

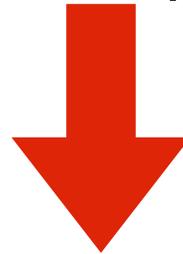
# Los datos de alto valor y su reutilización en la ciencia, la empresa y la administración.

Antonia Ferrer Sapena

# La esfera de los datos

El volumen global de datos pasará

**33 Zettabytes (ZB) en 2018**



**175 Zettabytes en 2025**



Dataesfera más grande del mundo

# Estrategia europea de datos

GAFA

Google

amazon



*“La **innovación basada en los datos** reportará enormes beneficios a los ciudadanos, por ejemplo, mediante la mejora de la medicina personalizada, la nueva movilidad y su contribución al Pacto Verde Europeo”*

# Estrategia europea de datos

Publicada por la Comisión para hacer frente y poder **competir en igualdad de oportunidades** con otros países/continentes.

Europa tiene que estar preparada para un nuevo marco donde **los productores de datos tengan un mayor control** sobre los datos que generan.

Los **portales de datos públicos deben conocer este nuevo entorno y adaptarse** lo más rápidamente posible a él.

# El valor de los datos

**Reside en su uso y reutilización.**

Una de las ventajas que representan los datos es que se pueden **replicar a un coste casi nulo**, y además pueden utilizarse simultáneamente por distintas personas o instituciones

# Problemas

- **No hay suficientes datos disponibles** para una reutilización innovadora, en particular en el caso del desarrollo de la inteligencia artificial. Datos tanto públicos como privados para intercambio entre empresas y administración, o a la inversa, o entre ambas.
- Existen **desequilibrios** en el poder de mercado. Son unos pocos los que lo dominan.
- La **interoperabilidad y calidad de los datos**, su estructura, autenticidad e integridad son clave para su explotación, especialmente en el contexto de la inteligencia artificial.

# Necesidades

**Reforzar la gobernanza de los datos.** Se precisan enfoques organizativos y estructuras que permitan la innovación basada en datos.

**Necesidad de infraestructuras y tecnologías de datos que permitan reducir su dependencia.**

**Existe escasez de mano de obra cualificada en materia de datos.**

# Definición de datos abiertos

**Por datos abiertos se entiende los datos en formatos abiertos que puede utilizar, reutilizar y compartir libremente cualquier persona con cualquier fin.**

- Con **disponibilidad y la reutilización generalizadas** con fines privados o comerciales, con restricciones mínimas o nulas de carácter jurídico, técnico o económico.

***«abiertos desde el diseño y por defecto»***

- Deben garantizar la **protección de los objetivos de interés público**: la seguridad pública, la información sensible sobre infraestructuras críticas, garantizar la protección de los datos personales.

# Principios FAIR

- Se adaptaran a los principios FAIR.
- Se facilitará también la información sobre qué datos pueden ser utilizados, cómo y por quién, cuando el fin sea la investigación científica.
- Se proporcionará un mayor volumen de datos de calidad del sector público para su reutilización, teniendo especialmente en cuenta el potencial que puede tener para las pyme.

**F**indable 

**A**ccessible 

**I**nteroperable 

**R**eusable 

# Tipos de datos

- Los **datos generados por el sector público**, así como el valor creado, deben estar disponibles para el bien común garantizando, en particular mediante un acceso preferencial, que estos datos sean utilizados por investigadores, otras instituciones públicas, pymes o empresas emergentes
- Utilización de la información del sector público por empresas
- El intercambio y el uso de datos de titularidad privada por otras empresas (intercambio de datos entre empresas o B2B). A pesar del potencial económico, el intercambio de datos entre empresas aún no ha despegado lo suficiente.
- Uso de datos de titularidad privada por las autoridades públicas
- El intercambio de datos entre las autoridades públicas

# Nuevas categorías de datos

**Los datos de investigación generados por la comunidad científica.**

También incluyen metadatos, especificaciones y otros objetos digitales.



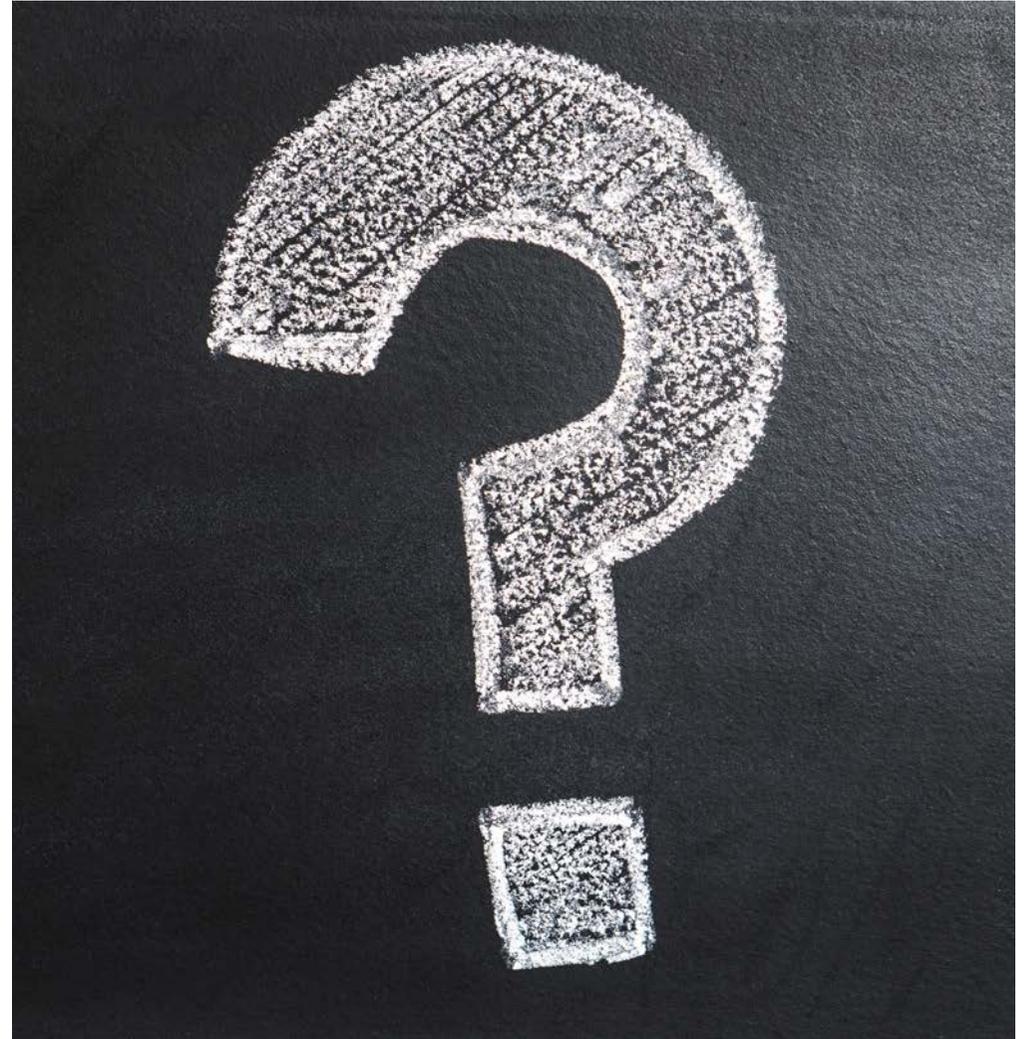
# ¿Cuáles son los datos abiertos de alto valor?

1. Geoespacial
2. Observación de la Tierra y medio ambiente
3. Meteorología
4. Estadística
5. Sociedades y propiedad de sociedades
6. Movilidad

Mas los tipos de datos que considera de relevancia: **los datos de investigación financiados con fondos públicos**, y se extienden las obligaciones de reutilización también a los metadatos

# ¿Donde nos encontramos?

**Empresa, administración y ciudadanos.  
A partir de la reutilización.**



# Informe Aseedie

***Informe 2020 de las empresas infomediarias de España*** conocen el valor de analizar y tratar la información tanto del sector público como del sector privado, creando a partir de ellos servicios de valor añadido.

## **Problemas del sector:**

- la dificultad en el acceso a los datos por que no se encuentran actualizados o disponibles.
- las restricciones al acceso que representa, en algunos casos, la normativa.

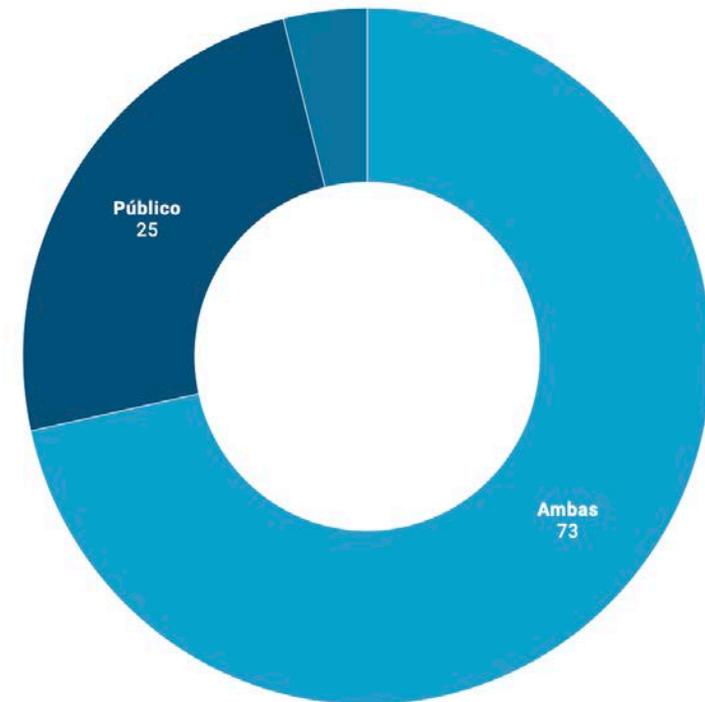
falta una ***“cultura del dato”*** consolidada

# Informe Aseedie

- 73% de las empresas encuestadas hacen uso tanto de fuentes públicas como privadas para la elaboración de sus productos o servicios.
- Un 23% hace un uso exclusivo de fuentes públicas y
- Un 4% explota solamente datos privados.

