como positiva o negativa. Este gesto, aparentemente sencillo, genera un flujo de señales que, debidamente explotado, abre cuatro vías de mejora continua.

La primera es la **incorporación al corpus de respuestas validadas**. Cuando un especialista marca una respuesta como positiva, esa respuesta constituye un producto de conocimiento en sí misma: es la síntesis que el asistente ha construido combinando múltiples fuentes del corpus, filtrada por los guardrails y alineada con los posicionamientos de ECODES, y que un experto humano ha considerado correcta y útil. Estas respuestas validadas podrían incorporarse al almacén como un

nuevo tipo de contenido —un "gold standard"— que enriquecería el corpus con conocimiento derivado, no solo con fuentes primarias.

La segunda vía es la **detección de patrones de fallo**. Las valoraciones negativas no solo señalan una respuesta concreta que no ha sido útil, sino que, analizadas en conjunto, revelan carencias estructurales: áreas temáticas donde el almacén carece de cobertura suficiente, tipos de consulta donde el re-ranking no prioriza los documentos más pertinentes, o formulaciones que el pipeline interpreta de forma incorrecta. Identificar estos patrones permite al Comité de Gobernanza actuar de forma proactiva: ampliar el corpus en las áreas deficitarias, ajustar los parámetros de búsqueda o refinar las estrategias de chunking para los documentos problemáticos.

La tercera es el **refinamiento de los prompts del sistema**. El análisis de las respuestas negativas proporciona información directa sobre qué instrucciones del system prompt no están produciendo el comportamiento esperado. Si un asistente tiende a dar respuestas demasiado genéricas ante consultas técnicas específicas, o a omitir la normativa aplicable cuando se le pregunta por recomendaciones prácticas, esos patrones pueden corregirse ajustando las instrucciones que recibe el modelo.

La cuarta es la construcción de **métricas de calidad basadas en datos reales**: ratio de satisfacción por asistente, por subárea temática, por modelo de lenguaje utilizado y por tipo de consulta. Estas métricas permitirían al equipo tomar decisiones de configuración fundamentadas —qué modelo asignar a cada asistente, qué corpus ampliar, qué roles priorizar— sin depender exclusivamente de la intuición o la experiencia.

La infraestructura para este ciclo de feedback ya está implementada en la base de datos. Su explotación completa es una de las líneas de desarrollo prioritarias de la siguiente fase del proyecto.

7.3.3. Base de datos de centro de información

La integración de búsqueda web en los asistentes de IA resuelve un problema real —el desfase temporal del corpus curado—, pero introduce uno nuevo: cada consulta a la API de Tavily consume recursos económicos, depende de la disponibilidad de un servicio externo y añade latencia a la respuesta.

En la práctica, una parte significativa de estas búsquedas son recurrentes: los usuarios preguntan con frecuencia por las mismas normativas vigentes, los mismos informes de referencia o los mismos portales institucionales. Repetir la búsqueda web cada vez que un usuario pregunta por el PNIEC, por una directiva europea o por los datos de la Red Eléctrica de España es un gasto evitable.

La solución propuesta es la creación de una **base de datos de URLs frecuentes** que funcione como caché inteligente entre los asistentes y la API de búsqueda web. El sistema operaría en cuatro fases:

1. **Consulta local prioritaria.** Cuando un asistente determina que necesita información web para completar una respuesta, antes de invocar la API de Tavily consulta la caché interna buscando contenido relevante ya almacenado.
2. **Evaluación de frescura.** Si la caché contiene un resultado relevante, el sistema verifica su antigüedad contra un umbral de vigencia configurable por tipo de fuente: 24 horas para datos que cambian con frecuencia (precios de energía, noticias), 7 días para informes institucionales, o 30 días para normativa consolidada que se actualiza con poca frecuencia. Si el contenido está dentro del umbral, se devuelve directamente sin realizar ninguna llamada externa.
3. **Búsqueda y almacenamiento.** Cuando no hay coincidencia en la caché o el contenido ha expirado, el sistema realiza la búsqueda web convencional a través de Tavily y, antes de devolver los resultados al asistente, almacena el contenido en la caché para futuras consultas, indexándolo vectorialmente para que sea localizable por similitud semántica.
4. **Actualización proactiva programada.** Para las fuentes identificadas como prioritarias —portales de normativa como EUR-Lex o el BOE, repositorios de datos como el de la Red Eléctrica, webs institucionales de ECODES—, un proceso periódico automatizado actualiza su contenido en la caché sin esperar a que un usuario lo solicite, garantizando que la información más consultada está siempre disponible y actualizada.

