

INFORME DE RESULTADOS DEL ESTUDIO TECNOLÓGICO SOBRE ESPACIOS DE DATOS EN ESPAÑA

Mayo 2026

1. INTRODUCCIÓN Y CONTEXTO

El CRED (Centro de Referencia de Espacios de Datos de España) ha llevado a cabo un estudio dirigido a proyectos de espacios de datos en España, con el objetivo de obtener una visión actualizada del estado tecnológico del ecosistema nacional. Este estudio evalúa nueve dimensiones clave que abarcan desde el modelo de construcción de la solución hasta el nivel de madurez alcanzado.

El presente informe recoge y analiza 139 proyectos que participaron en el estudio (con los datos disponibles hasta la fecha) pertenecientes a quince sectores de actividad. Los datos permiten identificar tendencias dominantes, grados de alineación con los estándares predominantes (Gaia-X, IDS, DCAT-AP, ODRL, eIDAS2) y áreas donde el ecosistema aún requiere mayor desarrollo.

1.1 Distribución por Sector

La muestra del estudio cubre una variedad representativa de sectores de la economía española. Los de mayor presencia son el Agroalimentario (32 proyectos, 23%), Salud (24, 17%), Movilidad Sostenible (20, 14%) e Industria (17, 12%), que conjuntamente suman más del 65% del total. Esta amplitud sectorial otorga solidez estadística a las conclusiones del informe y permite identificar diferencias de enfoque tecnológico entre dominios con distinta cultura de digitalización y regulación de datos.

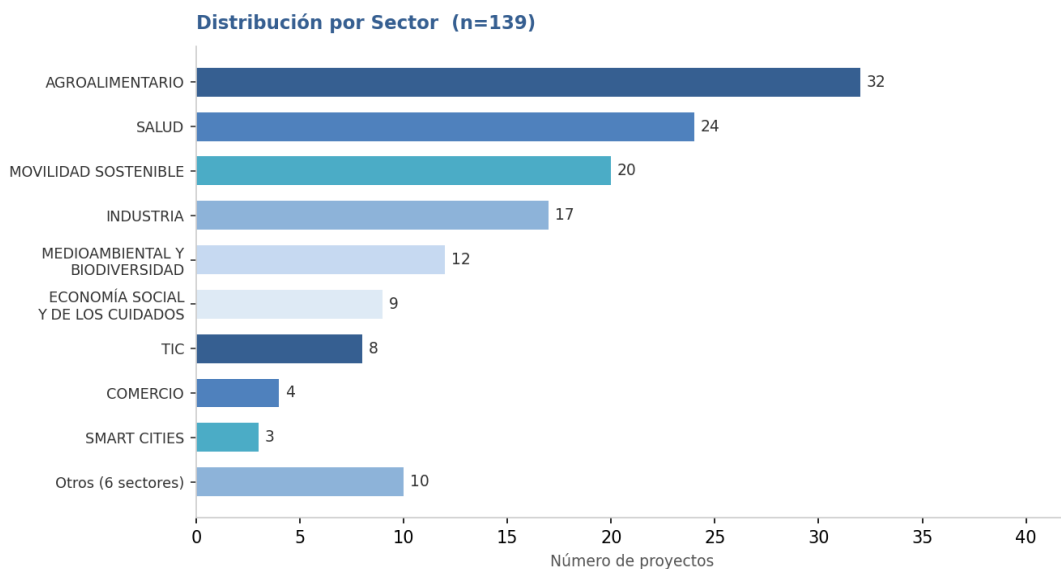


Figura 1. Distribución de proyectos por sector (n=139)

2. RESUMEN EJECUTIVO

La siguiente tabla sintetiza la opción dominante en cada una de las nueve dimensiones tecnológicas analizadas, ofreciendo una visión panorámica del estado del ecosistema.

Dimensión	Opción dominante	Respuestas	%
Modelo de solución	Desarrollo propio basado en estándares	83 / 138	60%
Despliegue	Cloud público (AWS, Azure, GCP...)	77 / 134	57%
Conector	EDC Eclipse	96 / 136	71%
Identidad y confianza	Credenciales verificables (Gaia-X / eIDAS2)	80 / 133	60%
Catálogo de datos	Basado en DCAT-AP	77 / 111*	69%
Políticas de uso	ODRL / estándar	91 / 132	69%
Auditoría	Logs técnicos	92 / 134	69%
Intercambio de datos	Sí, en piloto	54 / 135	40%
Nivel de madurez	Operativo básico	63 / 134	47%

* Respuestas con selección múltiple consolidadas.

