



Radiografía del Espacio de Datos de Turismo en España (Hub Gaia-X España)



Financiado por
la Unión Europea
NextGenerationEU



VICEPRESIDENCIA
PRIMERA DEL GOBIERNO
MINISTERIO
DE ASUNTOS ECONÓMICOS
Y TRANSFORMACIÓN DIGITAL

SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN E
INTELIGENCIA ARTIFICIAL



Plan de
Recuperación,
Transformación
y Resiliencia

España | digital

20
26





“

Los espacios federados de datos son instrumentos clave en el desarrollo de una Economía del Dato capilar y transformativa. El Turismo es una industria entrelazada con muchos otros ámbitos, y por ello encontrará grandes beneficios en los ecosistemas soberanos de compartición y explotación de datos. Aquellos que sean capaces de relacionar diferentes contextos entre sí, generarán una inteligencia de mercado mucho más dinámica y robusta.

”

Alberto Palomo

Chief Data Officer (Gobierno de España)

“

Nos encontramos en un momento clave, en el que el liderazgo de los espacios de datos de turismo reposicionará a España como un referente mundial mejorando la competitividad de la industria turística.

”

Dolores Ordóñez

Directora (AnySolution)

ÍNDICE

1. Contexto	3
2. La necesidad y el valor añadido de un espacio de datos de turismo	5
3. Retos y oportunidades del sector turístico	6
Del Destino al Turista	7
4. Casos de uso y plataformas de datos más comunes en la actualidad	10
Con datos abiertos	10
Con datos privados y soberanos	11
5. Gobernanza de la compartición y explotación de datos B2B en el sector turístico	13
Retos para la escalabilidad en la compartición de datos privados para el Turismo	13
Principios de diseño de espacios de datos industriales turísticos	16
Correspondencia entre criterios de gobernanza y arquitecturas técnicas	23
6. Necesidades técnicas	24
Seguridad	24
Soberanía	25
Interoperabilidad	26
Interoperabilidad semántica	26
Interoperabilidad de sistemas	27
Interoperabilidad de infraestructuras	27
Convergencia para la creación de valor	29
Building blocks para dataspace	31
7. Conclusiones y siguientes pasos	34
ANEXO – Metodología general para la construcción de espacios de datos	36



1

Contexto

Con el fin de promover un entorno propicio para la creación de **espacios de datos sectoriales**, se ha impulsado la creación del **Hub español de Gaia-X** ([Gaia-X Hub Spain](#)), que busca que los diferentes sectores de la economía creen comunidad en torno al dato. El objetivo del hub es que sirva para desarrollar e implantar soluciones innovadoras basadas en datos e inteligencia artificial, y que contribuya a impulsar la competitividad de las empresas, y la sociedad en general. Estos espacios son además tractores para generar beneficios a lo largo de toda la cadena transformativa del dato, y se presta especial atención a que en la misma estén también representadas PYMEs y microPYMEs. El hub se genera a partir del impulso de la estrategia [España Digital 2026](#) y con el [Plan de Recuperación, Transformación y Resiliencia](#) como hoja de ruta para la transformación digital de España.

El hub español de Gaia-X cuenta, entre otros, con un Grupo de Trabajo (GT) del espacio de datos de turismo, que el [pasado 19 de enero](#), dentro del contexto de la feria internacional **FITUR 2022**, llevó a cabo un **Taller de trabajo (Workshop)** con el objetivo de avanzar en la creación de dicho dataspace. Este primer taller colaborativo se centró **en el uso de los datos en la industria del Turismo**, uno de los sectores productivos y clave para España. En el mismo, se analizaron los retos y oportunidades en materia de gobernanza de dicho espacio de datos, las tecnologías y arquitecturas disponibles para facilitar la compartición de datos, así como casos de uso tractores para el sector. El taller contó con la participación de diferentes representantes de la política, de oficinas técnicas gubernamentales, de empresas a lo largo de toda la cadena de valor del Turismo, universidades e institutos tecnológicos, así como integradores del mundo TIC.

Junto con este workshop, se han realizado también consultas al Grupo de Trabajo de Turismo, para producir **la primera radiografía de dicho espacio**

de datos en España. El objetivo de este documento es resaltar en qué momento de desarrollo se encuentra el diseño de dicho dataspace, las diferentes oportunidades para el sector, tanto a nivel estatal como europeo, y qué retos deben superarse para conseguir el despliegue del mismo.



Como industria transversal, el turismo es una parte integral e importante de la economía europea, directamente relacionada con otros sectores como la movilidad, la logística, la salud, la agricultura, la cultura, los medios de comunicación, la automoción, la alimentación y bebidas, y otras.

Las **interdependencias del turismo con otras industrias** pudieron observarse de forma evidente durante la pandemia del COVID-19, cuando muchos de los sectores anteriormente mencionados se vieron directamente afectados por las restricciones de viaje y los cierres de empresas turísticas. Por el lado positivo, esto demuestra también las enormes posibilidades que existen en torno a la creación de valor por vía intersectorial, así como la necesidad de intercambio de datos eficiente en todas las direcciones. Diferentes aplicaciones pueden así acceder a los datos y combinarlos, para proporcionar valor añadido a los usuarios y a los responsables de la toma de decisiones de los diferentes sectores.

El sector turístico se ha visto directamente afectado por la pandemia del COVID 19 debido, entre otras cosas, a las restricciones de viaje y los cierres de empresas turísticas

Además, la creciente necesidad y demanda de datos en el sector turístico -y sus mercados intersectoriales interconectados- requiere un esfuerzo coordinado para acelerar el desarrollo y el despliegue de un Espacio de Datos de Turismo. Entendemos que todos los actores en el mercado podrán beneficiarse de valor añadido que proporciona un mejor acceso a los

datos de turismo, gracias la interconexión de las diferentes tecnologías relacionadas con los datos, la participación de todo el ecosistema, la generación de un sistema de gobernanza multinivel entre las múltiples partes interesadas, permitiendo la privacidad, seguridad, procesamiento, análisis de extremo a extremo y la inclusión.

La creciente necesidad y demanda de datos en el sector turístico requiere un esfuerzo coordinado para acelerar el desarrollo y el despliegue de un Espacio de Datos de Turismo

Estas estrategias de extremo a extremo deben crear "espacios" inteligentes que utilicen diferentes niveles de inteligencia entre los actores heterogéneos, así como una seguridad basada en el contexto y la confianza para compartir datos. Este entorno permitirá a los actores interactuar sin problemas entre ellos, así como con otros espacios de datos, y proporcionará una poderosa base para las innovaciones impulsadas por los datos, y un mercado digital europeo independiente en el futuro. Así mismo, otra cuestión importante es la necesidad de **generar valor para el propio sector turístico**, que pueda así resolver problemas existentes actualmente, sin centrarse exclusivamente en la cuestión de las ventas o los ingresos. Aportar valor a los procesos de la industria turística aparece por tanto como una cuestión principal en la que trabajar.

3 Retos y oportunidades del sector turístico

En este sentido, los retos relacionados con la creación de un espacio común de datos turísticos guardan similitudes con los de otras industrias, como pueden ser la **falta de confianza**, la **falta de modelos de datos compartidos**, o la **insuficiencia de normas de interoperabilidad** para un intercambio de datos eficiente entre las diferentes plataformas existentes y los actores de la cadena de valor (por ejemplo, operadores turísticos, el mundo HORECA, o las compañías aéreas).

Además, existen retos clave más específicos del Turismo, como la necesidad de combinar los espacios de datos B2B (compartición entre empresas y organizaciones privadas) con los C2B y G2B (compartición de usuarios a empresas, y sector público a empresas) a diferentes niveles: local, regional, nacional e internacional. Los datos proporcionados por las distintas partes interesadas presentan grandes **diferencias en cuanto a su nivel de granularidad y calidad**, y además utilizan diferentes formatos y una amplia gama de licencias, lo que da lugar a un panorama de datos desconexo, teniendo en cuenta que el turismo -por naturaleza- depende en gran medida del intercambio eficiente de datos.

Entre los retos relacionados con la creación de un espacio común de datos turísticos se identifican la falta de confianza, la falta de modelos de datos compartidos, o la insuficiencia de normas de interoperabilidad para un intercambio de datos eficiente



Otro de los principales retos del espacio de datos del turismo es incluir y comprometer a todo el ecosistema, teniendo en cuenta tanto a los proveedores de tecnología, por un lado, como las partes interesadas del propio sector turístico, por otro, incluyendo también a los visitantes y a la población local. En este sentido, para el correcto levantamiento de casos de uso en el sector turístico, se pone de relevancia que **las necesidades de la industria deben ser identificadas escuchando a sus participantes**, analizando casos de uso en destinos turísticos, así como las necesidades/ particularidades de cada uno de los actores que sobre dicho destino operan, así como entrevistando a los gestores con experiencia, para extraer características y especificidades que serán relevantes para el diseño de estos casos de uso.