El impacto estimado de esta funcionalidad es una reducción del 40-60% en las llamadas a la API de Tavily para patrones de consulta recurrentes, lo que se traduce en un ahorro directo en costes de API, una disminución de la latencia percibida por el usuario (la consulta local es prácticamente instantánea frente a los 1-2 segundos de una búsqueda web), y una mayor resiliencia del sistema ante posibles caídas o limitaciones del servicio externo.

7.3.4. Alertas proactivas de normativa

Aprovechando el servidor MCP del BOE ya implementado, se podría desarrollar un sistema de alertas proactivas:

- **Monitorización diaria:** Escaneo automático del BOE buscando normativa relevante para las áreas de descarbonización.
- **Notificación a curadores:** Cuando se detecta nueva normativa relevante (por ejemplo, una directiva sobre renovables), el sistema notifica a los especialistas de ECODES para que evalúen su incorporación al almacén.
- **Actualización automática del corpus:** Para normativa actualizada, el sistema podría proponer automáticamente la sustitución del documento anterior por la nueva versión.

7.3.5. Comparador de posicionamientos

Una herramienta visual que muestre cómo el posicionamiento oficial de ECODES se alinea (o diverge) de la normativa vigente y la evidencia científica:

- **Vista comparativa:** Tabla que cruce temas de descarbonización con las posiciones de ECODES, la normativa aplicable y la evidencia científica disponible.
- **Detección de brechas:** Identificación automática de temas donde ECODES no tiene posicionamiento explícito, pero existe regulación o evidencia relevante.
- **Coherencia narrativa:** Verificación de que los posicionamientos de ECODES son coherentes entre sí y con la evidencia más reciente.

7.3.6. Evaluación de calidad RAG (Benchmarking)

Actualmente, la calidad de las respuestas de los asistentes se evalúa de forma cualitativa: el equipo formula consultas, revisa las respuestas y valora subjetivamente si los documentos recuperados son pertinentes y si la síntesis es correcta. Este enfoque es útil en las primeras fases de desarrollo, pero no escala ni permite comparar de forma objetiva el impacto de los cambios técnicos que se realizan en el pipeline.

Para profesionalizar esta evaluación, se propone la implantación de un sistema de benchmarking estructurado que permita medir la calidad del sistema RAG con rigor cuantitativo. Este sistema se articularía en tres componentes.

El primero es la **construcción de un corpus de evaluación anotado**: un conjunto de consultas representativas para cada asistente (energía, alimentación, movilidad) junto con las respuestas esperadas y los documentos que deberían recuperarse para fundamentarlas. Este dataset, elaborado por los especialistas del Comité de Gobernanza, actuaría como referencia objetiva —un examen con respuestas correctas— contra la que medir el rendimiento del sistema. Las métricas calculadas serían el estándar en recuperación de información: precisión (qué proporción de los documentos recuperados son relevantes), recall (qué proporción de los documentos relevantes se han recuperado) y F1 (la media armónica de ambas).

El segundo componente es un sistema de **A/B testing de configuraciones**. Cada vez que el equipo técnico considera un cambio en el pipeline —un tamaño de chunk diferente, un nuevo modelo de embeddings, un umbral de re-ranking distinto, una estrategia de limpieza de texto alternativa—, el corpus de evaluación permite ejecutar ambas configuraciones sobre las mismas consultas y comparar los resultados de forma objetiva. Esto elimina la incertidumbre que acompaña a los cambios técnicos: en lugar de preguntarse "¿habrá mejorado?", el equipo puede afirmar "la precisión ha pasado del 78% al 84% en consultas de energía eléctrica".

El tercer componente es un **dashboard de métricas de calidad** que visualice la evolución de estos indicadores a lo largo del tiempo, segmentados por asistente y por área temática. Este panel permitiría detectar regresiones de calidad tras actualizaciones del corpus, identificar las áreas temáticas donde el rendimiento de la búsqueda es más débil y fundamentar con datos las decisiones de inversión en ampliación o curación de contenidos.