Conclusión global: el ecosistema español de espacios de datos muestra una orientación firme hacia los estándares predominantes (EDC Eclipse, Gaia-X, DCAT-AP, ODRL), aunque la mayoría de los proyectos se sitúa aún en fases de piloto o madurez operativa básica.

En este sentido cabe decir que los espacios de datos en España están marcados por un desarrollo propio basado en estándares consolidados. En este aspecto se puede inferir que el ecosistema tiene un gran potencial de interoperabilidad con otros ecosistemas tanto a nivel nacional como europeo.

3. ANÁLISIS DETALLADO POR DIMENSIÓN

3.1 Modelo de Solución

¿Cómo está construida la solución de espacio de datos?

Este apartado indaga sobre la estrategia de construcción tecnológica del espacio de datos. Con la información recogida se puede constatar que la opción claramente dominante es el Desarrollo propio basado en estándares, elegida por 83 proyectos (60%). Esta cifra refleja la voluntad de los equipos de adoptar los marcos de referencia consolidados, al tiempo que mantienen el control sobre la implementación.

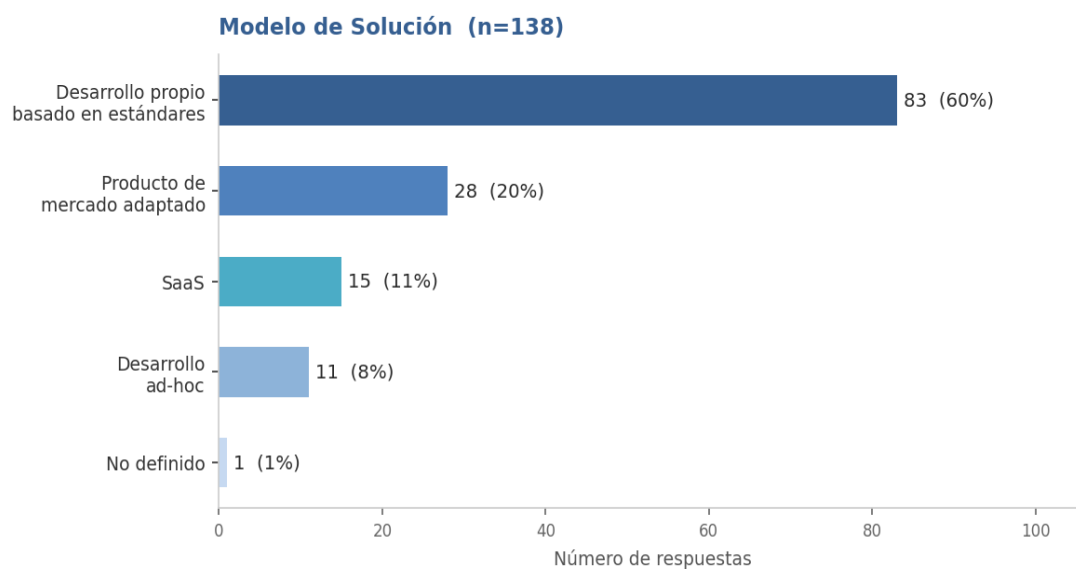


Figura 2. Distribución por modelo de solución (n=138)

- **Desarrollo propio basado en estándares (60%):** Es la estrategia preferida porque permite adaptar la arquitectura a las particularidades del sector y el caso de uso concreto, garantizando al mismo tiempo la interoperabilidad con el ecosistema europeo de datos. Los equipos que la eligen suelen contar con capacidades técnicas consolidadas o se apoyan en integradores con experiencia.
- **Producto de mercado adaptado (20%):** Una quinta parte opta por soluciones comerciales u *open source* configuradas o extendidas para el caso de uso. Esta opción equilibra velocidad de despliegue y alineación con estándares.
- **SaaS (11%):** Porcentaje relativamente bajo, posiblemente porque las soluciones SaaS específicas para espacios de datos aún son escasas en el mercado.

- **Desarrollo ad-hoc (8%):** Proyectos que, por sus particulares requerimientos o por estar en fases exploratorias iniciales, han optado por soluciones sin seguir estándares establecidos. Este grupo representa un desafío para la interoperabilidad en el sistema.

3.2 Modelo de Despliegue e Infraestructura

¿Cómo está desplegado el espacio de datos?

El cloud público es el entorno de despliegue preferido con 77 respuestas (57%), seguido del cloud privado/corporativo (17%) y el entorno híbrido (15%). La infraestructura on-premise representa solo el 10%.