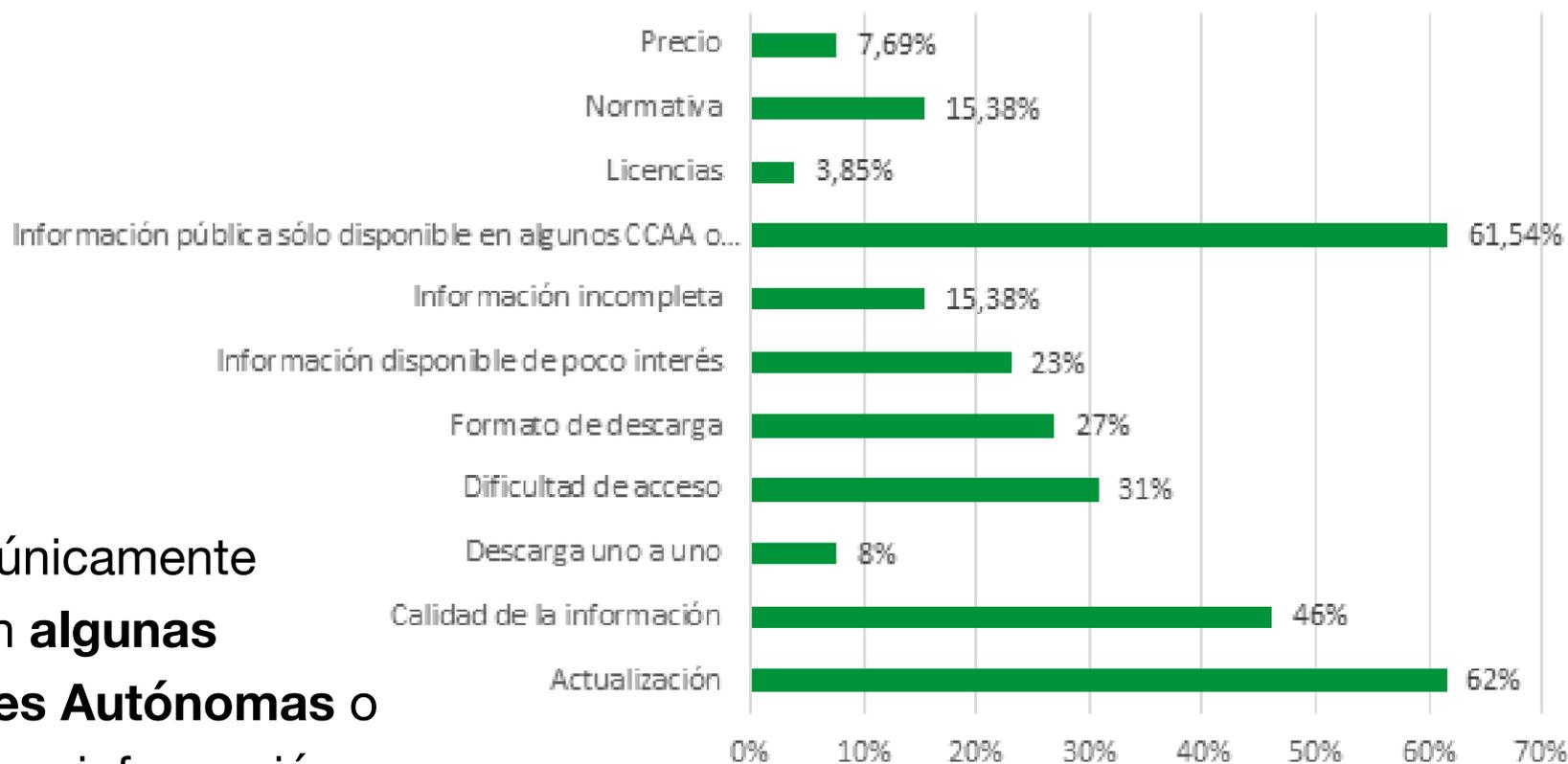
[ Uso de fuentes de datos ]

Ambas Público Privado



Fuente: Aseedie, 2020

# Informe Asedie. Barreras reutilización

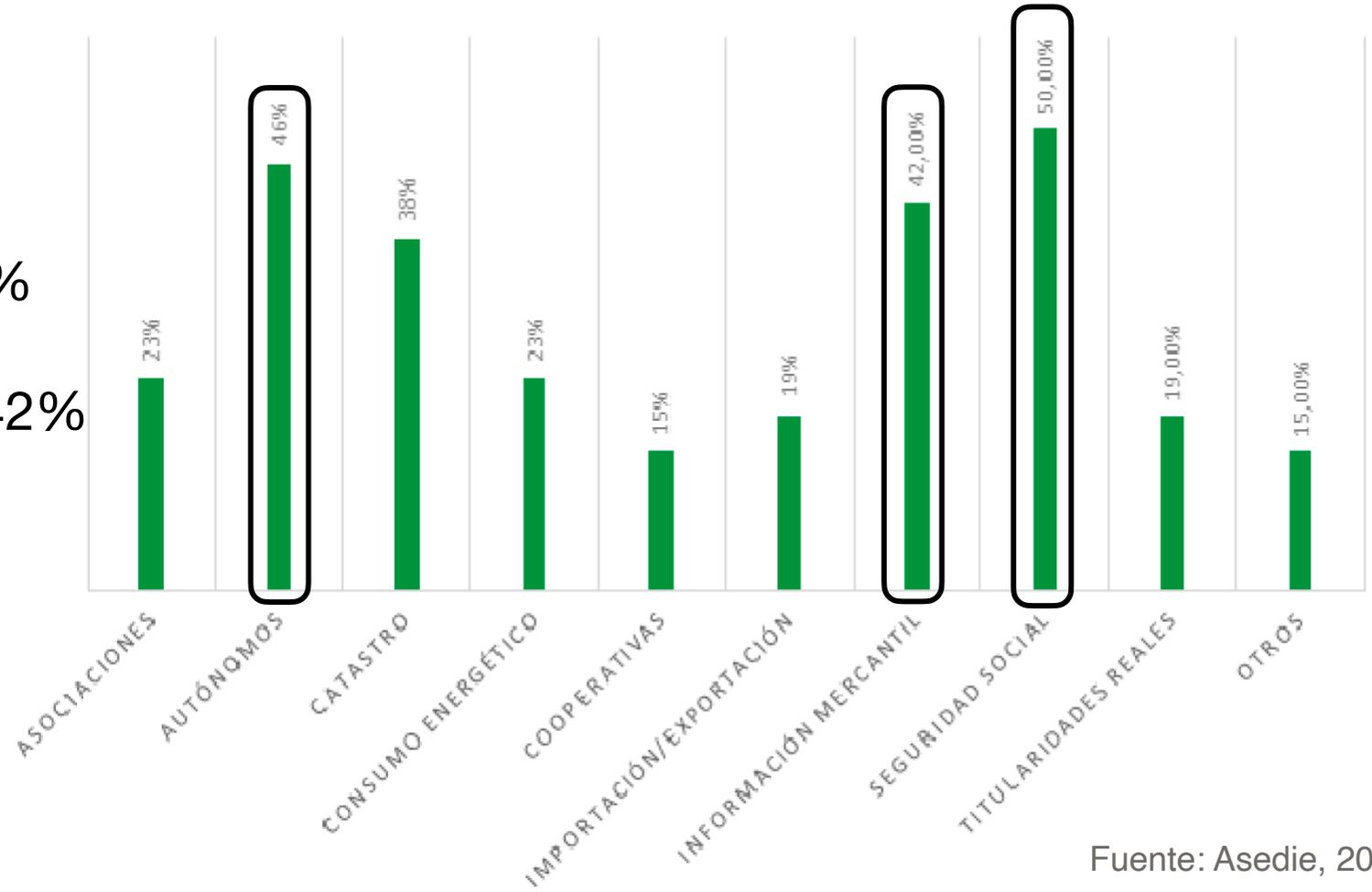


Información únicamente disponible en **algunas Comunidades Autónomas** o Ayuntamientos, información **desactualizada** e información de **baja calidad**.