7.3.7. Modo investigador

Los asistentes actuales de la EKP están diseñados para responder preguntas concretas en una única ronda de interacción: el usuario formula una consulta, el pipeline recupera los documentos relevantes y el modelo genera una respuesta. Este flujo es eficaz para preguntas directas ("¿Qué dice la normativa europea sobre eficiencia energética en edificios?"), pero resulta insuficiente cuando el usuario necesita un análisis en profundidad que abarque múltiples dimensiones de un mismo tema.

El Modo Investigador ampliaría esta capacidad permitiendo al asistente realizar procesos de investigación autónoma en múltiples rondas. Ante una consulta compleja —por ejemplo, "Elabora un análisis comparativo de las estrategias de descarbonización del transporte urbano en España, Francia y Alemania, incluyendo marco normativo, evidencia de impacto y posición de ECODES"—, el asistente descompondría automáticamente la consulta en sub-preguntas más específicas (normativa española, normativa francesa, estudios de impacto, posicionamientos institucionales) y ejecutaría búsquedas sucesivas en el corpus y en la web para cada una de ellas.

Cada ronda de búsqueda refinaría la siguiente: los hallazgos de las primeras consultas orientarían las sub-preguntas posteriores, permitiendo al asistente profundizar progresivamente en los aspectos más relevantes o cubrir lagunas detectadas en los primeros resultados. El proceso culminaría con una síntesis estructurada que integre los hallazgos de todas las rondas en un informe coherente, con secciones diferenciadas, referencias completas a cada fuente consultada y un resumen ejecutivo de las conclusiones.

Este modo transformaría al asistente de una herramienta de consulta puntual en un verdadero colaborador de investigación, capaz de producir borradores de informes técnicos que el especialista de ECODES podría revisar, completar y publicar, reduciendo significativamente el tiempo dedicado a la recopilación y síntesis inicial de información.

7.3.8. Impacto

La EKP ha demostrado su viabilidad técnica, pero su valor real se medirá por el efecto que produzca en las personas que la utilicen. Las siguientes fases del proyecto contemplan tres líneas de evaluación orientadas a cuantificar ese impacto y a compartir los aprendizajes obtenidos.

La primera es la medición del impacto en la toma de decisiones. Más allá de las métricas técnicas de calidad RAG, interesa saber si los asistentes están influyendo de forma efectiva en las decisiones de los usuarios: si un técnico municipal modifica su plan de movilidad tras consultar al asistente, si un nutricionista incorpora criterios de huella de carbono en sus recomendaciones, o si un gestor energético identifica opciones de descarbonización que desconocía. Esta evaluación requiere metodologías específicas —encuestas pre/post uso, seguimiento de acciones derivadas, entrevistas cualitativas— que permitan trazar el vínculo entre la consulta al asistente y la decisión tomada.

La segunda línea son los estudios de usabilidad con usuarios reales de los tres sectores de descarbonización. La plataforma ha sido desarrollada y probada por el equipo de ECODES, pero su despliegue efectivo requiere validar que profesionales con perfiles diversos —técnicos, decisores, comunicadores— pueden interactuar con los asistentes de forma natural y productiva. Estos estudios permitirían detectar fricciones en la interfaz, identificar expectativas no cubiertas y ajustar tanto el tono de las respuestas como las funcionalidades de la plataforma a las necesidades reales de cada perfil de usuario.

La tercera es la publicación de la metodología como recurso de referencia para otras organizaciones. El enfoque de la EKP —almacén curado con gobernanza experta, pipeline RAG multi-corpus, auditoría de alineamiento institucional, estimación de huella de carbono— es replicable por cualquier organización que necesite construir asistentes de IA alineados con su misión y fundamentados en conocimiento verificado. Documentar y compartir esta metodología multiplica el impacto del proyecto más allá de ECODES, contribuyendo a que otras entidades del sector ambiental, social o institucional puedan adoptar un enfoque similar sin partir de cero.