En esta dimensión cabe destacar que la tendencia debería ser hacia infraestructuras que garanticen la soberanía del dato a los participantes pero considerando el estado de desarrollo de muchos espacios de datos la facilidad y rapidez de despliegue mediante cloud público representa la única opción.

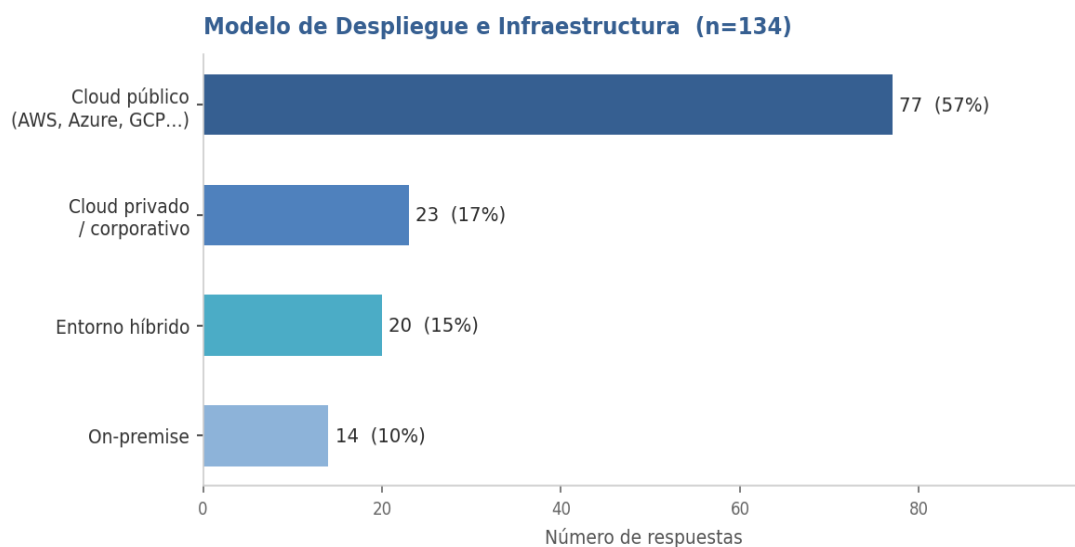


Figura 3. Distribución por modelo de despliegue (n=134)

- **Predominio del cloud público (57%):** La elasticidad, la velocidad de puesta en marcha y la disponibilidad de servicios gestionados hacen del cloud público la plataforma natural para proyectos innovadores con equipos de tamaño reducido. AWS, Azure y GCP concentran la mayoría de estos despliegues.
- **Cloud privado y entornos híbridos (32%):** Sectores como Salud, Industria y Administración Pública muestran mayor preferencia por estas modalidades, motivados por requisitos de soberanía del dato, cumplimiento normativo (GDPR, ENS) o integración con sistemas legados.
- **On-premise (10%):** Residual pero presente, especialmente en proyectos que gestionan datos sensibles o críticos. La tendencia europea apunta a soluciones de soberanía del dato que compatibilicen cloud con control territorial.

3.3 Conectores

¿Qué tecnología de conector se utiliza?

La tecnología de conector es quizás el indicador más revelador de madurez tecnológica. El conector EDC Eclipse (Eclipse Dataspace Components) ha sido adoptado por 96 proyectos, representando el 71% del total de proyectos analizados. Este dato es extraordinariamente significativo dentro del contexto europeo de espacios de datos.