Fuente: Asedie, 2020

# Informe Asedie. Datos más beneficiosos

Seguridad Social 50%  
Autónomos 46%  
Información Mercantil 42%



# Informe Asedie. Iniciativas que fomentan la apertura para las CCAA

## [ Iniciativas que fomentan la apertura ]



Las autonomías reconocen la necesidad de abrir mayor número de bases de datos, y consideran como factores de éxito:

- 12 la difusión de casos de éxito y la publicidad ayudan a movilizar a los equipos para abrir datos
- **La formación** 10
- La asistencia técnica 8

# Informe Asedie. Barreras para abrir

Según las CCAA :

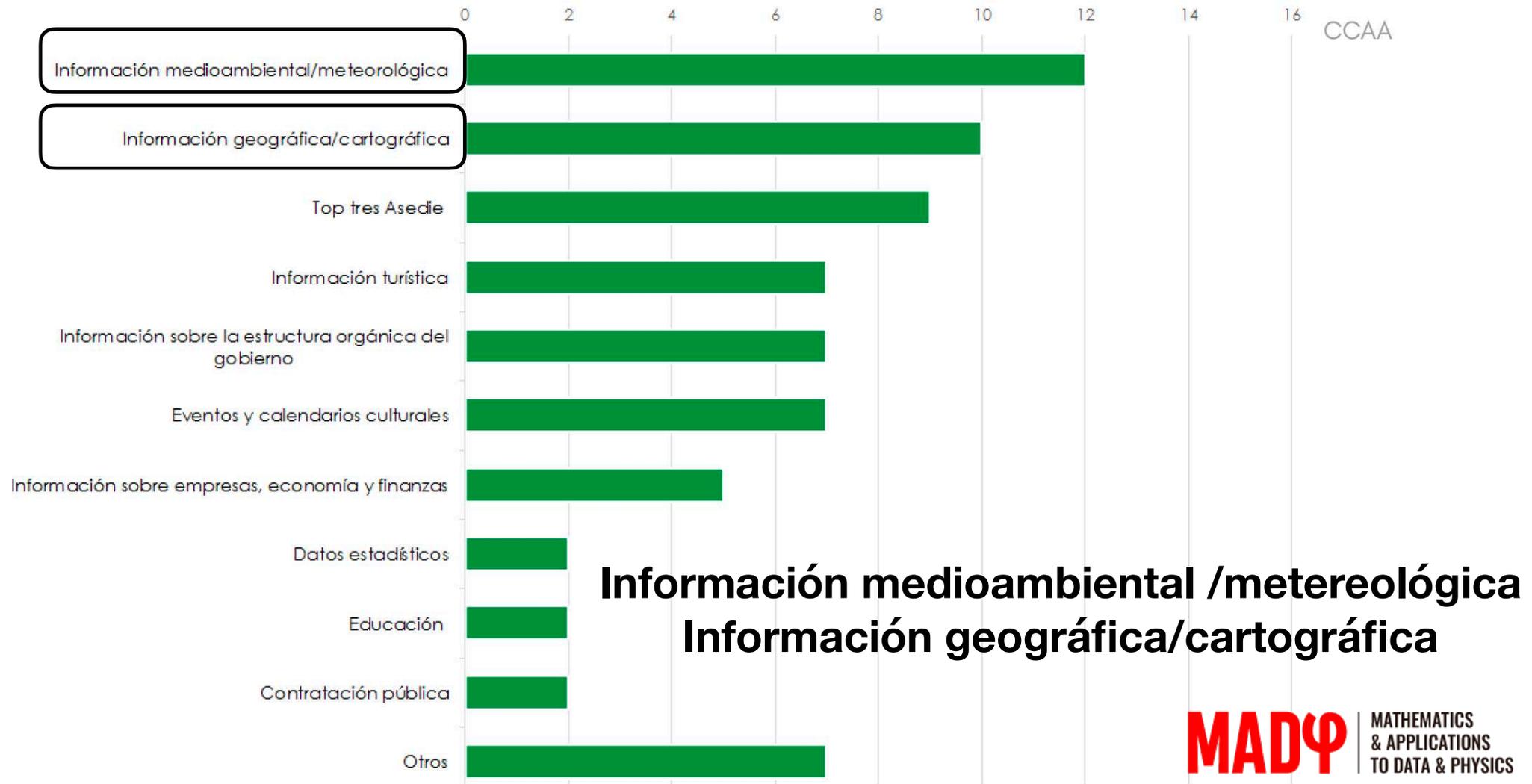
- **Tecnología:** 11 de 17 indican que “los sistemas de información se crean y se mantienen sin pensar que los datos se deben abrir, lo que implica que se deben rediseñar o aplicar costosas transformaciones para que lo datos sean reutilizables”.
- **Normativa:** 10 de las 17.
- **Cuestiones financieras:** 6 de 17.

*“existe la necesidad de un componente cultural que ayude a las organizaciones a entender el valor de la apertura y a los equipos a trabajar alineados para conseguirlo”*

# Informe Asedie. Barreras para abrir

*Se indica además el NIF como identificador único y dato prioritario a la hora de identificar y reutilizar los conjuntos de datos.*

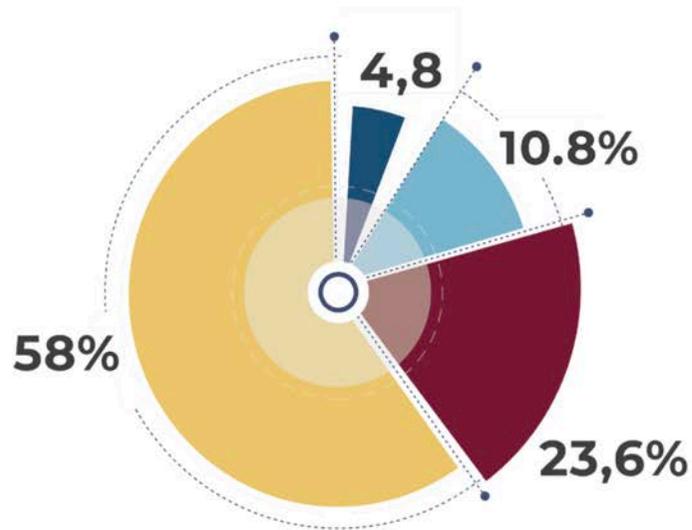
# Informe Asedie. Información de interés para la ciudadanía



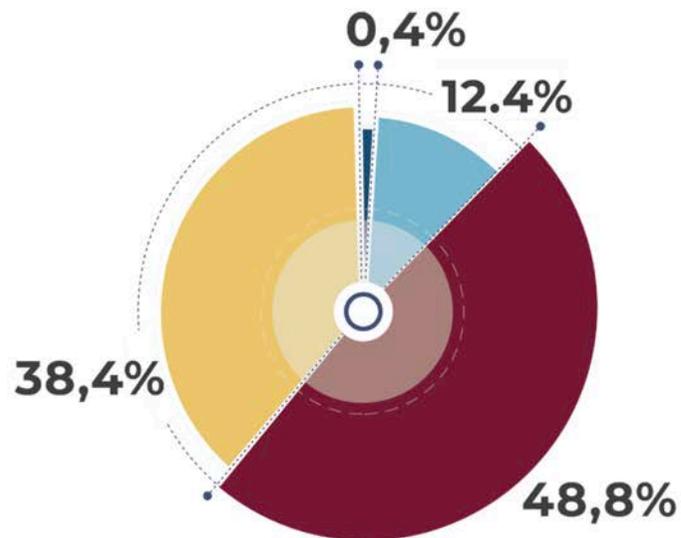
# Informe Asedie. No interesa a la ciudadanía

- **Información turística, o la estructura orgánica del gobierno, en 7 de las 16 CCAA**
- **Información sobre eventos o cultura, la información sobre empresas, economía o finanzas, en 5 de las 16**
- **Datos estadísticos e información sobre contrataciones públicas.**
- **Otros son el transporte, y los servicios de información del estado de las playas.**

# Reutilización de los datos abiertos en España. Abella et.al.



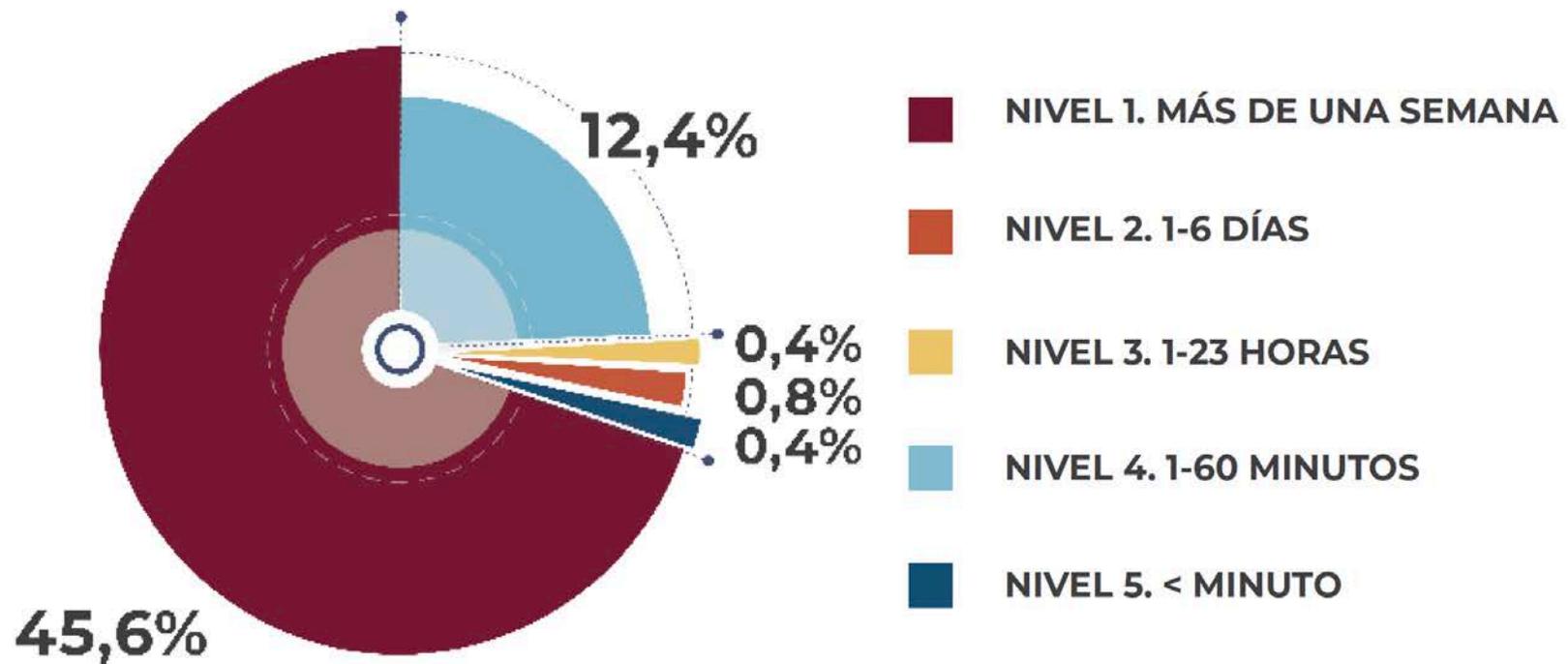
- NIVEL 1. ESTÁNDAR CERRADO NO REUTILIZABLE
- NIVEL 2. ESTÁNDAR CERRADO REUTILIZABLE Y ESTÁNDAR
- NIVEL 3. ESTÁNDAR ABIERTO
- NIVEL 4. ESTÁNDAR ABIERTO CON METADATA