7.3.9. Comité de Gobernanza de Inteligencia Artificial

La EKP cuenta ya con un Comité de Gobernanza de Datos que supervisa la calidad, clasificación y ciclo de vida del corpus documental. Sin embargo, a medida que la plataforma escala en número de asistentes, modelos y usuarios, emerge la necesidad de establecer un segundo órgano complementario: un **Comité de Gobernanza de IA** centrado no en los datos de entrada, sino en los resultados que produce la inteligencia artificial y en las implicaciones éticas, legales y operativas de su uso.

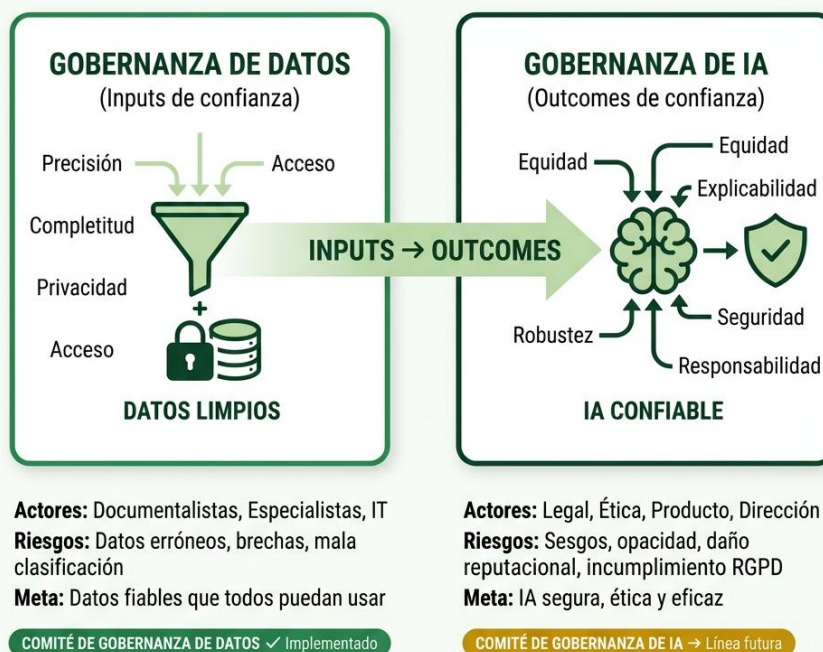
Mientras que la gobernanza de datos garantiza que lo que entra en el sistema es fiable (precisión, completitud, privacidad, acceso), la gobernanza de IA debe garantizar que lo que sale es seguro, justo, explicable y alineado con los valores de la organización (equidad, trazabilidad, robustez, responsabilidad). Ambos comités operarían de forma coordinada: el primero asegura *inputs* de confianza, el segundo asegura *outcomes* de confianza.

Funciones del Comité de Gobernanza de IA:

- **Supervisión ética y de alineamiento.** Definir y actualizar los principios éticos que rigen el comportamiento de los asistentes: límites de lo que pueden y no pueden responder, tratamiento de temas sensibles, equilibrio entre objetividad y posicionamiento institucional.
- **Cumplimiento normativo (RGPD, Ley de IA de la UE).** Garantizar que el uso de modelos de lenguaje cumple con el Reglamento General de Protección de Datos y con la Ley Europea de Inteligencia Artificial (AI Act), incluyendo la clasificación de riesgo de cada asistente, la documentación técnica exigida y los requisitos de transparencia hacia el usuario.
- **Control de costes y sostenibilidad.** Supervisar el gasto en tokens, llamadas a APIs y servicios cloud, estableciendo presupuestos por asistente y alertas de consumo. Evaluar periódicamente si los modelos utilizados son los más eficientes para cada tarea, en línea con el compromiso de ECODES con la minimización de la huella ambiental.
- **Evaluación de sesgos y equidad.** Diseñar y ejecutar auditorías periódicas para detectar sesgos en las respuestas de los asistentes: ¿favorecen determinadas perspectivas sobre otras? ¿Representan equitativamente las distintas realidades territoriales, socioeconómicas o culturales?
- **Gestión de incidentes.** Establecer protocolos de actuación ante respuestas incorrectas, alucinaciones, filtraciones de información privada o cualquier comportamiento inesperado de los asistentes, incluyendo mecanismos de desactivación temporal y comunicación al usuario.
- **Selección y evaluación de modelos.** Aprobar la incorporación de nuevos modelos de lenguaje a la plataforma, evaluando su rendimiento, coste, condiciones de licencia, localización del procesamiento y política de privacidad del proveedor.
- **Transparencia y rendición de cuentas.** Publicar informes periódicos sobre el uso de los asistentes, los costes generados, las incidencias registradas y las acciones correctivas adoptadas, garantizando que el uso de IA en ECODES es auditable y comunicable a los stakeholders.
- **Formación y cultura organizativa.** Promover la alfabetización en IA entre el equipo de ECODES, asegurando que todos los perfiles que interactúan con la plataforma comprenden sus capacidades, sus limitaciones y las implicaciones éticas de su uso.