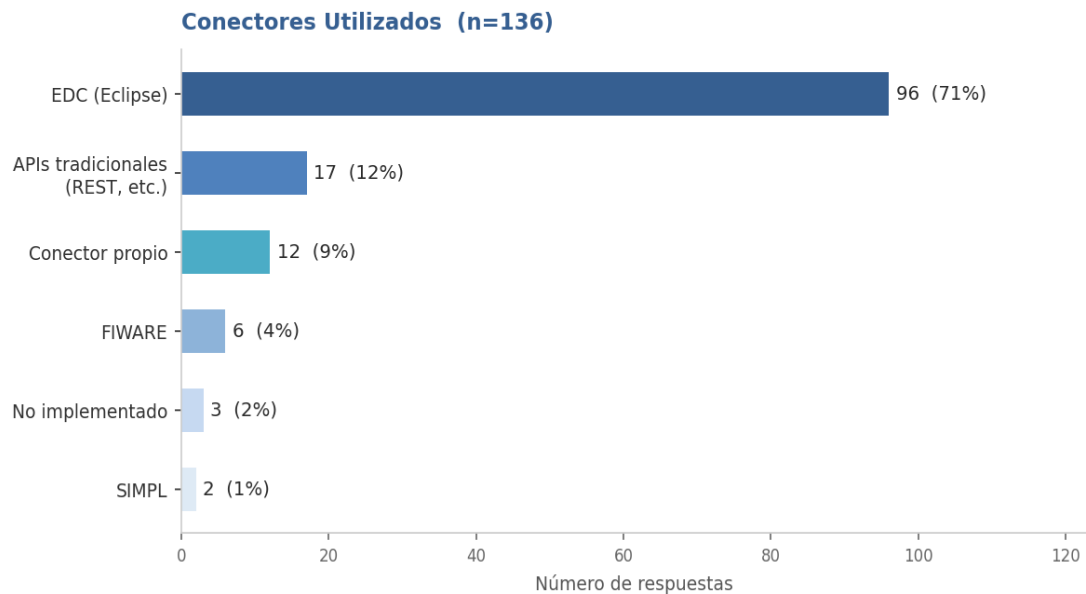


Figura 4. Distribución por tipo de conector utilizado (n=136)

La adopción del EDC Eclipse como conector estándar de facto en el 71% de los proyectos es el dato más destacado de todo el estudio. Refleja la consolidación de un componente tecnológico maduro, respaldado por la Fundación Eclipse y adoptado masivamente en el ecosistema europeo (Catena-X, Mobility Data Space, entre otros).

En este aspecto cabe destacar que el EDC (Eclipse Dataspace Connector) es el componente de referencia del protocolo IDS (International Data Spaces). Su adopción en España indica que:

- Los equipos técnicos han optado por un componente con comunidad activa, mantenimiento garantizado y alineación con el marco regulatorio europeo (Data Governance Act, Data Act).
- Existe una base tecnológica sólida para la interoperabilidad entre espacios de datos nacionales y transfronterizos.
- La industria española converge con el ecosistema europeo, lo que facilita la participación en proyectos Horizon y en el IDSA (International Data Spaces Association).

- **APIs tradicionales (13%):** Un 13% sigue utilizando APIs REST convencionales. Aunque funcionales, no incorporan las capacidades de control de políticas, negociación de contratos ni trazabilidad definidas por IDS.
- **FIWARE (4%) y SIMPL (1%):** Marcos con nicho específico, FIWARE en ciudades inteligentes y medio ambiente, SIMPL en el contexto europeo de Common European Data Spaces. Su presencia muestra la diversidad del ecosistema.

3.4 Identidad y Confianza

¿Cómo se gestionan la identidad y la confianza?

La gestión de la identidad es un pilar fundamental de los espacios de datos. Los resultados muestran una clara orientación hacia mecanismos avanzados basados en credenciales verificables alineadas con los estándares apoyados por la Comisión Europea.

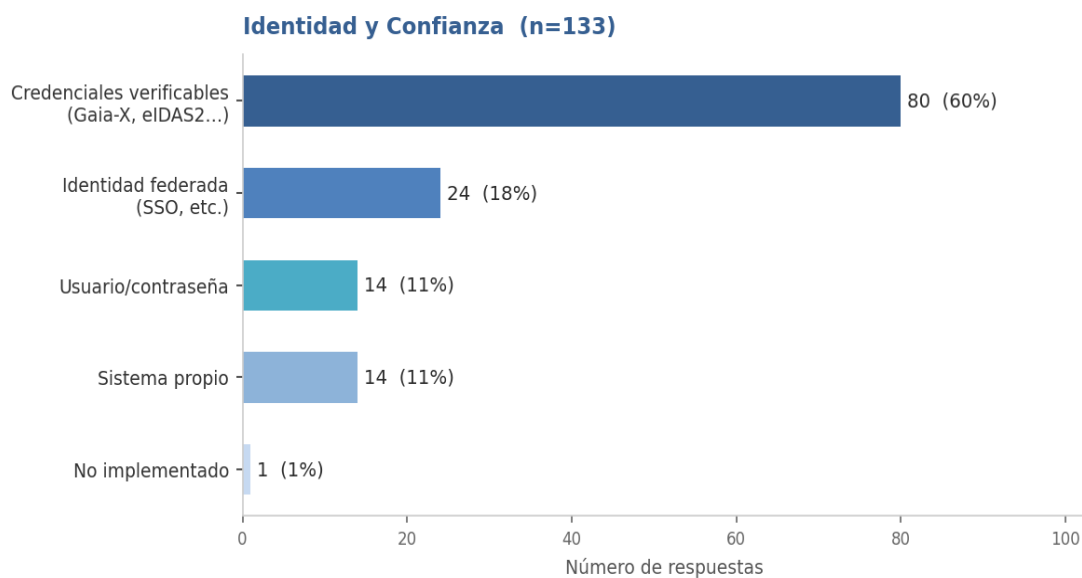


Figura 5. Distribución por mecanismo de identidad y confianza (n=133)