La principal característica del turismo, que marca una diferencia real con otros espacios de datos, es que **el turismo se hace por y para las personas**, generando impactos tanto positivos como negativos en las comunidades locales y sus ecosistemas. Esto lleva a la necesidad de tener en cuenta las implicaciones relacionadas con **el turismo como actividad social, cultural y económica**, con un impacto visible en el medio ambiente.

Del Destino al Turista

Hasta el momento, se ha dado una gran importancia al **destino turístico** como parte relevante del proceso turístico, y este enfoque ha permitido el desarrollo de importantes plataformas turísticas (ver Sección 4). Estas plataformas ofrecen:

- Una herramienta que permite priorizar la intervención en el destino en función de los objetivos y requisitos de un *Destino Turístico Inteligente*.
- Un **aumento de la competitividad**, gracias al mejor aprovechamiento de sus recursos turísticos y a la identificación y creación de otros
- Una **mejora en la eficiencia** de los procesos de producción y comercialización.
- Un impulso al **desarrollo sostenible** del destino.
- Una **mejora de la calidad** de la estancia de los visitantes y de la calidad de vida de los residentes.
- Hacer de la **estrategia turística** la base para la dinamización económica del territorio garantizando sus efectos positivos en el largo plazo.
- Garantizar un servicio de **calidad turística**.

Sin embargo, se ha detectado la importancia de **ampliar esta orientación** incluyendo la **medición de la experiencia del cliente**, para optimizar su satisfacción en todo el circuito del viaje y la creación de campañas de marketing cada vez más personalizadas, en base al análisis de los intereses de su segmento de mercado. Todo esto, unido a una clara tendencia al alza del canal online en la compra de productos y servicios de turismo, constituye a su vez una potente fuente de generación de datos, principalmente en torno a la facilidad, amplia oferta, y comodidad que supone para el usuario.



En base a este nuevo enfoque, las prioridades incluyen abandonar la vista del turismo únicamente centrado en el destino, **analizando los datos de forma integral**: destino, empresas turísticas, y los propios turistas, así como poner al **turista** (la persona) en el centro del análisis. Esto exige un espíritu de indagación, para el levantamiento de casos de uso tractores en la economía del turismo, así como la definición de sus datasets correspondientes, pudiendo identificarse los siguientes ejes:

- Emplear un **punto de vista turístico** desde el cual agregar/contextualizar los diferentes conjuntos de datos que generan las interacciones entre proveedores de servicios y turistas.
- Identificar **problemas comunes**, así como definir posibles soluciones a dichos problemas (economías de escala).
- **Capturar datos de proximidad**, especialmente importante en un marco turístico donde existe una fuerte interrelación entre la experiencia turística y el destino.

Estas oportunidades pueden resolverse más eficientemente en base al uso de las tecnologías y modelos de negocio innovadores. Se busca por tanto articular una cultura digital y de innovación para el sector, donde algunos de los objetivos a alcanzar incluyen:

Para el turista:

- Contribuir a la simplificación de los procesos de **organización previa al viaje**, a la vez que mejorar la fase de inspiración.
- Mejorar la **personalización de su experiencia** durante su estancia, aumentando la propensión al consumo turístico y su grado de satisfacción.
- Mantener una **relación post-viaje** entre la marca del destino y el turista, para incentivar nuevas visitas basadas en

recomendaciones personalizadas online con las motivaciones del cliente.

Para los proveedores de servicios turísticos:

- Activar un nuevo ecosistema de relaciones de mercado entre dos industrias clave de la economía española: turismo y tecnología.



➤ Estimular la **colaboración con los destinos** para identificar conjuntamente retos de gestión, conjuntos de datos, tecnologías e infraestructuras útiles, de cara a la resolución de los mismos.

➤ Potenciar la **soberanía e independencia tecnológica**, fomentando la creación de empresas y soluciones digitales del y para el sector turismo y de los viajes.

➤ Favorecer la sincronización y cooperación público-privada para la elaboración de un catálogo abierto, con soluciones específicas.

Para las Administraciones Públicas con competencias en el sector turísticos:

➤ Generar incentivos para hacer atractivo nuevos modelos de negocio digitales basados en datos.

➤ Articular la colaboración público-privada para que la intervención pública sea capaz de **generar oportunidades para el sector privado** (ya que estos agentes son los que despliegan su actividad, y tienen proximidad, con el turista).

➤ Mejorar las **capacidades de marketing y promoción** de destinos.

➤ **Entender mejor la interacción del turista con el destino**, la generación de modelos de datos en dicha interacción, y la identificación de fuentes granulares que permitan modelarla más eficientemente.

En resumen, la línea de desarrollo de casos de uso debe partir de un grupo de conocimiento que recoge una necesidad, que permita definir intervenciones públicas o privadas, y así desarrollar aplicaciones tecnológicas con que instanciar dicha intervención. Significativamente, no se debe olvidar tampoco la **sostenibilidad de los proyectos** (actualmente apoyadas por inversiones como los fondos europeos en el [Componente 14 del PRTR](#), o plataformas y proyectos actualmente financiadas por programas de destinos turísticos).



Un caso de uso de compartición de datos es una solución a un reto de negocio, donde se describen tanto el objetivo, los agentes involucrados, datasets a compartir, así como las interacciones y procesos que conlleva tanto el desarrollarlo, como su operación a posteriori. Estos casos de uso se generan partiendo desde la identificación de oportunidades o desafíos que se pueden abordar con el intercambio (y después explotación) de los datos, las responsabilidades que adquirirán los agentes involucrados, tipos de datos compartidos y aplicaciones que se montarán, así como la evaluación de la complejidad en el desarrollo e interacciones (una vez desplegado el caso de uso) entre las distintas partes, donde se determina qué servicios y medidas se necesitan para generar confianza e interoperabilidad entre las partes, así como la generación de un valor final nuevo. El desarrollo de los diferentes casos de uso que resuelven un problema o una necesidad de negocio darán lugar a plataformas.

Con datos abiertos

Actualmente existen un conjunto de iniciativas destinadas a la resolución de casos de uso concretos, pero se acusa todavía **la falta de plataformas de compartición de datos generales** en el sector turístico. Se están desarrollando diferentes iniciativas privadas para fomentar la optimización de actividades turísticas en base a la explotación de datos, siendo en su mayoría plataformas basadas en datos abiertos, algunas para la extracción y posterior uso de dichos datos por las empresas u organismos turísticos y otras que proveen de cuadros de mando de explotación y análisis de dichos datos. Algunas de estas iniciativas públicas son:

- **DATASTUR:** Un agregador de datos relacionados con la industria del Turismo, lanzado a finales de 2020. Ofrece una API para descarga de datos.
- **TurEspaña:** El objetivo de este organismo es atraer turismo internacional a España. Por ello, trabaja -a partir de la explotación y síntesis de smartdata relevantes desde la perspectiva turística- en la generación y difusión interna de conocimiento digital, para diseñar y optimizar estrategias de marketing y comercialización.
- **INVAT·TUR** genera informes sobre la información más relevante para el sector turístico de la Comunitat Valenciana.
- **NAC** es la plataforma digital de turismo de Andalucía para administrar los contenidos de empresas, servicios y recursos turísticos con el fin de construir un gran catálogo digital de destinos andaluces.
- **Smart Data**, de la Junta de Andalucía, aglutina los datos turísticos acumulados en tres niveles, agrega datos procedentes de diferentes fuentes, y lo hace saltándose a los intermediarios, tales como operadores hoteleros, permitiendo a sus usuarios el acceso directo a los datos.
- **SIT – Sistema de Inteligencia Turística de SEGITTUR**, basado tecnológicamente en una plataforma Big Data en modo cloud, que estudia, incorpora y analiza distintas fuentes de información poniéndolas a disposición del sector turístico. El sistema recopila, organiza y analiza diferentes fuentes de información elaborando cuadros de mando e informes dinámicos adaptados a las necesidades del usuario y posibilitando además la descarga de la información. Facilita el acceso y la difusión del conocimiento, contribuyendo a mejorar los procesos de planificación estratégica de todo el sector, incluidos los destinos turísticos específicos. Al mejorar las capacidades de medición y seguimiento del turismo, se facilita la adopción de decisiones estratégicas mejor informadas por parte de los agentes y operadores del sector turístico.