GOBERNANZA DE DATOS vs GOBERNANZA DE IA

De los inputs de confianza a los outcomes de confianza



La creación de este comité no es una necesidad futura abstracta, sino una respuesta anticipada a un marco regulatorio europeo que ya está en vigor. La Ley de IA de la UE exige a las organizaciones que despliegan sistemas de IA documentar su funcionamiento, evaluar sus riesgos y establecer mecanismos de supervisión humana. Para ECODES, contar con un Comité de Gobernanza de IA no solo es una buena práctica: es una garantía de que sus asistentes operan dentro de un marco de responsabilidad coherente con su misión.

7.3.10. Certificación ISO/IEC 42001: sistema de gestión de inteligencia artificial

La norma ISO/IEC 42001, publicada en diciembre de 2023, establece el primer estándar internacional para sistemas de gestión de inteligencia artificial (AIMS — Artificial Intelligence Management System). Su objetivo es proporcionar a las organizaciones un marco estructurado para desarrollar, desplegar y operar sistemas de IA de forma responsable, garantizando que se gestionan adecuadamente los riesgos, se cumplen los requisitos éticos y se mantiene una mejora continua de los procesos.

La EKP se encuentra en una posición favorable para abordar este proceso de certificación, ya que muchos de los requisitos de la norma están parcial o totalmente cubiertos por las funcionalidades ya implementadas:

- **Política de IA y objetivos (cláusula 5).** La filosofía de alineamiento, los principios de diseño y los posicionamientos institucionales codificados en la plataforma constituyen una base sólida para formalizar la política de IA de ECODES.
- **Evaluación de riesgos (cláusula 6.1).** Los guardrails multicapa, el auditor de alineamiento y el sistema de roles ya implementan controles de riesgo operativos. Sería necesario formalizarlos en un registro de riesgos documentado con su evaluación de impacto y probabilidad.
- **Gestión de datos (Anexo A, control A.7).** La gobernanza de datos con taxonomía de 5 ejes, curación experta, trazabilidad integral y panel de control cubre extensamente los requisitos de calidad, procedencia y ciclo de vida del dato.
- **Transparencia y explicabilidad (Anexo A, control A.8).** El sistema de traza del pipeline, la citación de fuentes, el panel de fuentes RAG y el resumen de consumo por conversación proporcionan un nivel de transparencia que excede lo habitual en sistemas de IA conversacional.
- **Seguimiento y medición (cláusula 9).** La auditoría de tokens, la estimación de emisiones de CO₂ y las métricas del panel de control ofrecen indicadores cuantitativos sobre el rendimiento y el impacto del sistema.

Áreas que requerirían desarrollo adicional para la certificación:

- **Documentación formal del AIMS.** Redacción del manual del sistema de gestión de IA, incluyendo alcance, política, roles, procedimientos y registros según la estructura exigida por la norma.
- **Análisis de impacto en derechos fundamentales.** Evaluación formal del impacto de los asistentes de IA en los derechos de las personas, con especial atención a la no discriminación, la privacidad y la accesibilidad.
- **Auditoría interna y revisión por la dirección.** Establecimiento de un ciclo de auditorías internas periódicas y de revisiones formales por parte de la dirección de ECODES, conforme al modelo PDCA (Plan-Do-Check-Act) que exige la norma.
- **Gestión de proveedores de IA.** Formalización de los criterios de selección, evaluación y seguimiento de los proveedores de modelos de lenguaje (OpenAI, AWS Bedrock, Cohere), incluyendo acuerdos de nivel de servicio, políticas de privacidad y garantías de localización de datos.
- **Formación y competencia.** Programa de capacitación documentado para todos los miembros del equipo que interactúan con el sistema de IA, desde los gobernantes de datos hasta los usuarios finales.