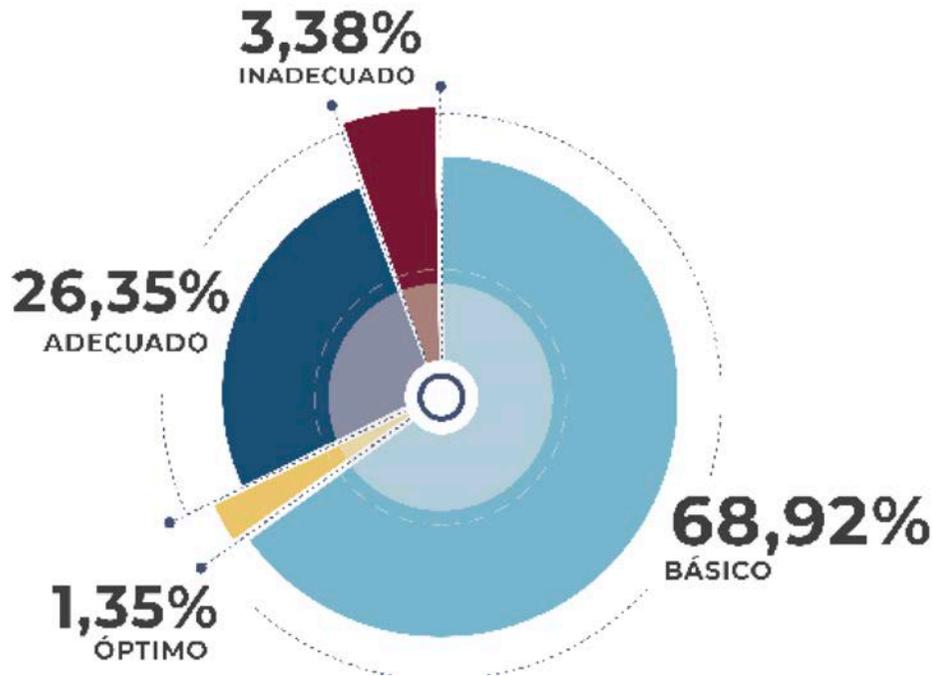
- NIVEL 1. NO ACCESO WEB O CONSULTA MANUAL
- NIVEL 2. ACCESO POR URL CON REGISTRO O CON INTERACCIÓN WEB
- NIVEL 3. ACCESO A TRAVÉS DE UNA URL ÚNICA A TODOS LOS DATOS DEL JUEGO DE DATOS
- NIVEL 4. ACCESO VÍA API O LENGUAJE DE INTERROGACIÓN

# Reutilización de los datos abiertos en España. Abella et.al.

## Nivel de actualización



# Reutilización de los datos abiertos en España. Abella et.al.



## Reusabilidad

uno de cada ocho responsables de portales no conocen a los reutilizadores de los datos que publican.

En la opinión de los responsables de los portales, se obtiene que **los ciudadanos no suelen ser reutilizadores** habituales de estos datos.

# Reutilizadores

- **Los ciudadanos** no suelen ser reutilizadores habituales de estos datos, en 2019 sólo un 19,6%
- **Reutilizadores profesionales sin ánimo de lucro** (ONG, fundaciones, desarrolladores individuales sin ánimo de lucro y otras iniciativas sociales), son del 31,7%
- **Investigadores y personal académico** 34,2%
- **Reutilizadores profesionales con ánimo de lucro** (infomediarios y desarrolladores individuales con ánimo de lucro) 36,6%
- **La propia organización** son los agentes que con más frecuencia utilizan los datos abiertos Son usuarios frecuentes o habituales **56,1%**

# Reutilización de los datos abiertos en España. Abella et.al.

## Ámbito de actuación de los reutilizadores de datos abiertos,

ámbito local 39,1%  
 autonómico 31,7%  
 nacional 29,3%  
 europeo 12,2%

Escala	Local		Autonómico		Nacional		Europeo		Otros	
	2017	2019	2017	2019	2017	2019	2017	2019	2017	2019
Nunca	3,7%	2,4%	0%	2,4%	7,4%	2,4%	11,1%	4,9%	7,4%	2,4%
Casi nunca	11,1%	7,3%	11,1%	4,9%	4%	2,4%	11,1%	4,9%	7,4%	4,9%
A veces	18,5%	14,6%		14,6%	14,8%	19,5%	7,4%	17,1%	0%	4,9%
Frecuente	11,1%	9,8%	7,4%	29,3%	14,8%	22%	3,7%	7,3%	3,7%	2,4%
Siempre	14,8%	29,3%	11,1%	2,4%	0%	7,3%	3,7%	4,9%	7,4%	4,9%
NS/NC	40,7%	36,6%	48,1%	46,3%	55,5%	46,3%	62,9%	61%	74,1%	80,5%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

# Servicios creados

## Temática

Las categorías más frecuentes son: de cultura y ocio (14,5%), seguida de medio ambiente (12,9%), transporte (12,9%) y urbanismo e infraestructuras (9,7%).

## Sostenibilidad

30,6% en 2019 de las aplicaciones no tenían ningún mecanismo de sostenibilidad (ni el patrocinio de una institución ni mecanismos de generación de ingresos, bien directos, coste o indirectos, pagos freemium o publicidad). Sólo 19,4% de los servicios presenta un modelo de negocio.

# Servicios creados

Autores de los servicios

Las organizaciones públicas (tanto la que publica los datos como otras relacionadas) con el 61,2%

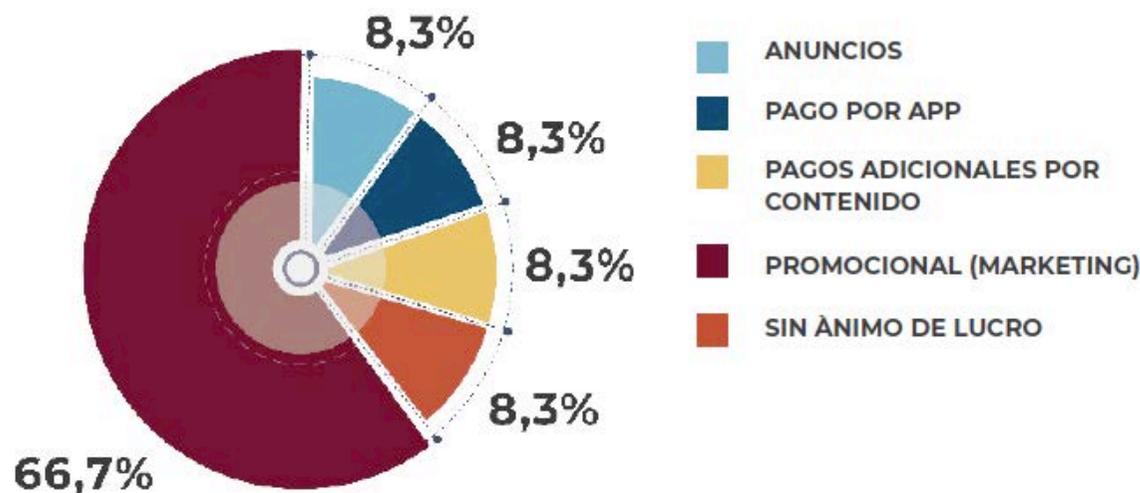
Las AA. PP. las que más servicios basados en datos abiertos generan. Por otra parte, casi un 25% son organizaciones con ánimo de lucro.

# Servicios creados

Otras características

Servicios geolocalizados 74,2%,  
Sólo el 25,8%, en 2019 son servicios con  
información en tiempo real.

Para los casos que tienen desarrollado un modelo de negocio



En 2019, el 83,3% de los servicios con un modelo de negocio están geolocalizados y sólo el 58,3% de ellos es en tiempo real

De las aplicaciones con modelo de negocio, en 2019, destacan las de salud (25%), economía, medio ambiente y transporte (16,7%)

# Datos abiertos de investigación



# Datos abiertos de investigación

Forma **parte de las estrategias que abogan por la Ciencia Abierta**, apoyadas tanto por instituciones de investigación como por responsables políticos.

Se sustenta en que el **acceso abierto** debe de estar dirigido **no solo a las publicaciones científicas sino también a los datos de investigación**.

Se basa en la certeza de que los **datos tienen un enorme potencial** que debe ser aprovechado por el bien de la ciencia y de la sociedad en general.

# Datos abiertos de investigación

“datos de investigación [es] todo aquel material que ha sido **registrado durante la investigación, reconocido por la comunidad científica** y que sirve para **certificar** los resultados de la investigación que se realiza. [...] debe provenir de una **fente única** y deben ser **difíciles o imposibles de obtener de nuevo**” (Torres-Salinas; Robinson-García; Cabezas-Clavijo, 2012)

Necesarios para **validar** los resultados de una investigación: “...factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated.” OECD (2007)

# Valor de los datos abiertos de investigación

Los **datos de investigación son clave para el avance económico y la innovación**, tal y como señalaba el comisario europeo Moedas en su discurso «Open Innovation, Open Science, Open to the World» en la apertura de la conferencia «Opening up to an ERA of Innovation» celebrada en junio de 2015 (Moedas, 2015)



[https://es.wikipedia.org/wiki/Carlos\\_Moedas](https://es.wikipedia.org/wiki/Carlos_Moedas)

# European Open Science Cloud

En abril de 2016 se lanzó oficialmente la European Open Science Cloud (EOSC), un entorno virtual común para que todos los investigadores europeos puedan almacenar, gestionar, analizar y reutilizar datos para la investigación, la innovación y la educación



**EUROPEAN OPEN  
SCIENCE CLOUD**

**Objetivo:** crear un entorno abierto de confianza en el que la comunidad científica pueda compartir y reutilizar datos y resultados de investigación

# European Open Science Cloud



Los datos de investigación despiertan gran interés por su potencial aprovechamiento y reutilización en sus distintos ámbitos: investigación, industria y sociedad.