Todas estas plataformas consumen una gran cantidad de **datos abiertos**. Este tipo de datos aportan a la innovación y la mejora de la competitividad de empresas turísticas, Administraciones Públicas, así como la calidad de vida tanto de los turistas como de los residentes locales, convirtiéndose en un elemento de gran valor. Por este motivo, ayuntamientos y otras AAPP muestran cada vez más interés en difundir los servicios que ofrecen, y así atraer al mayor número de visitantes posibles. Por ejemplo, los datos abiertos aplicados en el sector turismo aportan mejoras en la experiencia del viajero facilitando la planificación del viaje o ayudando a personalizar su experiencia; aumentan los beneficios de las empresas, sirviendo de activo para el desarrollo de servicios y aplicaciones dirigidos a los viajeros y permitiendo detectar mercados y destinos sin explotar; e incluso a la optimización de las gestiones de las AAPP, ayudando a reconocer el comportamiento de turistas y la definición de estrategias, así como permitiendo mejorar la eficiencia y eficacia de los servicios prestados.

Actualmente existen un conjunto de iniciativas destinadas a la resolución de casos de uso concretos, pero se acusa todavía la falta de plataformas de compartición de datos generales en el sector turístico.



Por todo ello, la apertura de datos en el sector turístico es una práctica común. Sólo en España existen más de 2.600 conjuntos de datos abiertos de diferente naturaleza. Algunos de ellos son la estancia media por tipos de alojamientos

por comunidades y ciudades, índices de precios por tipos de alojamientos, encuestas de coyunturas turística en distintas localidades, destinos turísticos, actividades, monumentos, itinerarios, restaurantes y bares, oficinas de turismo, estados de las playas, uso real de la bicicleta, número de visitantes o incluso información sobre rutas accesibles en sillas de ruedas, y otros como predicciones meteorológicas, o datos de geolocalización, etc.



Con datos privados y soberanos

Sin embargo, existen limitaciones en las plataformas y modelos analíticos contruidos en base a datos abiertos (conjuntos de datos limitados, bajo ritmo de refresco, frecuentemente descontextualizados, y/ o con una gran heterogeneidad de formatos y métodos de acceso). Por ello, se aspira al desarrollo de soluciones en base a conjuntos de datos más granulares, más contextualizados, y también con volúmenes más grandes. En este sentido, el paso natural es la **utilización de conjuntos de datos privados**, si bien

existen también fuertes barreras para disponer de este tipo de datos de forma general, como son los miedos a las responsabilidades por posibles fallos de privacidad, la potencial pérdida de competitividad que uno asocia a la compartición de recursos *core*, las inversiones adicionales que se hacen necesarias, la falta de talento especializado para acometerlas, o una regulación fragmentada y compleja.

En este sentido, una primera aproximación para mitigar barreras consiste en utilizar a organizaciones públicas (sin ánimo de lucro) como 3º de confianza; posición desde la que consolidar datos del ámbito empresarial de carácter privado (**B2G**), con los que generar indicadores que después compartir de nuevo con dichas organizaciones privadas (**G2B**). En estos ámbitos, encontramos casos como el de *Turismo de Andalucía* (en colaboración con la *Federación de Hoteles de Andalucía*), que aprovecha su infraestructura tecnológica para realizar pilotos en que unifican y consolidan datos de varios actores del sector turístico como *Skyscanner*, *AirBnb* -y otros- para medir (en base a tests A/B) las mejoras obtenidas entre dos campañas, una de las cuales utiliza técnicas avanzadas de explotación de datos, desarrollando además un estudio sobre qué fuentes de datos sirven para hacer crecer la oferta, o para realizar predicciones en alojamientos durante los siguientes 7 días, y así anticiparse a la demanda.

Este tipo de proyectos empiezan a vislumbrar una nueva generación de plataformas y –más genéricamente– de paradigmas de compartición de datos. La limitación que supone tener al Sector Público intermediando, es la falta de escalabilidad. Por eso surgen los **espacios de compartición de datos**. Estas estructuras son vehículos sobre los que desarrollar una economía del dato más eficaz y escalable, concebidas en base a **mecanismos de federación** (fuentes de datos que interoperan entre sí, pero que no están necesariamente

desplegadas sobre la misma infraestructura, ni bajo el control de la misma organización, ni por supuesto localización geográfica).

Si bien estos paradigmas novedosos están todavía por implementar en la práctica de forma extensiva, creemos que ya se han articulado algunos avances en dicha dirección. Concretamente, se están llevando a cabo diversas colaboraciones B2B, sobre cuyos resultados creemos interesante desarrollar acciones adicionales para generar primeras experiencias de **compartición de datos B2B**. Por ejemplo, se ha orquestado una cooperación entre varios centros tecnológicos de turismo que, de la mano de la plataforma tecnológica *Thinktur*, ha posibilitado la elaboración de un e-book de ciberseguridad¹ aplicada al sector, en el que se refleja la importancia de la seguridad de los datos, y que propone diversas consideraciones al respecto de cómo abordarla. Además, se dan también proyectos en los que –en base a la compartición y explotación de datos relacionados con el comportamiento del turista y las tendencias de sus patrones de visita- se busca aumentar las ventas y la fidelización del público objetivo.

En una línea similar- proyectos para permitir que una cadena hotelera (en base a datos del canal de reserva, el número de habitaciones, fechas, y la duración de estancia) pueda desplegar un análisis de fiabilidad en las reservas, con que realizar una previsión real de la ocupación. Esto es debido al riesgo de cancelación fruto de la situación pandémica ocasionada por el COVID-19, así como la incertidumbre laboral, que hacen que el cliente demande una mayor flexibilidad en la política de cancelaciones. Por último, también hay proyectos para trazar la huella digital del usuario, tanto la del viajero como la del ecosistema que le da servicio, desde la reserva hasta el registro, en los viajes de corte turístico-cultural.

¹ https://www.thinktur.org/media/eBook_CIBERSEGURIDAD_2022.pdf

En los mencionados espacios de datos, la gobernanza de la compartición es una cuestión crucial, tanto en el acceso, como en la reutilización de los datos. Sólo creando un **ecosistema confiable y seguro**, las diferentes organizaciones del sector turístico se encontrarán dispuestas a **ceder/ compartir sus datos**.

Retos para la escalabilidad en la compartición de datos privados para el Turismo

Por un lado, en el sector turístico, los **datos se encuentran dispersos** entre los diferentes actores del sector. El sector turístico tiene una variedad de actores bastante elevada, desde el sector público a todos los niveles (Administración Central, regiones y ayuntamientos) hasta el sector privado. En particular, en el ámbito privado se pueden encontrar tanto PYMEs y microPYMEs (industria cultural, restauración, empresas de eventos, ...), como grandes empresas en la industria hotelera, o de la digitalización del patrimonio cultural nacional. Esta gran variedad de actores y datos genera barreras no sólo para acceder a los datos, sino también a la hora de encontrarlos.

Además, no debemos olvidar que el mundo del Turismo está formado por **organizaciones digitalizadas a diferentes niveles**. Esto impacta tanto en los volúmenes de datos gestionados, como en los tipos de datos que se pueden compartir, y también en las estructuras de dichos conjuntos de datos, y sus infraestructuras. Todo esto supone dificultades para el acceso,

razón por la que **normalización y estandarización** adquieren además especial relevancia, por cuanto simplifican los procesos necesarios en torno a la compartición y explotación de datos.

Por otro lado, **la calidad de los datos no es óptima** para muchos usos que se les quiere dar. Si bien es cierto que no es necesario una calidad óptima para todo tipo de usos, el objetivo de procesamiento debe tener una calidad determinada que evite sesgos o falta de información para el objetivo para el que se quieren procesar los datos. Éste no es un problema solamente del ámbito turístico, sino de la economía del dato en general. Por ello, se están impulsando varias iniciativas como la *datasheets for datasets*² para asegurar una información clara sobre la calidad de los datos, así como la idoneidad de los objetivos de procesamiento para esos datos. En este sentido, la segmentación de los mismos resulta de gran ayuda para optimizar el diseño metodológico.

La gobernanza de la compartición es una cuestión crucial, tanto en el acceso, como en la reutilización de los datos. Sólo creando un ecosistema confiable y seguro, las diferentes organizaciones del sector turístico se encontrarán dispuestas a ceder/ compartir sus datos



² <https://arxiv.org/abs/1803.09010>

Asimismo, el respeto a la **privacidad** y la **protección de los datos** de los ciudadanos (de manera que no se vean vulnerados sus derechos) es uno de los grandes retos de los espacios de datos, y más especialmente en relación a los espacios de datos del turismo, ya que estos pueden involucrar datos de carácter personal. Si bien estos datos pueden procesarse bajo distintas bases legales, puede ocurrir que ciertas categorías especiales de datos, como por ejemplo datos biométricos (usados en algunas ferias o en aeropuertos), origen étnico o racial o creencias religiosas u orientación sexual, puedan verse expuestos durante el procesamiento de numerosos y variados datasets, o incluso ser parte de la personalización de los productos. Esto, además de aspectos desde un punto de vista normativo, tienen también consideraciones reputacionales para las entidades.