- **Credenciales verificables –Gaia-X / eIDAS2– (60%):** La mayoría de proyectos ha adoptado o está adoptando el modelo de Self-Sovereign Identity (SSI) a través de credenciales verificables conforme a los estándares W3C, alineadas con Gaia-X y el reglamento eIDAS2. Este enfoque garantiza la portabilidad de la identidad y la descentralización de la confianza, aspectos críticos para ecosistemas multiparticipante.
- **Identidad federada –SSO– (18%):** Soluciones tipo OpenID Connect / SAML que permiten reutilizar identidades corporativas existentes. Son habituales en proyectos con participantes del mismo sector o ecosistema organizativo.
- **Usuario/contraseña y sistema propio (21% combinados):** Porcentaje a vigilar. Estos mecanismos implican una gestión más centralizada de la confianza y menor interoperabilidad con terceros, lo que puede limitar la escalabilidad del espacio de datos.

3.5 Catálogo de Datos

¿Cómo se implementa el catálogo?

El catálogo permite a los participantes descubrir los activos de datos disponibles. DCAT-AP (Data Catalog Vocabulary – Application Profile) emerge como estándar dominante en las respuestas.

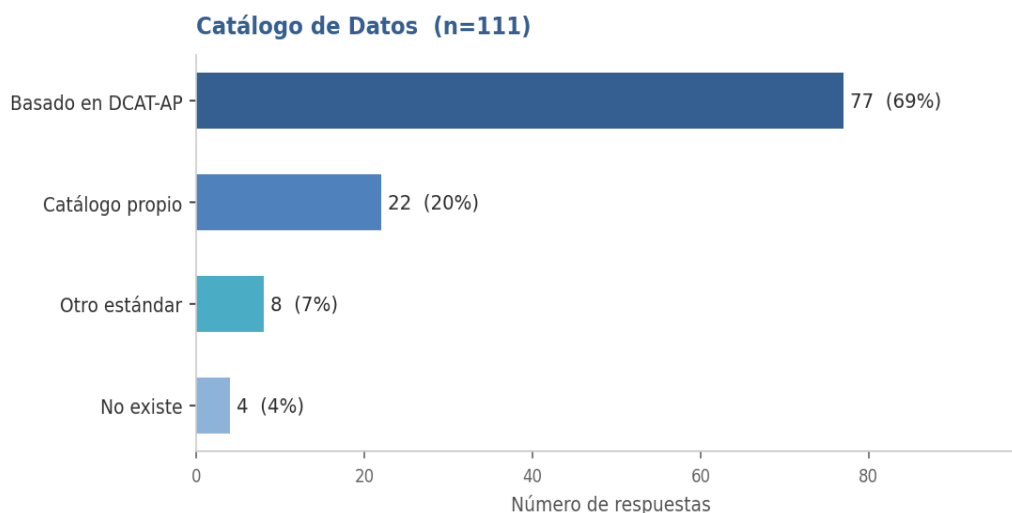


Figura 6. Distribución por tipo de catálogo de datos (n=111)

- **DCAT-AP (69% incluyendo combinaciones):** El estándar de la Unión Europea para la descripción de conjuntos de datos en portales de datos públicos europeos se ha convertido también en referencia en los espacios de datos privados. Su adopción garantiza la compatibilidad con el European Data Portal y con los catálogos federados propuestos en el marco de Common European Data Spaces.
- **Catálogo propio (20%):** Proyectos con soluciones de catalogación ad-hoc, normalmente con menor capacidad de interoperabilidad externa, aunque en algunos casos complementarias a DCAT-AP.
- **No existe (4%):** Un porcentaje pequeño pero relevante carece aún de catálogo. La ausencia limita significativamente la gobernanza y el descubrimiento de datos dentro del espacio a día de hoy. En este sentido esto puede ser debido al estado de madurez del proyecto o los datos recogidos en el estudio.

3.6 Políticas de Uso del Dato

¿Cómo se gestionan las políticas?

Las políticas de uso permiten al propietario del dato especificar las condiciones bajo las cuales puede ser accedido, transferido o procesado. Los resultados del análisis muestran una altísima adopción del estándar ODRL.