## Beneficios:

- Desde el punto de vista de las **instituciones**: proporciona transparencia en los procesos de obtención o generación de datos.
- Entre los **investigadores** promueve la colaboración entre grupos afines, y evita duplicidades, permite validar estudios.
- Para la **sociedad**, mejora la confianza en el sistema científico y supone un ejercicio de transparencia, rendición de cuentas y de responsabilidad en el uso de la inversión en I+D.

# European Open Science Cloud



Fomenta la **validación** de los métodos de investigación.



**Evita la repetición** de costosos proyectos que utilizan las mismas intervenciones que se utilizaron previamente.



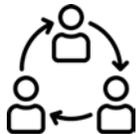
Aumenta la **visibilidad** y el impacto de la investigación.



Aporta **recursos** para la educación y la formación



**Agiliza** la producción científica



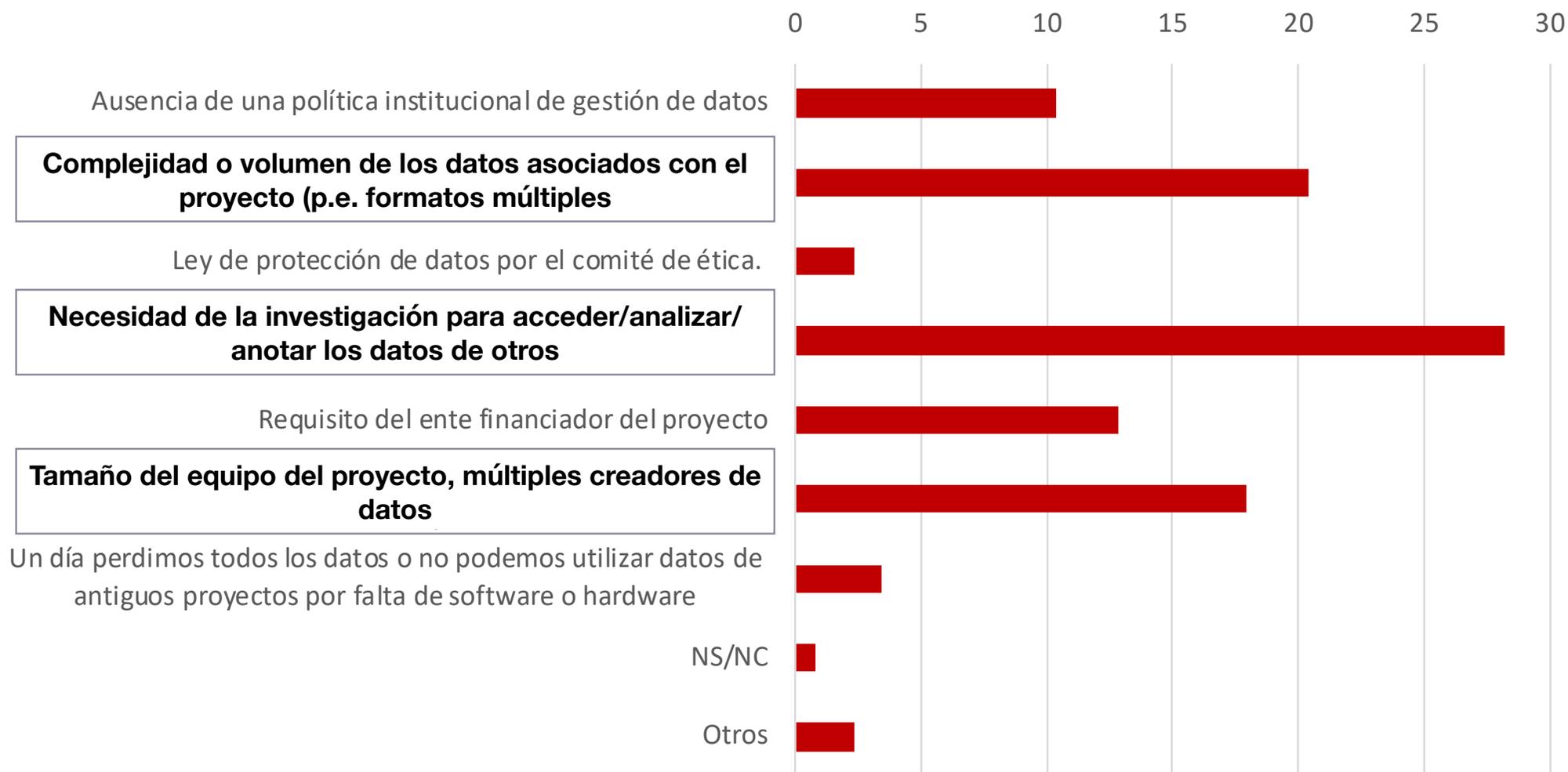
Fomenta la **participación y colaboración**

¿Qué conocen los investigadores?

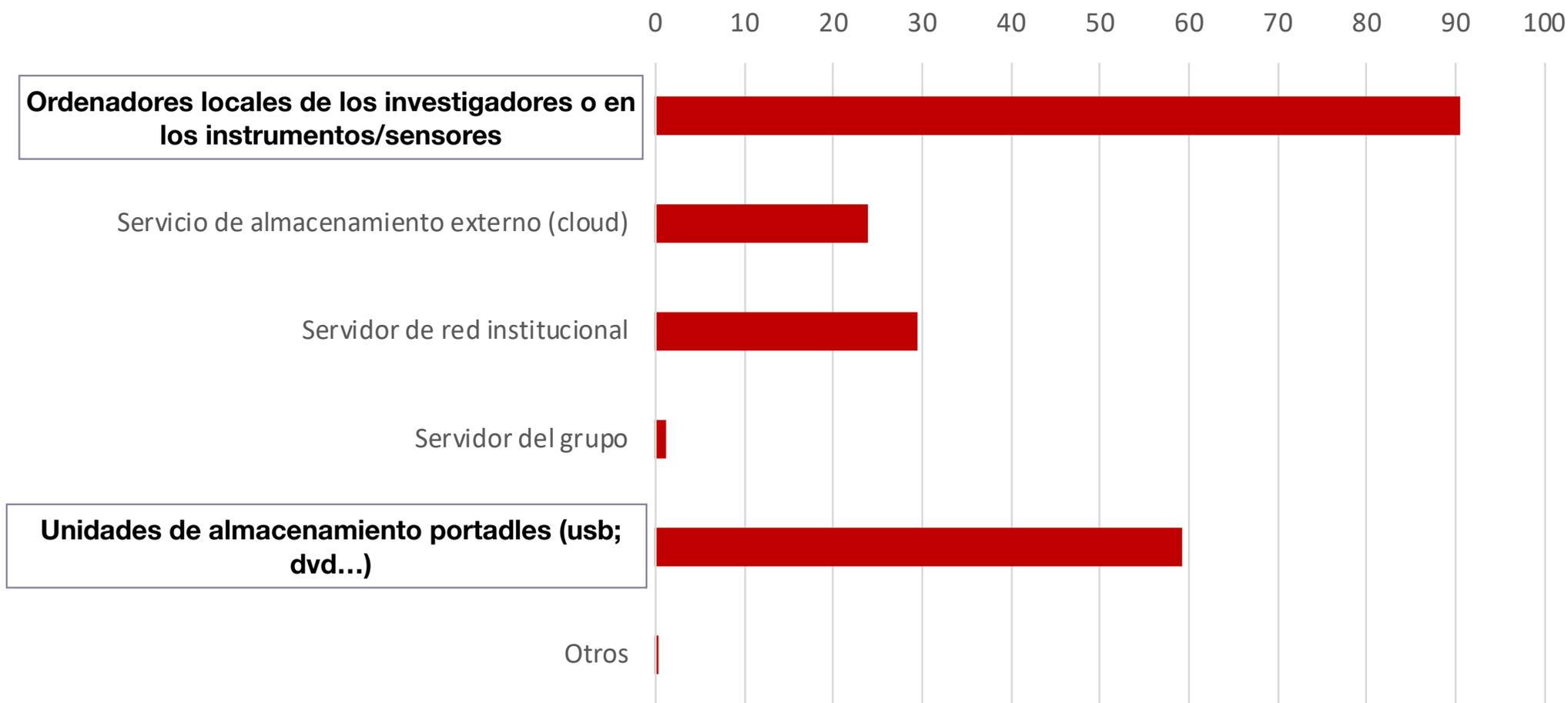


Foto de [Gustavo Fring](#) en [Pexels](#)