La gobernanza de datos en el sector turístico, por tanto, debe dar respuesta a todas estas reflexiones. En este sentido, existen ya herramientas de gestión de la oferta y la demanda turística (como los *channel managers*) que pueden dar respuestas parciales a estas cuestiones de gobernanza. Sin embargo, su uso no está muy extendido por el momento, por lo que se considera la misma se debe abordar todavía de manera más general.

Por otro lado, la **cultura sobre compartición del dato** es uno de los grandes retos de la economía del dato. Los actores turísticos se muestran reacios a compartir sus datos para evitar que esa compartición afecte negativamente a sus ventajas competitivas o a sus secretos comerciales. A pesar de que el sector turístico fue uno de los primeros que se

digitalizó con la venta de billetes de avión por internet, o la aparición de los marketplaces de alojamientos y vuelos, la necesidad de esta cultura de compartición de datos no se ha transmitido al conjunto de actores del sector y en especial a las PYMEs y microPYMEs. Uno de los motivos posibles de esta falta de cultura de compartición del dato puede ser debido a determinadas prácticas de concentración de datos en grandes compañías del sector. Para mitigar esta situación, sería de gran ayuda el establecimiento de agentes que funcionen como embajadores de la compartición de datos en el sector.

Para generar concienciación es necesario evangelizar que la compartición de los datos genera **rentabilidad para los agentes del sector**, no solamente en un ámbito económico de venta de los mismos, sino también como aportación de valor al negocio con la generación de nuevos productos o enriquecimiento de los ya existentes. Hay que acordar un modelo económico y de incentivos que permita establecer una recompensa para cada uno de los actores involucrados. En este sentido, se debe tener en cuenta que, para algunos, el modelo de negocio será el de poner información a cambio de una *fee*, mientras que -para otros- su valor añadido estará en extraer información de ese ecosistema de datos, elaborada a través del uso de tecnología, algoritmos, etc. Por tanto, para que los actores quieran participar las reglas del juego deben estar claras, e indicar las recompensas a la participación. Dicho análisis tiene también que ir de la mano de una evaluación de **qué datos aportan valor y cuáles no**, ya que el valor de los mismos dependerá del proyecto que se quiera abordar.

Otro aspecto relevante es la actualización de competencias en los trabajadores del sector. A este respecto, se están desarrollando iniciativas de **formación de perfiles directivos en materia de datos**, que posteriormente se posicionaran para la toma de decisiones estratégicas (que trasladen el aporte que supone de valor de los datos). Aun así, es un reto importante en el que habrá que seguir aunando esfuerzos, dada la importante brecha existente en materia de competencias digitales. Brecha aún mayor en el caso de microPYMEs que todavía no han digitalizado sus procesos más básicos. En este sentido, creemos que la Administración Pública puede cumplir un papel de apoyo y preparación al tejido empresarial.

En vista de la diversidad de perspectivas y temáticas desde las que abordar -de forma complementaria- el diseño y creación de dicho ecosistema confiable, seguro, y práctico para la industria turística, se hace necesario el consenso y la unificación en torno a unos mecanismos de gobernanza, en base a un entendimiento común acerca de lo que se quiere conseguir conjuntamente.



Principios de diseño de espacios de datos industriales turísticos

En este sentido, y en relación con la búsqueda de directrices comunes, creemos también de interés incluir la perspectiva europea respecto a unos marcos generales para la compartición y explotación de datos, según recoge la *Oficina del Dato* española (dependiente del *Ministerio de Economía y Transformación Digital*). Estos marcos proveen de una serie de principios y criterios básicos sobre los que desplegar la construcción de espacios de datos, también de utilidad en el caso específico del sector turístico, así como la ‘Metodología General para la Construcción de Espacios de Datos’ (ver *Anexo*).

Estos principios son los siguientes:

1. **Soberanía de datos**, el organismo turístico que ostenta el rol de propietario de los datos, cómo el legítimo dueño del derecho de uso sobre los mismos, debe poder determinar quién, cómo y durante cuánto tiempo se emplean dichos datasets. Un espacio de datos que provea de capacidades soberanas sobre los datos debe disponer de herramientas para que la empresa proveedora pueda definir e instanciar políticas de acceso y de uso sobre los mismos. Estas herramientas se deben basar en estándares para la definición de derechos de uso –cómo, por ejemplo, el lenguaje de marcado *ODRL*, o *Rego*–, o para la definición de políticas de control de acceso –como el *XACML*– y en tecnologías que permitan hacer efectivas las políticas de uso protegiendo la confidencialidad de los datos, cómo las denominadas PET (*Privacy Enhancing Technologies*).

Este principio se puede trasladar además a los criterios arquitectónicos de:

- a. **Empoderamiento para la compartición soberana**: garantizando que los organismos turísticos involucrados en la compartición de los datos pueden adoptar decisiones según:
 - ↗ Herramientas y procesos para la implementación de la gobernanza de la compartición escogida.
 - ↗ Definición de políticas de acceso y utilización de los datos (acuerdos offline y online).
 - ↗ Que incluya tanto a los organismos del sector turístico como a los ciudadanos (turistas, residentes, ...).
 - ↗ Desde el diseño, orientado hacia la federación.
- b. **Confianza en la compartición**: garantizando desde el diseño que los espacios de datos turísticos funcionan de acuerdo a los requisitos previstos de:
 - ↗ Seguridad de los activos del espacio de datos, y el apoyo a los acuerdos no-repudiables e inequívocos.
 - ↗ Aspectos de privacidad que deben ser tenidos en cuenta, tanto en el desarrollo de las plataformas como en las aplicaciones de intercambio de datos.
 - ↗ Garantías de seguridad y privacidad.

2. **Apertura**, los espacios de datos deberán diseñarse de tal forma que las distintas tipologías de organismos del sector turístico no encuentren barreras de entrada ni de salida del entorno de compartición de datos, así como en el diseño o uso de servicios de valor añadido, y con independencia de la solución tecnológica subyacente, así como disponibilidad de herramientas comunes y espacios de pruebas (testing). Una solución abierta permitiría al organismo turístico propietario de los datos no sólo tener total control sobre sus datos en el espacio, sino también determinar sus modelos de incentivos/ retribución/ creación de valor, e incluso dónde alojarlos y bajo qué modelos técnicos. De esta manera, el principio de apertura se puede trasladar además a los criterios arquitectónicos de:

a. **La flexibilidad para la obtención de valor** en base a la ingeniería y la ciencia de datos:

- ↗ Para interoperar datasets en base a diferentes modelos (tanto organizativos, como de negocio, como técnicos).
- ↗ Para extender los requerimientos de confianza según deseo de cada participante.
- ↗ Para desarrollar o contratar ad-hoc servicios y modelos de valor añadido.

b. **Los principios FAIR**, que extienden criterios para conjuntos de datos de carácter científico a otros más generales:

- ↗ *Findability* (encontrabilidad), permitiendo que una vez publicados, y mediante el uso de catálogos y herramientas de búsqueda, los organismos turísticos puedan localizar fuentes de datos en que pueda estar interesadas, según sus propios criterios.
- ↗ *Accessibility* (accesibilidad) para que, una vez identificados, los datos puedan ser accesibles, bajo la aceptación de las reglas que marca el dueño de los mismos.
- ↗ *Interoperability* (interoperabilidad) para que los distintos conjuntos de datos de interés al sector puedan ser fácilmente re-utilizables. Para este objetivo, deben seguir una serie mínima de reglas (marcadas por el propio sector) que habiliten la interoperabilidad:
 - Semántica (significado de los datos aplicados a un contexto).
 - Comportamiento técnico (resultados esperados en el proceso de intercambio).
 - A nivel de ámbitos, de manera que el intercambio responda a la intersección de las diferentes perspectivas normativas, organizacionales, y de las políticas de datos.
- ↗ *Re-usability* (reusabilidad) para que los datos puedan ser re-utilizados por cualquier organismo de la industria, quedando clara su procedencia y las condiciones a cumplir.