La obtención de esta certificación posicionaría a ECODES como una de las primeras organizaciones del tercer sector en Europa con un sistema de gestión de IA certificado bajo estándar internacional, reforzando su credibilidad como referente en el uso responsable de la tecnología al servicio de la descarbonización.

7.3.11. Memoria conversacional y perfilado del usuario

Actualmente, los asistentes de la EKP mantienen el historial de cada conversación y almacenan metadatos asociados (modelo utilizado, tokens consumidos, fuentes citadas), pero cada sesión es esencialmente independiente: el asistente no recuerda lo que el usuario preguntó la semana pasada, no conoce su perfil profesional ni puede adaptar progresivamente sus respuestas al nivel de conocimiento o los intereses recurrentes de quien le consulta.

Avanzar en la gestión de la memoria de los asistentes abriría una dimensión completamente nueva en la personalización y la eficacia de las respuestas. Se proponen tres niveles de desarrollo:

Nivel 1 — Memoria de sesión extendida. Permitir que el asistente acceda al historial de conversaciones anteriores del mismo usuario, identificando temas recurrentes, preguntas ya resueltas y líneas de consulta habituales. De este modo, si un técnico municipal lleva semanas preguntando sobre zonas de bajas emisiones, el asistente podría contextualizar automáticamente sus nuevas consultas sin necesidad de repetir los antecedentes.

Nivel 2 — Perfil de usuario semántico. Construir, con el consentimiento explícito del usuario, un perfil temático basado en el análisis de sus interacciones: áreas de interés predominantes, nivel de profundidad técnica con el que suele preguntar, sector profesional, preferencias de formato de respuesta (tablas, listas, texto narrativo). Este perfil se almacenaría como un embedding semántico que el asistente consultaría al inicio de cada sesión para calibrar el tono, la extensión y el nivel de detalle de sus respuestas.

Nivel 3 — Recomendación proactiva de contenidos. A partir del perfil acumulado, el asistente podría sugerir al usuario documentos nuevos incorporados al almacén que se alineen con sus intereses, alertas de cambios normativos en las áreas que consulta habitualmente, o conexiones entre temas que ha explorado por separado y que podrían enriquecer su comprensión (por ejemplo: un usuario que pregunta sobre eficiencia energética en edificios y sobre pobreza energética podría recibir la sugerencia de explorar la rehabilitación energética como nexo entre ambas preocupaciones).

Consideraciones de privacidad y consentimiento. Cualquier avance en esta dirección estaría condicionado al cumplimiento estricto del RGPD y a la obtención del consentimiento informado del usuario. La plataforma debería ofrecer control total al usuario sobre su perfil: consultar qué datos se han almacenado, modificarlos o eliminarlos en cualquier momento, y decidir el nivel de personalización que desea activar. Este enfoque se alinea con los principios de privacidad por diseño que ya

rigen el sistema de permisos de la EKP y con los requisitos de la norma ISO/IEC 42001 sobre transparencia y derechos del usuario.

La implementación de esta funcionalidad transformaría al asistente de una herramienta de consulta genérica en un colaborador que aprende del usuario y mejora con cada interacción, manteniendo siempre al usuario como propietario y garante de sus propios datos.

7.4. Conclusión

El trabajo desarrollado en el marco del Objetivo Específico OE2.1 ha ido más allá de la creación de un prototipo tecnológico. Lo que comenzó como el diseño de un almacén de datos curado y una prueba de concepto de asistentes de IA ha cristalizado en una plataforma integral de gestión del conocimiento para la descarbonización —la ECODES Knowledge Platform (EKP)— que incorpora una visión propia sobre cómo debe construirse, gobernarse y utilizarse la inteligencia artificial cuando está al servicio de una organización con una misión tan concreta como urgente.