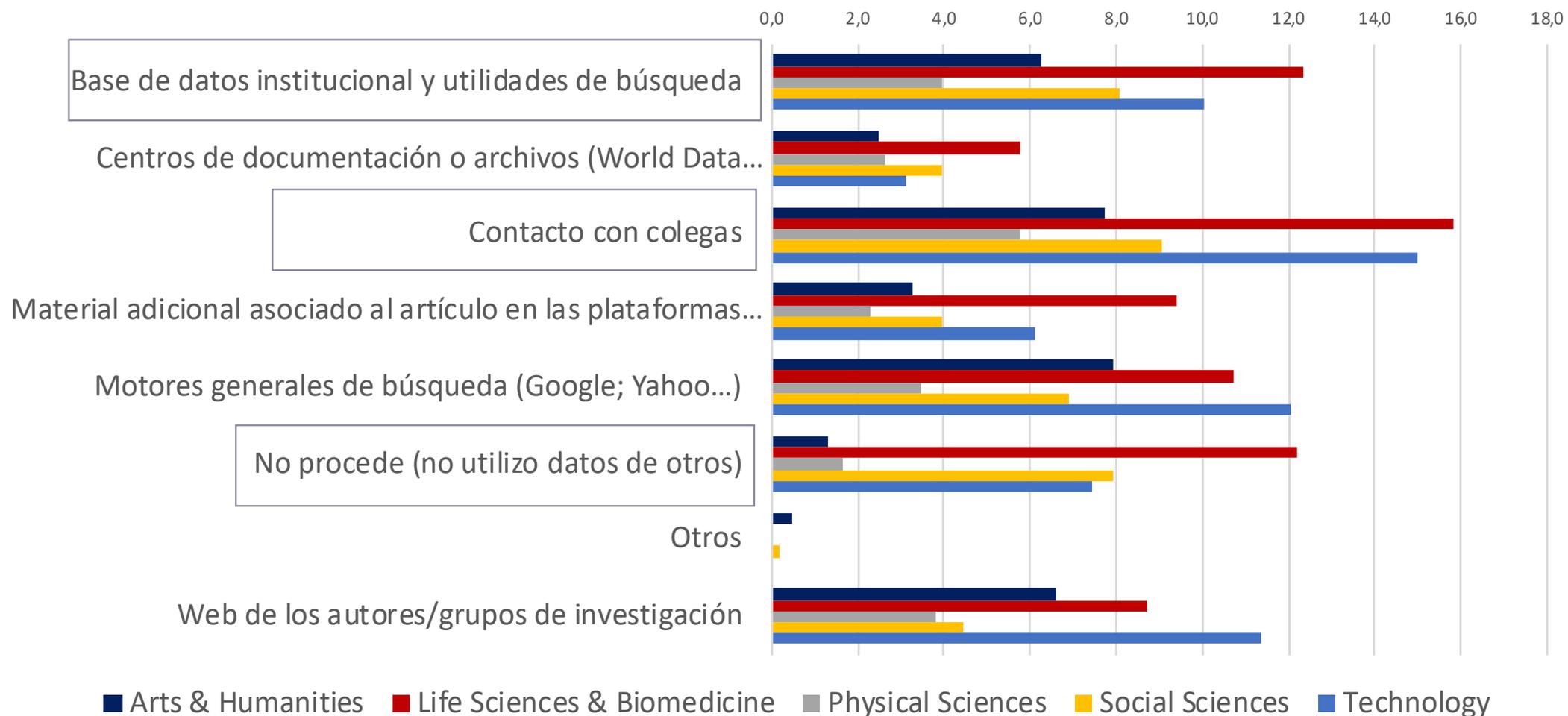
## Motivos por los que desarrollar una política de gestión de datos



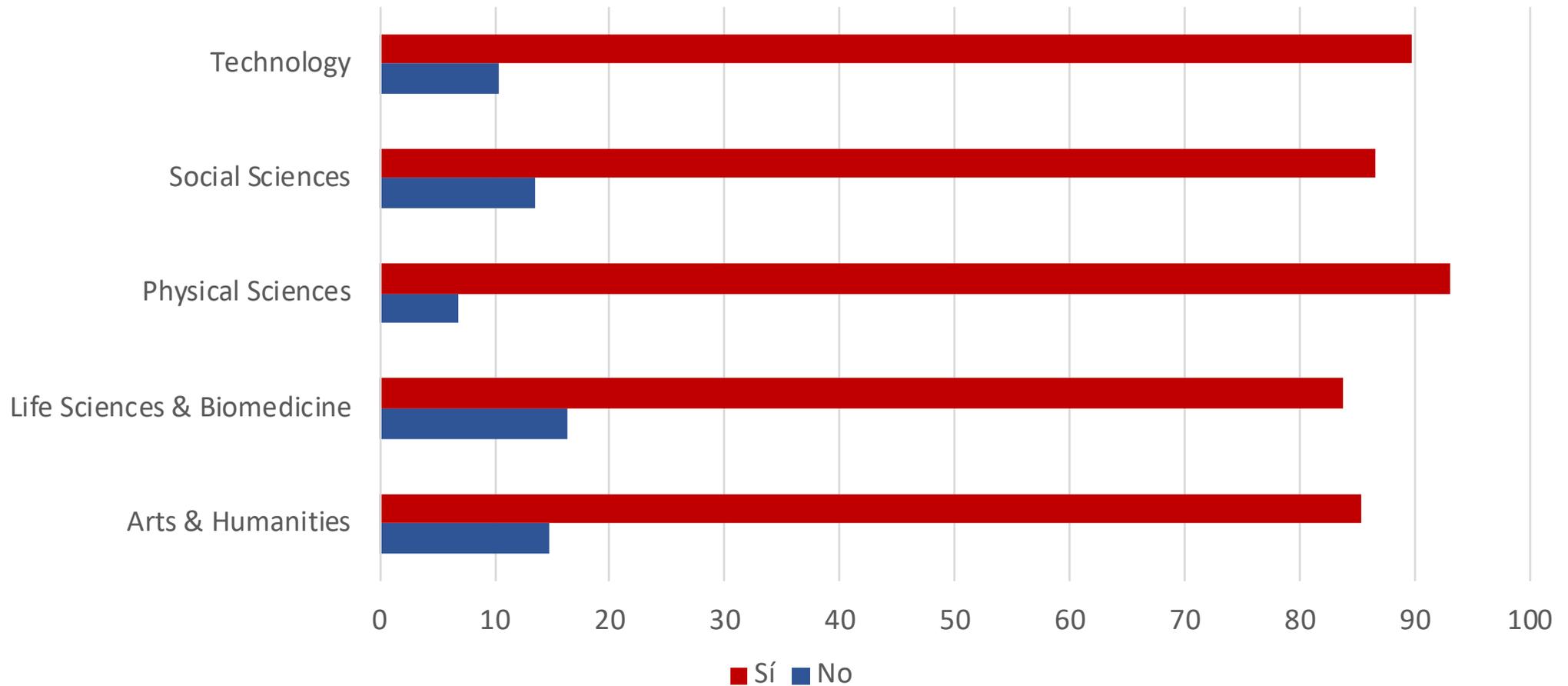
## Donde suele almacenar los datos de investigación



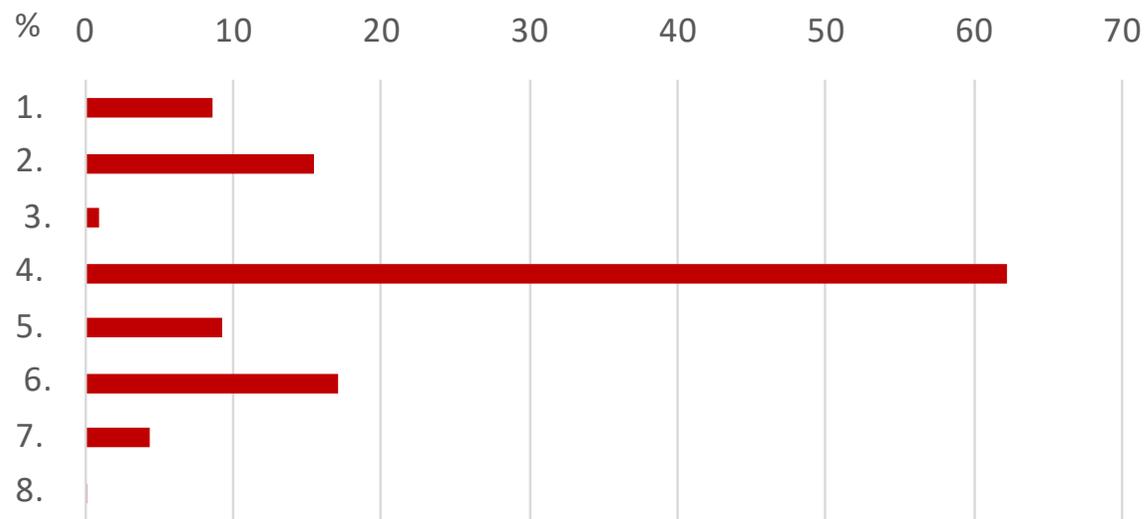
## Medios utilizados para localizar y acceder a los datos de otros investigadores