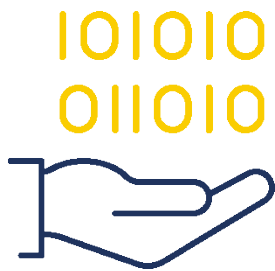


- c. **La creación de comunidad**, aprovechando las distintas aportaciones de las partes implicadas en el espacio de datos. Esto toma forma en torno a la/s:

- ↗ Soluciones basadas en estándares abiertos.
- ↗ Reusabilidad de productos pre-existentes.
- ↗ Fuentes de código abiertas (cuando así se estime).
- ↗ Sostenibilidad de las soluciones.

3. **Gobernanza compartida**, para el diseño, la construcción, operación y el mantenimiento de estos entornos abiertos, debe existir una sólida –aunque flexible– gobernanza en la que todos los organismos turísticos pertenecientes al espacio de datos se puedan ver representados y se encuentren implicados (gobernanza participativa). Este aspecto es clave para la creación de valor y la adopción masiva de estas estructuras. Además, este principio está muy ligado al criterio arquitectónico del:

- a. **Fomento de la Economía del Dato**, que sirva para crear las condiciones que aporten incentivos a la compartición, tanto económicos como de otra índole.



- ↗ Incentivos económicos: modelos de valoración y monetización de datos.
- ↗ Otros incentivos adicionales, más allá de intercambios pecunarios.
- ↗ Mecanismos de negociación, consenso, y aseguramiento del cumplimiento.

4. **Infraestructura ‘soft’ descentralizada**, en el que la infraestructura total es la suma de todos los dominios interoperables, cada uno con sus propias reglas (y también infraestructuras ‘hard’). Esta parte ‘soft’ acaba siendo la colección de facto de acuerdos, y no una estructura monolítica, que se basa en capacidades (neutrales tecnológicamente, de ahí que no sea una estructura monolítica) de cara a cumplir con una serie de requisitos funcionales y no-funcionales como la encontrabilidad, interoperabilidad, portabilidad, seguridad, privacidad y la confianza. Verdaderamente, esta infraestructura descentralizada es el resultado de la conjunción de los criterios arquitectónicos de la compartición de datos en base a la confianza, del valor que la misma aporta, tanto en base a incentivos monetarios, como a la capacidad de crear valor de negocio innovador (ágilmente y de forma flexible), así como de la interoperabilidad y federación por diseño, que se fomentan y mantienen gracias al desarrollo de una comunidad abierta.

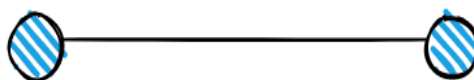
Esta interoperabilidad entre e intra-dominios es clave para que las distintas iniciativas que hay en desarrollo, y las que se pondrán en marcha en un futuro, den forma conjunta a entornos abiertos. Lo contrario llevaría a la fragmentación del mercado, así como al establecimiento de silos y monopolios sectoriales. Esta interoperabilidad implica algún nivel de consenso en el uso de estándares que afectan a la identidad y autenticación de participantes, al control de acceso, a

la definición de derechos de uso, de contratos digitales, de funcionalidades a instanciar, de definición del ecosistema y las reglas de acceso y baja, y los modelos de interacción, de gestión de incidencias y disputas, de cuestiones de marketing y crecimiento de los dominios, y de sostenibilidad financiera y operativa, así como aspectos técnicos de seguridad, de APIs y modelos de metadato, de ontologías, o de arquitecturas tecnológicas para la compartición y explotación de los conjuntos de datos, y el desarrollo y mantenimiento de los componentes necesarios. En el caso del sector turístico, esta interoperabilidad resulta además de añadida relevancia en base a la interrelación del mismo con otros sectores adyacentes, como el de movilidad, salud, logística (entre otros) y los diferentes dominios que estos puedan generar.



Teniendo en cuenta estos principios, se puede realizar una clasificación simplificada³ de las familias arquitectónicas más típicas con que abordar la compartición y explotación de datos. En el eje vertical, caracterizamos la potencialidad de extracción de valor de cada arquitectura, en base a la disponibilidad de conjuntos de datos, su interoperabilidad nativa entre ellos, así como el impacto positivo en la soberanía y la flexibilidad tecnológica. En el eje horizontal categorizamos la complejidad que supone la implantación de estas arquitecturas en la práctica. Así, modelizamos familias que van desde:

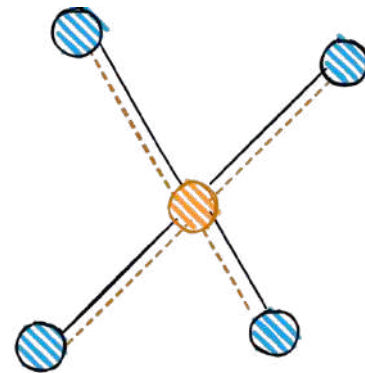
- Un **intercambio bilateral ad-hoc** en el que ambas entidades se conocen, comparten datos bajo un acuerdo entre partes (punto-a-punto), y en el que se ha acordado previamente el uso que se va a dar a dichos datos, así como el modelo de negocio. Es la arquitectura más sencilla, comúnmente empleada, pero que sufre de falta de escalabilidad natural, y basa todas sus capacidades de soberanía del dato en la confianza y fe entre las partes.



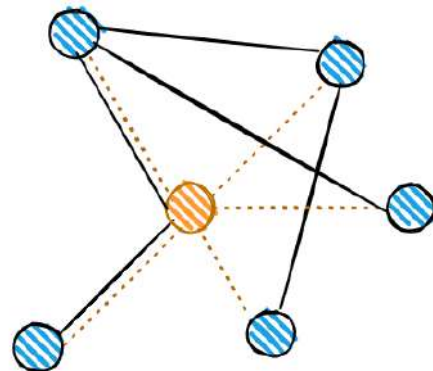
³ Ésta es una modelización genérica, pero también aplicable al desarrollo de casos de uso del sector del Turismo.

[Leyenda modelos⁴]

➔ A un nivel más complejo, en el que pueden intervenir más actores con sistemas más robustos en el control de acceso, trazabilidad y explotación de los datos, pero todavía sin herramientas que ejerzan el control soberano de los datos de forma generalizada, encontramos el modelo **centralizado**. Éste se caracteriza por disponer de un operador y una infraestructura centralizada destinada a cargar y explotar los datos procedentes de múltiples fuentes, normalmente a través de una red corporativa. Se suele disponer también de herramientas de ingestión, limpieza de datos, analítica, Machine Learning e incluso herramientas de Business Intelligence y de explotación en base a la elaboración de informes y cuadros de mando. El caso típico de este tipo de arquitecturas es un Data Warehouse, un Data Lake, o –más recientemente– un LakeHouse.

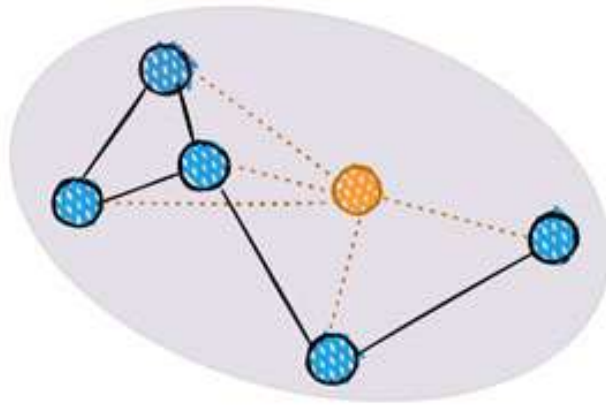


➔ Los modelos de **Hub de datos / Data Marketplace** se caracterizan por disponer de una infraestructura centralizada (gestionada típicamente por un operador único) que ofrece un catálogo de metadatos con la oferta de datos proporcionada por los proveedores, y que sirve para hacer de broker entre proveedores y consumidores de datos, normalmente con un coste asociado al uso del servicio, y donde se garantiza desde el diseño la trazabilidad y auditabilidad de los intercambios. Son soluciones orientadas a la compartición de datos, en que los consumidores y proveedores no tienen por qué conocerse de ante mano, sino que se identifican en el momento del intercambio. El intercambio de datos en sí ocurre bien punto-a-punto, o se puede también emplear el propio hub. Estas plataformas no están generalmente destinadas a interoperar fuertemente con otras arquitecturas, y suelen ser de carácter propietario.