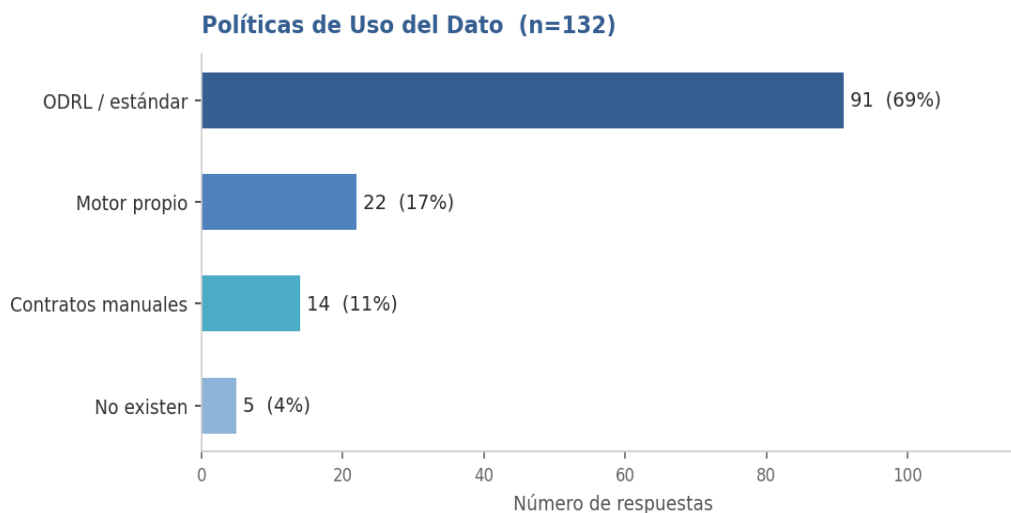


Figura 7. Distribución por mecanismo de políticas de uso del dato (n=132)

- **ODRL / estándar (69%):** El Open Digital Rights Language (ODRL) es el lenguaje de expresión de políticas adoptado por IDS. Su adopción en casi siete de cada diez proyectos indica un nivel de sofisticación notable: los espacios de datos no solo intercambian datos, sino que automatizan el cumplimiento de condiciones de uso mediante contratos inteligentes codificados en ODRL.
- **Motor propio (17%):** Funcional pero con riesgo de interoperabilidad cuando se integre con otros participantes que usen ODRL nativo.
- **Contratos manuales (11%):** La gestión manual implica menor automatización y mayor riesgo operativo a medida que el número de participantes escala.
- **No existe (4%):** En este aspecto cabe decir que esta situación puede ser debida al estado de desarrollo de los espacios de datos específicos (EDS IQVIA, PrivacyShield, DATAFORGE IGNITIO, Hestia) o la falta de datos en el momento de realización del estudio.

3.7 Auditoría y Trazabilidad

¿Cómo se implementa la trazabilidad?

La trazabilidad de los accesos y transferencias de datos es un requisito tanto técnico como regulatorio. Los resultados muestran que la mayoría de los proyectos se apoya en soluciones de logging técnico.

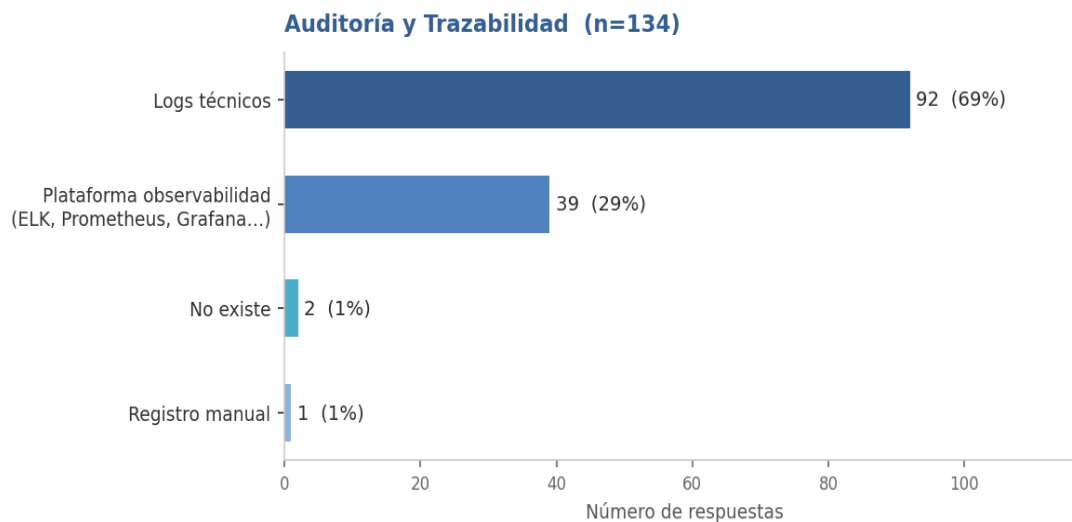


Figura 8. Distribución por mecanismo de auditoría y trazabilidad (n=134)