## ¿Le gustaría utilizar datos de investigación de otros investigadores?



## En relación a la disponibilidad de los datos



**1.** Mis datos están disponibles en abierto para mi disciplina científica

**2.** Mis datos están disponibles en abierto para todo el mundo

**3.** Mis datos están disponibles mediante el pago de una tasa

**4.** Mis datos están disponibles para mi grupo de investigación y/o colegas colaboradores en la investigación

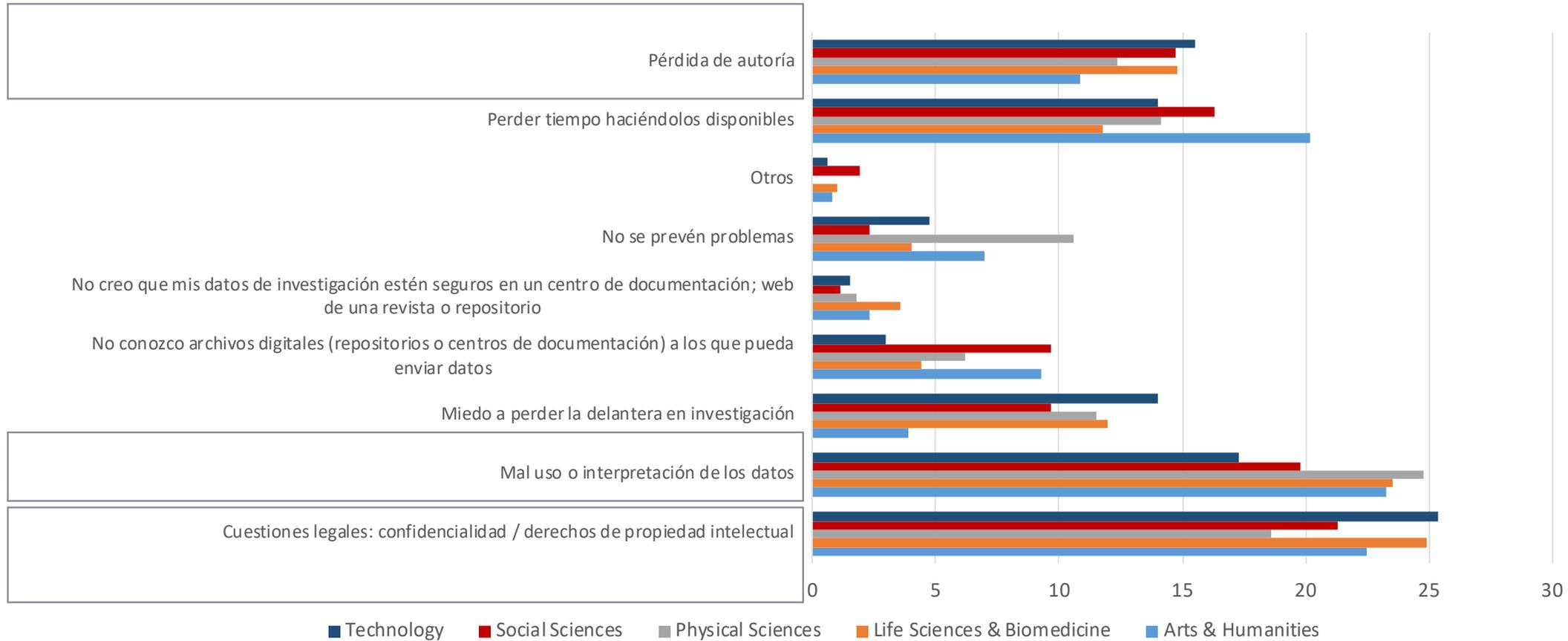
**5.** Mis datos podrían estar disponibles con los cambios apropiados (por ejemplo: datos clínicos anónimos o con la obligación de citarlos)

**6.** No comparto mis datos pero me gustaría hacerlo en un futuro

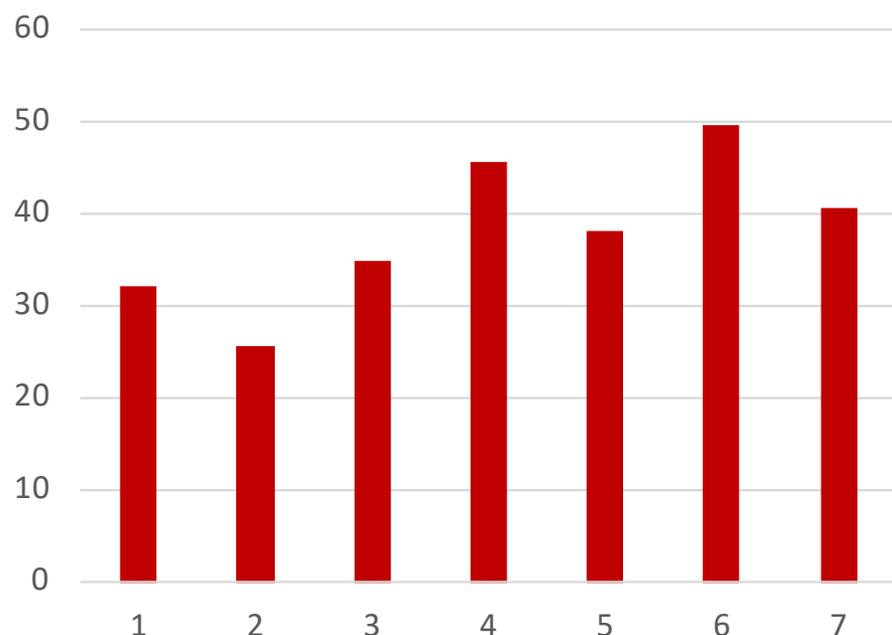
**7.** No comparto mis datos y no quiero hacerlo en el futuro

**8.** Otros

# ¿Qué le preocupa a la hora de compartir?

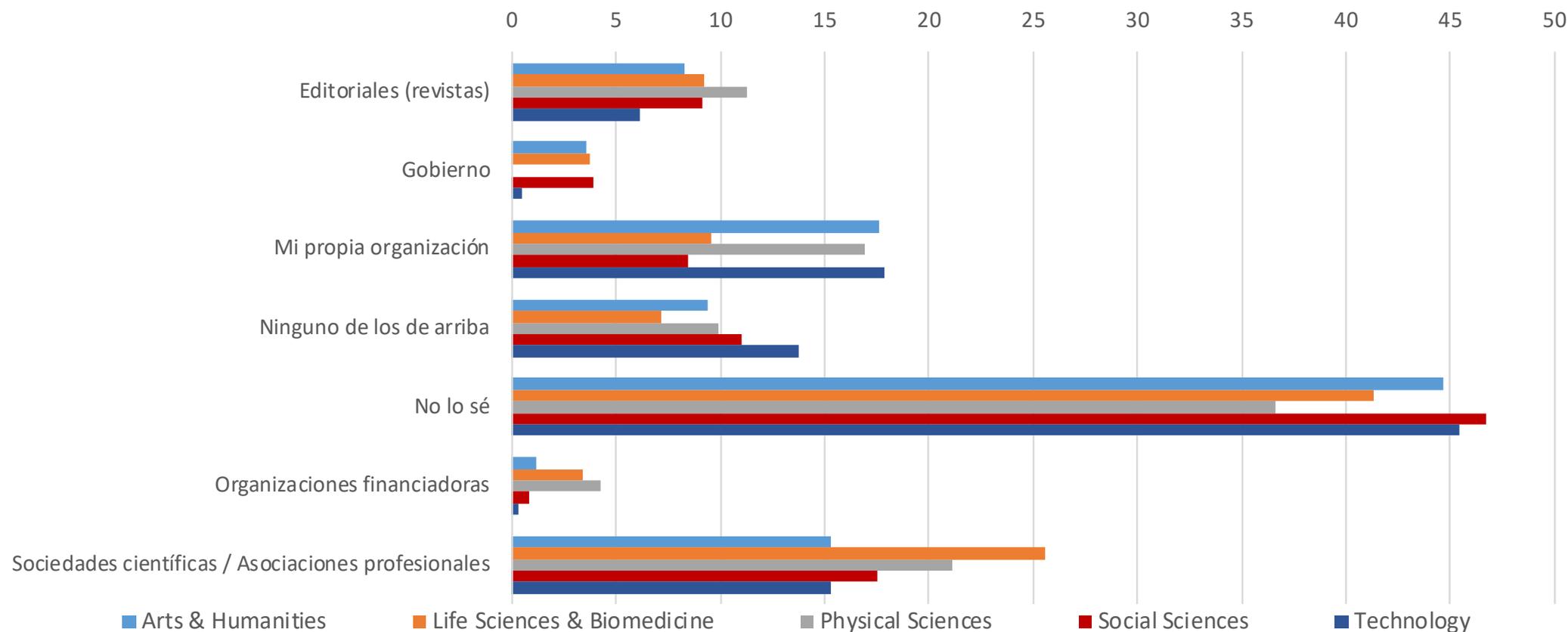


## Principales amenazas a compartir

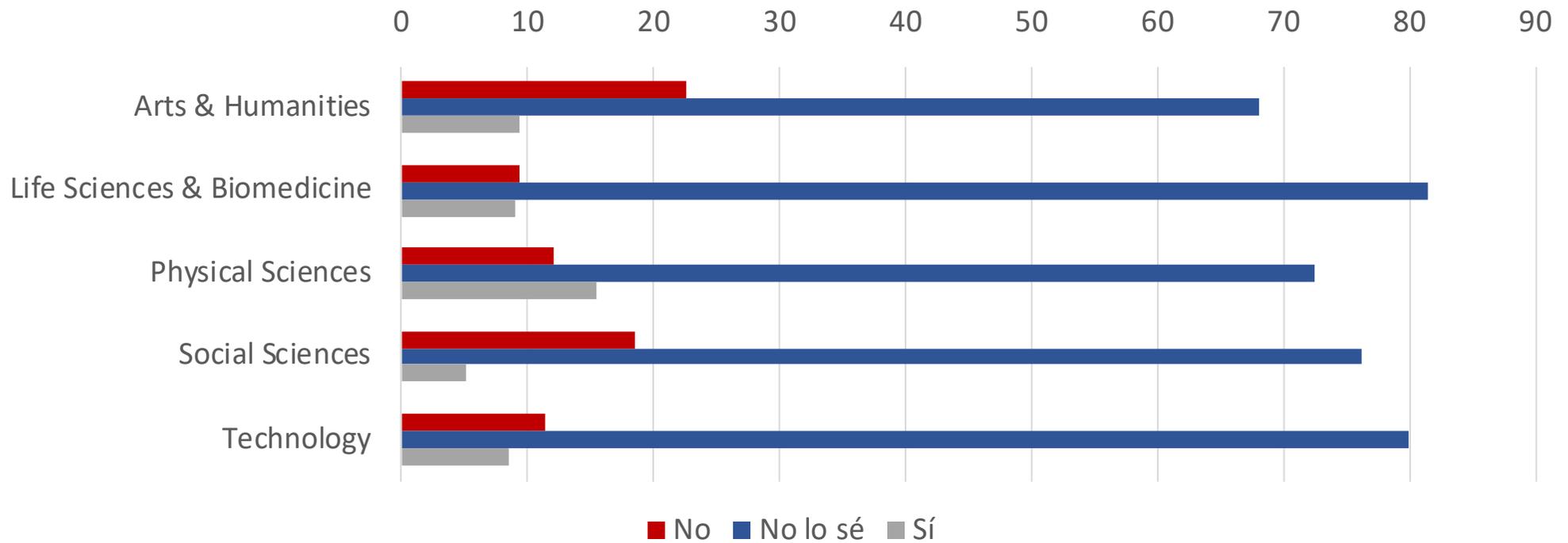


- 1** Los usuarios pueden ser incapaces de comprender o utilizar los datos (semántica, formatos o algoritmos implicados)
- 2** La falta de software, hardware o soporte sostenibles de los ordenadores
- 3** El origen y autenticidad de los datos puede ser incierto
- 4** Las restricciones de acceso y uso (propiedad intelectual)
- 5** Pérdida de capacidad para localizar los datos
- 6** La custodia actual de los datos, en una organización o proyecto, puede dejar de existir en algún punto en el futuro
- 7** Aquellos en quienes confiamos para cuidar los datos, pueden fallarnos