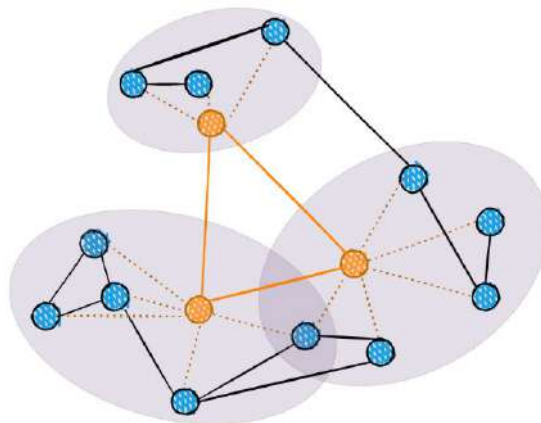
- 4
- Proveedor o consumidor de datos
 - Infraestructura centralizada
 - Proveedor o consumidor de datos y/o aplicaciones
 - Operador ecosistema de datos
 - Datos entre participantes
 - Metadatos entre operador y participante
 - Metadatos entre operadores
 - Ecosistema

-  Por el contrario, en los modelos de **ecosistemas de datos** (también típicamente conocidos como los **espacios de datos**) se da forma a una red constituida por actores autónomos que consumen, producen y/o proporcionan datos, así como servicios de valor añadido (como identificación, soberanía, trazabilidad de las operaciones, catálogo de la oferta de datos, vocabularios, aplicaciones, etc.) para que otros desarrollen las actividades de compartición. Cada actor ejecuta uno o más roles y está conectado con los demás mediante una gobernanza 'soft' que hace que su intercolaboración y competición promueva la autorregulación, y prevenga situaciones de dominio sobre el entorno de compartición. Normalmente son soluciones sectoriales. En el caso de turismo no existe ninguno de relevancia hasta el momento.



-  El modelo más complejo es el del **ecosistema de espacios de datos**. Éste es el modelo aspiracional de Gaia-X, en el que se desarrollan ecosistemas interconectados en base a redes descentralizadas, consiguiendo un grado máximo de interoperabilidad y apertura.

En una arquitectura de este tipo, un organismo del sector turístico consultaría los catálogos comunes con la oferta de servicios disponible en su espacio de datos y del resto de los espacios de datos de otros sectores, e.g. del ámbito de la movilidad, con los que estará conectado al ser (por diseño) espacios de datos interoperables entre sí. De igual manera, un proveedor de datos o de otros servicios de valor añadido puede poner a disposición sus recursos a todos los consumidores del resto del ecosistema de ecosistemas. Cada espacio de datos o dominio es –a su vez– una federación propia, operada por un agente federador que es el responsable de ofrecer los servicios esenciales para el funcionamiento de tal espacio de datos, y también responsable de conectar su dominio con el resto de las federaciones adjuntas.



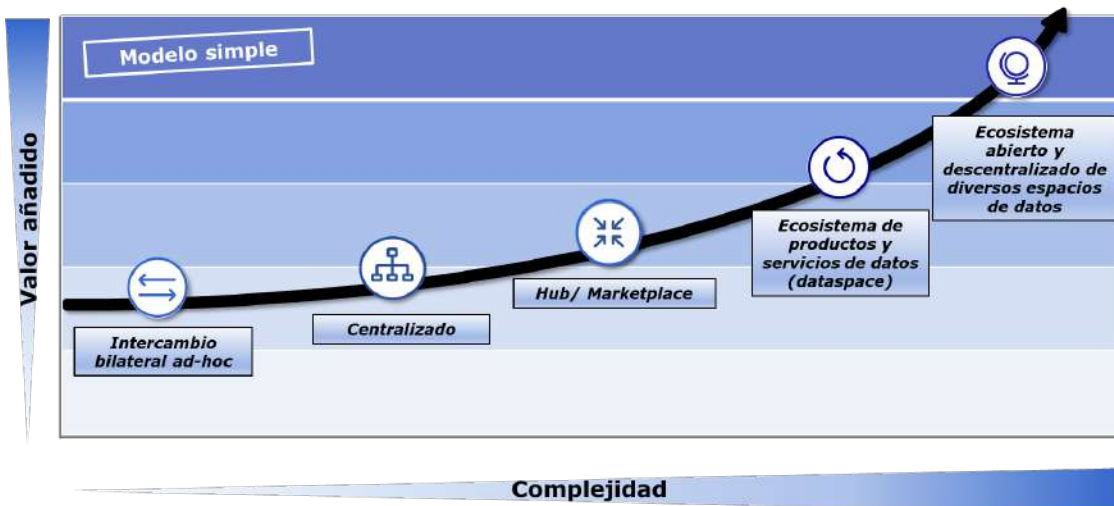
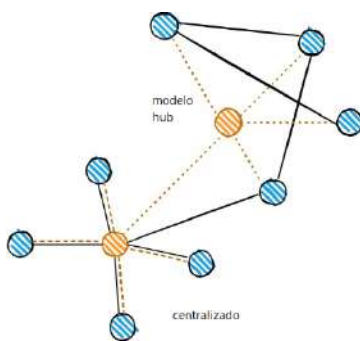


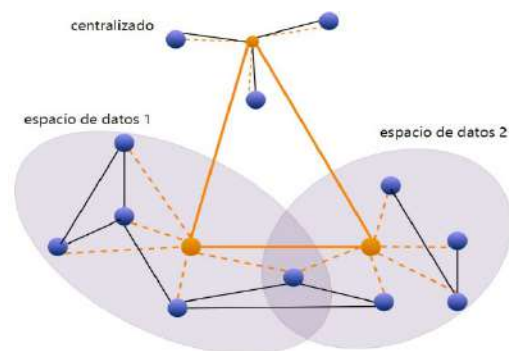
Figura 1. Clasificación simple de familias arquitectónicas para la compartición y explotación de datos



Es importante resaltar que, y aunque estos modelos no son directamente incrustables uno sobre otro –en el sentido de que no se puede asumir que las características y funcionalidades de que goza un caso de uso construido en base a un modelo, sean inmediatamente compartidas por otro caso de uso bajo un modelo más complejo– se pueden ligar de forma sencilla. Más concretamente, se pueden construir sistemas compuestos en base a federar distintos tipos de nodos y arquitecturas entre sí. De esa

manera, un modelo concreto puede formar parte de un sistema más grande, por ejemplo, un sistema centralizado que se encuentre conectado y aportando datos (procedentes de sus diferentes nodos, pero previamente consolidados a través de su operador) bajo un modelo tipo *hub*.

O que el operador (de dicho modelo centralizado) haga también funciones de federador con espacios de datos adjuntos, y por tanto pueda aportar y recibir datos de todo el ecosistema (si bien sus propios nodos deben siempre transitar los datos entrantes/salientes por medio de dicho operador).



Correspondencia entre criterios de gobernanza y arquitecturas técnicas

El desarrollo del espacio de datos del Turismo debe ser un proceso evolutivo, construido en base a diferentes casos de uso sectoriales que busquen maximizar la interoperabilidad y las funcionalidades comunes entre sí. Como es de esperar, cada caso de uso contará con unas características y especificaciones propias, fruto de los diferentes criterios de sus respectivos participantes. Las necesidades propias de negocio, sus estructuras internas – tanto funcionales y jerárquicas como sus arquitecturas de sistemas– e incluso la Cultura del Dato de cada organización y su apetito particular respecto de cada caso de uso, determinarán el consenso final en torno a la gobernanza del mismo.

En este sentido, creemos que puede ser de utilidad contar con una carta general de ámbitos y dimensiones⁵ en la compartición y explotación de datos. Al tratarse de una lista general, entendemos que no todos ellos serán aplicables a cada caso de uso; por el contrario, se trata de un inventario general con que asegurar que no quedan aspectos por considerar que pudieran ser de utilidad. Con esta carta, se busca identificar fidedignamente los criterios de cada caso de uso propuesto.

Dicha identificación da pie a una correspondencia (mapeo) entre los requisitos y las familias arquitectónicas disponibles (según modelizamos en la sección anterior). De esta manera, buscamos afiliar de manera rápida y sencilla clases de criterios (de compartición y explotación de datos) a modelos arquitectónicos comunes, que cuentan con una serie de componentes

tecnológicos habituales. Estos componentes habrían sido previamente ligados a plantillas y scripts para su despliegue semi-automático sobre infraestructuras cloud. Esta metodología, conocida como “*infrastructure as code*”, permite desplegar de manera sencilla y rápida bloques tecnológicos comunes interlazados y pre-configurados para su uso inmediato.

Así, desplegamos building blocks tecnológicos (ver siguiente sección) según los criterios y requerimientos de cada caso de uso, dando forma a arquitecturas semi-prefabricadas (con un alto porcentaje de componentes ya listos para su uso) que responden a los acuerdos de los participantes en el mismo. Es un concepto de “*architecture as a service*” (arquitectura como servicio), que facilita el desarrollo de pruebas rápidas (experimentación con modelos de compartición de datos), la evangelización y creación de cultura alrededor de los mismos, fomenta el aprendizaje práctico sobre funcionalidades de negocio y componentes técnicos necesarios, y sirve incluso para tomar el pulso a la industria tecnológica, de cara a cubrir posibles gaps existentes, y desarrollar paradigmas nuevos.