Un almacén que no solo almacena, sino que gobierna. Los 592 documentos que componen el corpus de la EKP no son una acumulación de archivos. Cada uno ha sido identificado por un especialista, evaluado en su calidad y pertinencia, analizado por inteligencia artificial, validado por un experto humano y clasificado según una taxonomía de cinco ejes diseñada específicamente para la descarbonización. Este proceso de curación —supervisado por un Comité de Gobernanza multidisciplinar que reúne competencias documentales, sectoriales, comunicativas y tecnológicas— es lo que transforma un repositorio de documentos en un activo estratégico de conocimiento. La distinción en tres corpus complementarios (Evidencia científica, Normativo y ECODES) no es una cuestión de organización interna: es la estructura que permite a los asistentes fundamentar cada respuesta en datos, en legislación y en la perspectiva institucional de la organización, de forma simultánea y trazable.

Asistentes que no inventan: razonan sobre evidencia. Los tres asistentes desarrollados como prueba de concepto —energía eléctrica (158 documentos), alimentación (95 documentos) y movilidad (75 documentos)— demuestran que es posible construir herramientas conversacionales que no generan información, sino que la recuperan, la contrastan y la citan. Cada respuesta lleva asociada la referencia al documento fuente, al corpus de origen y al fragmento exacto del que se extrae la información. Esta trazabilidad integral, desde el PDF que un especialista ingresó en el sistema hasta la cita que aparece en la respuesta al usuario, no es un añadido técnico: es el fundamento de la confianza que una organización como ECODES necesita depositar en sus herramientas de IA para poder respaldarlas públicamente.

Una arquitectura diseñada para la coherencia institucional. La contribución técnica más significativa de este proyecto no es ningún componente individual, sino la forma en que se articulan. El pipeline PRISMA orquesta la búsqueda simultánea en los tres corpus. Los guardrails basados en embeddings filtran consultas irrelevantes o malintencionadas sin consumir un solo token adicional. El auditor de alineamiento narrativo contrasta cada respuesta con los posicionamientos oficiales de ECODES antes de entregarla al usuario. Los servidores MCP extienden las capacidades de los asistentes más allá del corpus curado. Y el sistema de roles permite que un mismo asistente se comporte como analista de política energética, como asesor técnico o como comunicador ambiental, según la necesidad del momento. Es esta arquitectura completa —no la suma de sus partes— la que hace que la EKP sea algo más que un chatbot con documentos.

Responsabilidad con la propia herramienta. Para una organización cuya razón de ser es la descarbonización, medir el impacto ambiental de sus propias herramientas de IA no es un ejercicio opcional. La EKP registra cada token consumido, cada modelo utilizado y cada llamada a proveedores externos, y permite estimar —aunque sea de forma aproximada— la huella de carbono asociada a cada conversación. Las estrategias de optimización implementadas (guardrails sin tokens, modelos ligeros para tareas auxiliares, caché de embeddings, selección de modelo por tarea) reflejan un diseño consciente: obtener la máxima calidad con el mínimo consumo de recursos.

Un modelo replicable. La arquitectura, las decisiones de diseño y las lecciones aprendidas que documenta este informe no son específicas de ECODES. Cualquier organización que necesite construir asistentes de IA alineados con su misión institucional —fundaciones, centros de investigación, administraciones públicas, organizaciones del tercer sector— encontrará en la EKP un modelo de referencia: cómo gobernar el conocimiento, cómo garantizar la trazabilidad, cómo proteger la coherencia narrativa y cómo medir el impacto de lo que se construye.

El camino por delante. Las líneas futuras descritas en este capítulo —aprendizaje por feedback, grafo documental, caché de URLs, alertas proactivas, benchmarking, modo investigador— no son aspiraciones teóricas, sino extensiones naturales de una infraestructura que ya está preparada para soportarlas. La base de datos ya registra las valoraciones de los usuarios. El tracking de tokens ya permite calcular costes y emisiones. El pipeline ya soporta múltiples rondas de búsqueda. Lo que queda por delante es activar estas capacidades latentes y, sobre todo, validarlas con usuarios reales en los tres sectores de descarbonización.

El resultado del OE2.1 no es solo una plataforma funcional. Es la demostración de que la inteligencia artificial puede ponerse al servicio de la descarbonización de forma rigurosa, trazable, alineada con los valores de una organización y consciente de su propio impacto.



Con el apoyo de:



GRACIAS

Contacto

 www.ecodes.org

 +34 976 29 82 82

 Plaza San Bruno, 9, 1º Oficinas, 50001 Zaragoza.