- **Logs técnicos (69%):** La mayoría genera registros técnicos de las operaciones. Si bien esto cubre los requisitos básicos, los logs técnicos puros son menos adecuados para auditorías regulatorias o para la demostración de cumplimiento frente a terceros, ya que carecen de la inmutabilidad y el contexto semántico que ofrecen plataformas especializadas.
- **Plataformas de observabilidad –ELK, Prometheus, Grafana– (29%):** Una parte significativa ha dado el paso hacia plataformas que permiten no solo almacenar logs sino correlacionarlos, visualizarlos y establecer alertas. Este nivel es más compatible con los requerimientos de trazabilidad del Data Act y del IDSA Clearing House.
- **Oportunidad de mejora:** Solo el 29% utiliza plataformas de observabilidad avanzadas. El incremento de esta cifra es clave para alcanzar niveles de madurez "Gobernado" y "Alineado con estándares".

3.8 Intercambio de Datos

¿Se realiza intercambio real entre participantes?

Este elemento mide el grado real de operatividad del intercambio de datos entre participantes, que es el objetivo último de cualquier espacio de datos.

Intercambio de Datos (n=135)

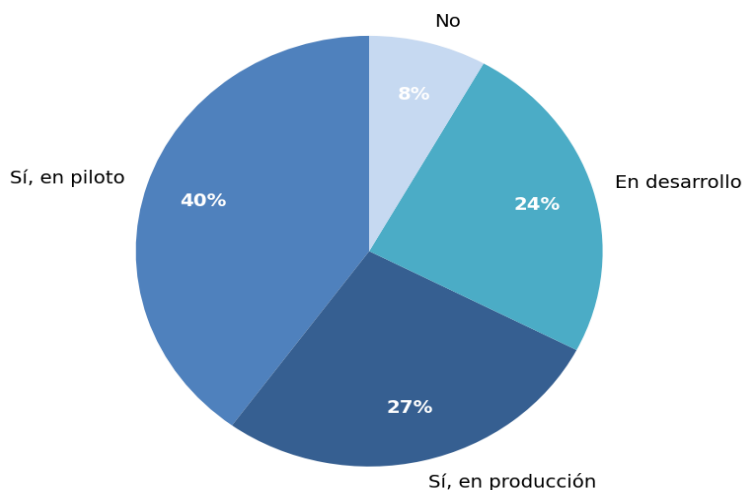


Figura 9. Estado del intercambio real de datos entre participantes (n=135)

- **Sí, en piloto (40%):** El grupo más numeroso está en fase piloto: el intercambio existe pero en un entorno controlado, con participantes limitados y funcionalidades aún en validación. Es el estadio previo a la producción.
- **Sí, en producción (27%):** Poco más de un cuarto ha alcanzado la producción real. Aunque positivo, revela que el ecosistema español está aún en proceso de maduración.
- **En desarrollo (24%):** Una cuarta parte no ha iniciado aún el intercambio real, ya que está construyendo la infraestructura o completando integraciones.
- **No (8%):** Un pequeño grupo no realiza intercambio de datos, esto se debe o porque se encuentra en fases muy iniciales o porque su modelo es fundamentalmente de catalogación.

Solo el 27% de los proyectos ha alcanzado intercambio real en producción. El 67% combinado (producción + piloto) confirma que el ecosistema ya está activo y generando valor, aunque mayoritariamente en entornos controlados.

3.9 Nivel de Madurez

El nivel de madurez es el indicador sintético que integra el resto de dimensiones. Los datos recabados nos muestran un ecosistema en transición entre las fases básicas y las fases avanzadas.