⁵ Listados en el Anexo de este documento.

Necesidades técnicas

Como describíamos arriba, la gobernanza común que debe regir estos dominios turísticos debe abordar muy diferentes perspectivas, incluyendo la del propósito o misión de cada dominio, la de negocio, la normativa, la funcional, la relativa a la calidad y estructuras de los datos, la operacional, y por supuesto la tecnológica, que sirve para dar forma y sostener las restantes. Esta dimensión tecnológica se ve también afectada por las **peculiaridades de sector del Turismo**, como son la **atomización**, que supone una dificultad para encontrar las fuentes de datos de interés, de entre toda la información generada y disponible; la **transversalidad** del sector, pues se trata de una industria que implica a una gran multiplicidad de terceros sectores (esto incide en la atomización y señala en la importancia de las taxonomías y categorización de datos, con que desplegar nexos entre sectores), y la alta **heterogeneidad en los niveles de madurez digitales** de los actores turísticos, que va desde alta madurez en organizaciones transnacionales, como a extremadamente bajo, para aquellas microPYMEs que operan en base a modelos más físicos/ analógicos.

Estas características hacen que, desde un punto de vista tecnológico, centremos la atención en cuatro bloques principales:

- **Ciberseguridad.** Para que prosperen los espacios de datos, es necesario garantizar los aspectos clásicos de seguridad. Esto es algo que se aborda de forma natural en base a la tecnología.
- **Soberanía y control del uso.** Más allá de la seguridad de un sistema, es clave también abordar los aspectos de soberanía del dato, entendiendo ésta como la capacidad de controlar el uso que se hace sobre los datos propios, en base a la definición de criterios y reglas.
- **Interoperabilidad.** La practicidad y escalabilidad de estos entornos federados, donde confluyen diferentes perspectivas y herramientas, van indudablemente ligadas a la compatibilidad técnica y semántica.
- **Creación de valor.** Por último, resulta esencial incorporar la gestión de los recursos de datos como activos con valor comercial. Activos que pueden ser publicados, descubiertos y, eventualmente, comercializados. De este modo se impulsa la creación de mercados en los que se puedan desarrollar servicios innovadores.

Seguridad

Como decimos, de cara a la construcción de estos espacios de confianza, donde sus participantes puedan colaborar sin miedo, la **ciberseguridad** adquiere una especial importancia. Ya no vale sólo con tener en cuenta los riesgos y amenazas convencionales a los que día a día se enfrentaba (y sigue enfrentando) el sector, sino que además ha entrado en juego la seguridad del ciberespacio. Un espacio anárquico donde no existen fronteras, altamente dinámico, todavía débilmente regulado en algunos aspectos, y completamente accesible a nivel global. Por ello, el sector debe ser consciente de

asegurar la gestión de la información de sus clientes, integrar la tecnología innovadora de manera segura, así como fomentar la concienciación y la cultura de ciberseguridad, haciendo conocedores a los participantes de los espacios de datos de los riesgos y amenazas que conlleva la conectividad exponencial que se genera dentro de dichos espacios.



Soberanía

En este sentido, y además de los aspectos de seguridad de los sistemas de información más tradicionales, el factor novedoso en estos espacios de datos es el relativo a la compartición de datasets y algoritmos. No sólo deben ser compartidos de forma segura, sino que se debe asegurar además que el uso que se hace coincide con las expectativas originalmente definidas para los mismos. Éste es un campo más incipiente, y por ello se hace necesario alcanzar cuotas más altas de desarrollo tecnológico, que permitan compartir recursos asegurando en todo momento **que se cumple con las políticas de soberanía** del dueño de los derechos de actuación sobre ellos. Actualmente se trabaja a lo largo de tres líneas tecnológicas principales:

- **Pruebas de conocimiento cero.** Estructuras que garantizan cualquier atributo del usuario, sin que éste tenga que informar al respecto. Por ejemplo, si un servicio turístico sólo es accesible para mayores de edad, una organización certificadora validaría que el turista cumple con esa condición, sin que éste deba comunicarlo al servicio en cuestión.
- **Computación homomórfica**, donde los datos de los turistas sólo se encuentran disponibles de forma cifrada (no legible), pero -aun así- los algoritmos son capaces de operar eficientemente sobre ellos.
- **Aprendizaje federado.** Los modelos se construyen secuencialmente, viajando a lo largo de diferentes fuentes de datos, donde van progresivamente optimizándose. La clave aquí es que cada dataset no sufre copia o modificación alguna, así como que el algoritmo entrenado (con dicha fuente de datos) no

contiene ninguna información que pueda permitir identificar específicamente las fuentes originales.

- **Combinación de varias tecnologías.**



Interoperabilidad

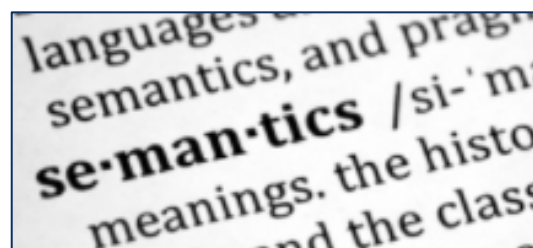
Así mismo, además de la seguridad y soberanía en el intercambio y explotación de los datos, otro aspecto fundamental a la hora de que dicha compartición tenga utilidad práctica, es la interoperabilidad entre sistemas, empezando por los propios conjuntos de datos, que –al provenir de diferentes ámbitos y organizaciones- no suelen estar preparados para dialogar entre sí. Buscamos por tanto generar una **interconexión entre silos de datos**. Para ello, creemos que los estándares semánticos resultan una pieza de gran utilidad. Es algo en que además España tiene gran experiencia, en base al avance de referencia mundial en materia de normalización semántica de las *Smart Cities*, que además guarda una fuerte relación con la compartición de datos en turismo.

Interoperabilidad semántica

En este sentido, **las normas UNE 178504** -sobre hotel digital, inteligente y conectado (HDIC), que generan una serie de requisitos y recomendaciones para convertir un alojamiento en un hotel digital inteligente, conectado con el destino turístico o ciudad inteligente- **y la UNE 178503** -sobre semántica aplicada a modelizar cómo se estructuran los contenidos que constituyen la capa digital del destino, con el objetivo de definir la semántica base que permite representar la información relevante del destino turístico, asegurando la interoperabilidad entre las plataformas turísticas y la ciudad, o territorio, y con desarrollos de terceros- son herramientas que **sirven para articular la interrelación entre sectores y sus respectivas fuentes de datos**. Dentro de esta última norma, se define además un vocabulario que recoge una interpretación precisa de los conceptos manejados por turistas, destinos y la propia industria. Este vocabulario asegura el tratamiento homogéneo de la información, facilitando la interoperabilidad y el análisis de los datos. Además, contiene una taxonomía aplicada a la actividad turística en destino, estructurada como una colección de clases o tipologías de objetos y elementos turísticos definidos desde el punto de vista del itinerario del viajero, atendiendo a sus necesidades

de información, y con el objetivo de cubrir el espectro completo del comportamiento turístico en destino: organización de viaje e información del propio destino, paquetes, alojamientos, restaurantes, ocio, atractivos turísticos, eventos, y experiencias; y -por último- metodologías y herramientas para facilitar la serialización, etiquetado y validación de los datos.

En resumen, estas normas son parte del ecosistema de nuevas herramientas tecnológicas que sirven para profundizar en el paradigma del Turismo Inteligente. Abordan dos aspectos fundamentales del destino inteligente: en primer lugar, favorecer la comunicación bidireccional entre el hotel y el destino, y -en segundo lugar- proveer de una estructura semántica homogénea que garantiza que la **comunicación e interoperabilidad entre los diferentes sistemas y plataformas tecnológicas ocurre inequívocamente, bajo un mismo estándar**. Esto aspira a repercutir en una mejora de la experiencia del viaje del turista, y de la competitividad y planificación del hotel y el propio destino.