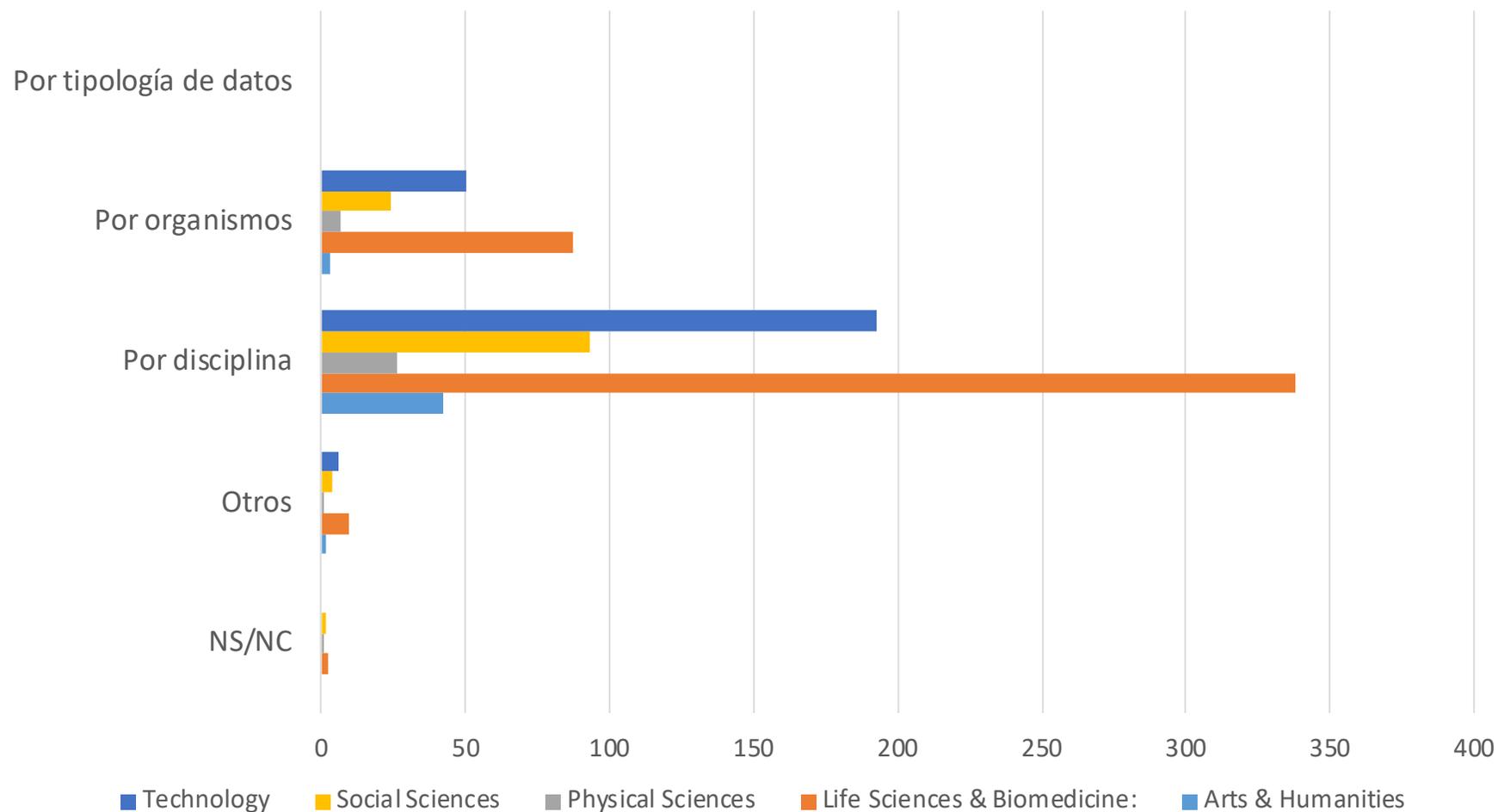
## En su disciplina, ¿Quién proporciona guías o recomendaciones para preservar datos?



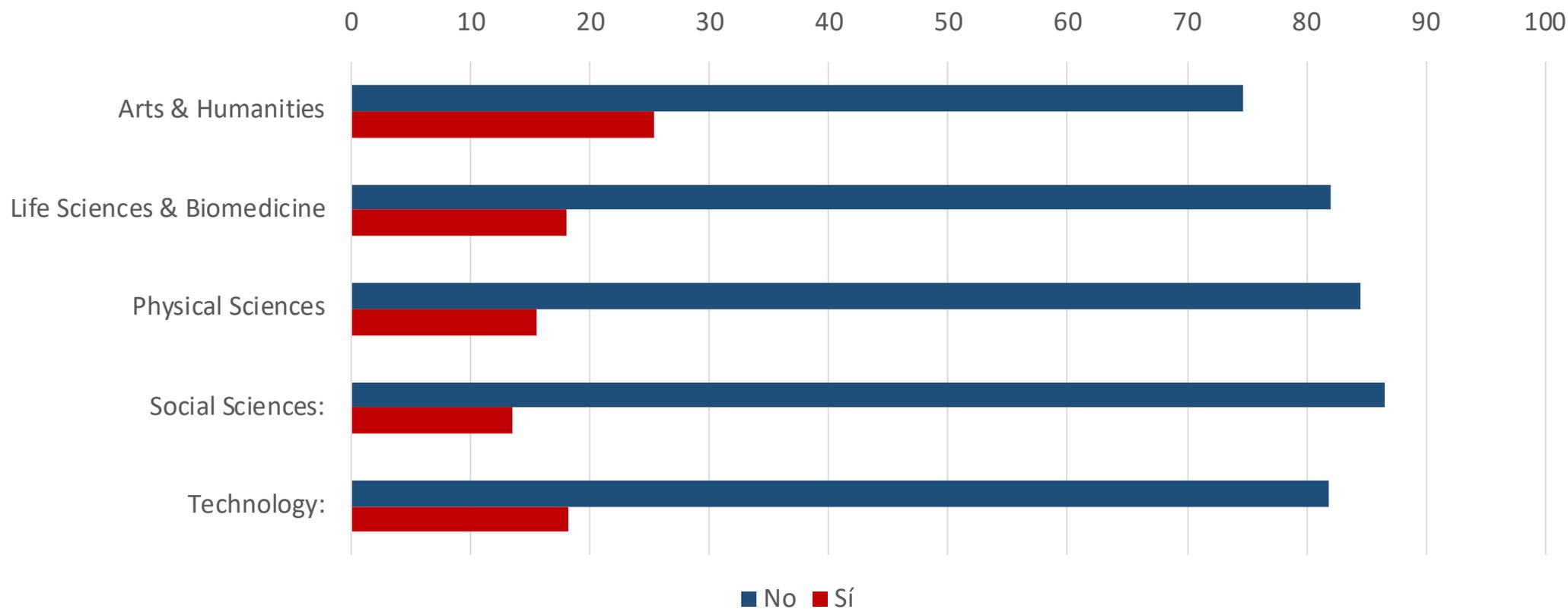
## Existe en su disciplina alguna instalación de datos digitales que pueda utilizar



## Cómo cree que debería ser esta infraestructura



## Su organización le ayuda con la preservación de datos



# Conclusiones. Sobre los datos

- Las temáticas de los datos han variado. Ahora los datos más publicados son los de sector público
- Existe una divergencia entre lo publicado y los servicios generados.
- El número de juegos de datos que no pueden reutilizarse por su formato se ha mantenido constante.

# Conclusiones. Sobre los gestores.

- Ha aumentado el porcentaje de conocimiento de los reutilizadores de datos.
- La propia organización es el que más reutiliza los datos por encima de académicos y entidades con ánimo de lucro.
- Existe un amplio desconocimiento de segmentos de usuarios que son reutilizadores.
- Los sectores económicos de transporte y almacenamiento y la propia Administración Pública son los sectores con mayor reutilización conocida..
- Las actividades de promoción del uso de los portales han experimentado un ligero descenso en las principales opciones (concursos, actos propios, actos externos, reuniones con reutilizadores).

## ¿Qué hacen los investigadores frente a sus datos?

- Suelen utilizar ordenadores locales o memorias digitales portátiles como sistemas de almacenamiento y depósito.
- A una amplia mayoría les gustaría utilizar los datos de investigación recogidos por otros investigadores.
- La principal amenaza a sus datos son es la restricción de acceso y uso relacionada con la propiedad intelectual.
- El principal motivo para desarrollar una política de datos fue por la necesidad de la propia investigación



# ¡GRACIAS!



DATAUSE: datos y agricultura. PID2019-105708RB-C21  
Proyectos I+D+i «Retos Investigación» (MICINN/FONDOS  
FEDER)