Nivel de Madurez (n=134)

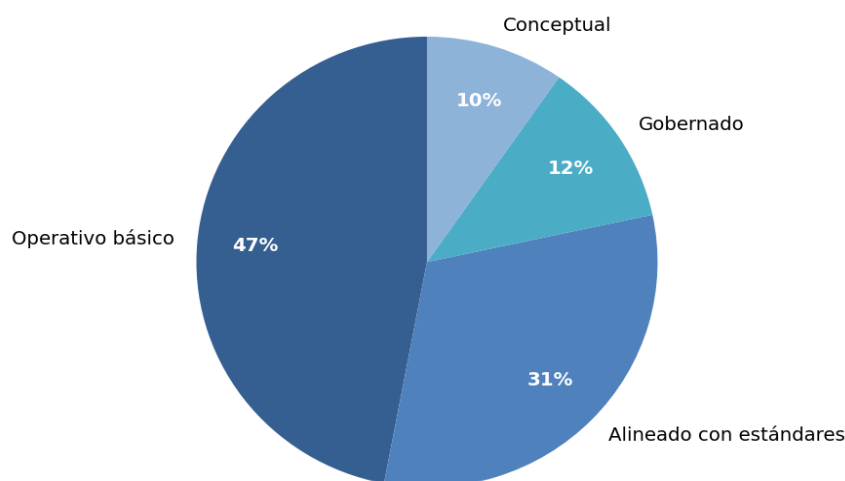


Figura 10. Distribución del nivel de madurez de los espacios de datos (n=134)

- **Operativo básico (47%):** La mayoría se definen en este nivel: el espacio de datos existe y funciona, pero sin una gobernanza robusta ni una alineación completa con estándares. Es el escalón natural tras la fase conceptual y el punto de partida para avanzar hacia niveles superiores.
- **Alineado con estándares (31%):** Un tercio declara estar alineado con los estándares relevantes en la diferentes dimensiones (Gaia-X, IDS, DCAT-AP, ODRL, eIDAS2). Este nivel implica interoperabilidad técnica y capacidad para participar en ecosistemas transfronterizos.
- **Gobernado (12%):** Proyectos con marcos de gobernanza formalizados, contratos de datos activos, auditoría regulatoria y gestión del ciclo de vida del dato. Nivel de referencia para los Common European Data Spaces.
- **Conceptual (10%):** Proyectos en fase de diseño o exploración, sin implementación operativa aún. Su presencia es esperable en un ecosistema en desarrollo.

4. CONCLUSIONES Y RECOMENDACIONES

El análisis de las diferentes iniciativas analizadas permite extraer las siguientes conclusiones sobre el estado del ecosistema español de espacios de datos.

4.1 Fortalezas detectadas

- Adopción masiva de EDC Eclipse (71%) como conector de referencia, lo que posiciona a España como ecosistema compatible con los principales Data Spaces europeos (Catena-X, Mobility Data Space, etc.).
- Orientación a estándares en identidad (60% credenciales verificables Gaia-X/eIDAS2), catálogo (69% DCAT-AP) y políticas (69% ODRL), reflejo de una madurez técnica notable.
- El 60% de los proyectos ha optado por desarrollo propio basado en estándares, mostrando capacidad técnica interna y compromiso con la interoperabilidad.
- El cloud público (57%) facilita la escalabilidad e innovación rápida, compatible con los requisitos de soberanía cuando se combina con otros mecanismos.

4.2 Áreas de mejora

- Solo el 27% ha alcanzado el intercambio de datos en producción real. Es necesario acelerar la transición de piloto a producción.
- El 69% utiliza únicamente logs técnicos para trazabilidad. La adopción de plataformas de observabilidad avanzadas (ELK, Prometheus, Grafana) mejoraría el cumplimiento regulatorio.
- El 8% de los proyectos que aún no realiza intercambio de datos y el 10% en fase conceptual requieren acompañamiento técnico para avanzar.
- El 21% que usa mecanismos de identidad no conformes deberían migrar hacia credenciales verificables para garantizar la interoperabilidad.

4.3 Recomendaciones del CRED

- Desarrollar un programa de acompañamiento para proyectos en nivel "Operativo básico" que quieran avanzar a "Alineado con estándares" o "Gobernado".
- Impulsar la adopción de plataformas de observabilidad y clearing houses para elevar los estándares de trazabilidad y auditoría.
- Fomentar la transición de proyectos en piloto a producción mediante casos de uso sectoriales demostrativos y grupos de trabajo interempresariales.

- Establecer un observatorio tecnológico que repita este estudio anualmente para medir la evolución del ecosistema.

El ecosistema español de espacios de datos muestra una orientación estratégica clara hacia los estándares europeos y una base tecnológica sólida muy bien encarada hacia la interoperabilidad entre espacios de datos. El reto principal en el próximo ciclo es consolidar la fase de producción y elevar el nivel de gobernanza en la mayoría de proyectos.