Interoperabilidad de sistemas

Pero, incluso más allá de los estándares semánticos o de la información, a un nivel inferior, es clave también asegurar una **interoperabilidad técnica entre sistemas**. Dado que los espacios de datos son ecosistemas abiertos, donde por definición se articulan diferentes funcionalidades y configuraciones que deben interactuar entre sí, esta interoperabilidad implica un reto complejo, pero en el que la estandarización puede desempeñar un papel importante. En este sentido, se espera que los espacios de datos hagan uso de iniciativas y **trabajos de normalización previos**. Por ejemplo, el *World Wide Web Consortium* (W3C) ofrece una lista⁶ (no exhaustiva) de estándares para la definición de políticas de control de acceso y de derechos sobre el uso de datos. Esta lista es interesante por cuanto en los dataspaces confluyen un gran número de estándares procedentes de distintos ámbitos, y -por ello- la ambición en torno a su estandarización no consiste en proponer un único estándar concreto, sino **una selección de estándares que ya interoperan entre sí**. Dicho de otra manera, el proceso de normalización de un espacio de datos aspira a seleccionar -de la forma más óptima posible- de entre un conjunto numeroso de estándares previamente disponibles. En determinados ámbitos, e.g. la arriba mencionada definición de políticas sobre el uso de datos, al no existir estándares 100% adecuados, lo más práctico resulta extender estándares ya comunes⁷, de manera que se hagan funcionar dentro del contexto de un determinado espacio de datos.

⁶<https://www.w3.org/Policy/pling/wiki/PolicyLangReview>

⁷ Por ejemplo, el del ODRL - *Open Digital Rights Language*

Interoperabilidad de infraestructuras

Estos estándares a los que hacemos referencia son del mundo del software, pero –de igual manera- uno ha de pensar en **las infraestructuras** que sustentan dichas funcionalidades informáticas. Es algo en que la [iniciativa Gaia-X](#) trabaja desde su creación, promoviendo la economía del dato y la soberanía y resiliencia digital en base a la conceptualización y diseño de una capa de servicios de federación (conocidos como los [Gaia-X Federations Services](#) – GXFS), cuyo objetivo es desacoplar la capa de datos de la de infraestructura (de dichos datos). Los GXFS están compuestos por 4 familias tecnológicas y un artefacto:

- Servicios de identidad, para asegurar el quién es quién de los participantes (clave en entornos distribuidos).
- Un catálogo federado de los servicios ofertados que aporten valor al sector.
- Tecnologías para asegurar la soberanía en el intercambio de datos (contratos, políticas de uso) tanto entre organismos del sector turístico, como de otros sectores con los que guarden relación.
- Mecanismos de conformidad/cumplimiento, que afectan a las distintas familias, y sobre los que se sustenta la confianza del sistema.
- Un portal (para el onboarding y acreditación de participantes), con capacidades para orquestar las diferentes funcionalidades necesarias.



El objetivo de Gaia-X es dar forma a ecosistemas soberanos de datos, desplegados sobre servicios de mercado (de datos e infraestructura de datos), contruidos estos en base a un *framework* técnico y arquitectónico que ha sido concebido para fomentar la interoperabilidad entre recursos de datos federados. Este paradigma posibilita no sólo la capacidad de conectar datos entre sí, sino también de capacidades computacionales (aplicaciones SW, e incluso componentes de infraestructura), habilitando de esa forma que los instrumentos sobre los que desplegar dichos ecosistemas de datos limiten su dependencia de sistemas tecnológicos monolíticos. Esto es importante ya que la utilidad práctica de estos diseños depende en gran medida de su capacidad de atracción, y de ofrecer soluciones específicas a necesidades, i.e. **flexibilidad de negocio, que es precisamente donde brillan los entornos distribuidos** (frente a los monolíticos).

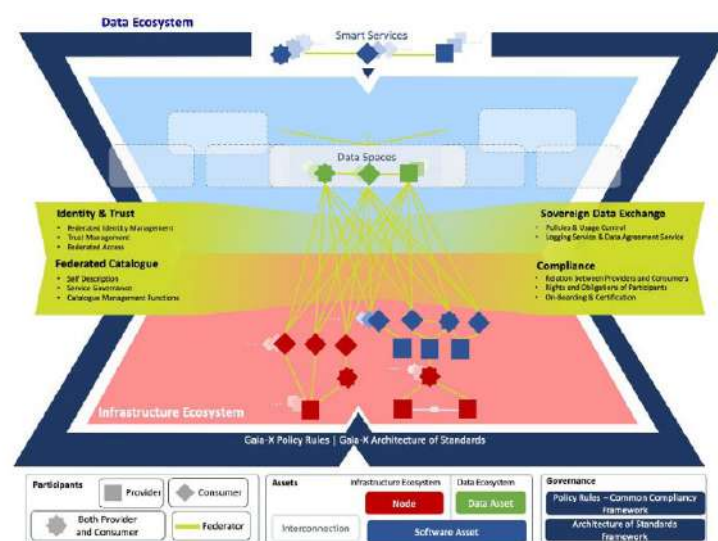


Figura 2. Modelo Gaia-X

Convergencia para la creación de valor

En este contexto de federación, resulta evidente que la construcción de espacios de datos requiere de la confluencia de diferentes perspectivas complementarias. Los mismos deben abordarse mediante un enfoque abierto, en el sentido de que tienen que ser capaces de aunar oferta y demanda, y dar respuesta a requisitos reales de negocio. Por tanto, su desarrollo no debe concebirse de manera aislada, o desde la visión exclusiva de una sola asociación, sino de la forma más holística posible, que sea capaz de crear tracción en el mercado, y por ende llegue a buen puerto con amplia capilaridad. Esto es algo que también han identificado las principales organizaciones europeas en materia de dataspaces, que en otoño de 2021 crean la *Alianza Empresarial para los Espacios de Datos* ([Data Spaces Business Alliance](#)).

Ésta se crea para impulsar la adopción del paradigma de espacios de datos, a través de una sola voz y bajo un próximo marco único y común. Juntas, las organizaciones fundadoras⁸ de la *Alianza* representan a más de 1.000 actores clave de las industrias, el mundo asociativo, de la investigación e innovación, así como los *policy-makers*. En base a su experiencia, recursos y conocimientos técnicos, la *Alianza* busca impulsar la toma de conciencia, la evangelización en estos paradigmas y tecnologías, dar forma conjunta a estándares, y –más genéricamente–interconectar y hacer **converger las diferentes iniciativas y visiones** en materia de espacios de datos.

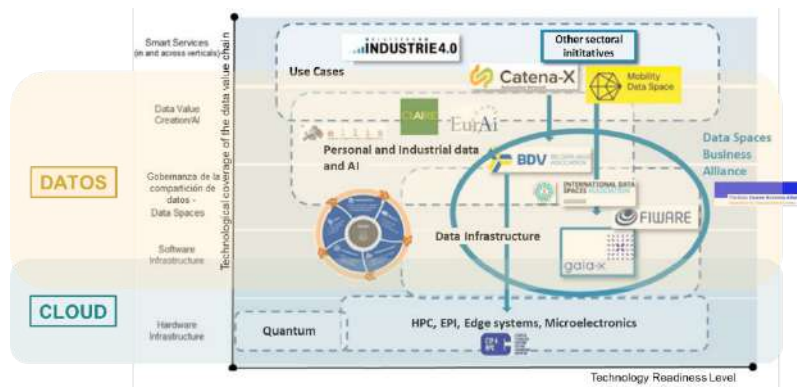


Figura 3. Diferentes iniciativas europeas en la cadena de valor de explotación del dato, clasificadas en torno a su madurez tecnológica (TRL) y a la tecnologías y capas de la cadena que aborda cada una (Basado en: BDVA)

Esta convergencia tiene sentido ya que los espacios de datos deben ser abordados desde diferentes perspectivas, en base a la compleja cadena de valor de explotación del dato. En el diagrama de arriba se hace una categorización de diferentes organizaciones a lo largo de dicha cadena, considerando también sus respectivos niveles de desarrollo tecnológicos. En la parte superior encontramos la perspectiva más cercana a Negocio, donde se encuentran los servicios inteligentes (*Smart Services*) de datos, y todas aquellas iniciativas sectoriales y aplicaciones de negocio. Aquí ubicamos casos de uso dentro de las iniciativas más maduras de espacios de datos como [Catena-X](#) (automoción) el [Mobility Data Space](#), la [Smart Connected Supplier Network](#) (centrado en las cadenas de suministro para la manufactura), etc.

⁸ BDVA (*Big Data Value Association*), IDSA (*International Data Spaces Association*), FIWARE Foundation, y Gaia-X AISBL.

En la siguiente capa encontramos la creación de valor en base al uso de los datos, e.g. mediante la creación de algoritmos, el uso de tecnologías Big Data, técnicas de Inteligencia Artificial, así como otros mecanismos de extracción de valor. En esta capa se ubican actuaciones de iniciativas como la BDVA, así como otros esfuerzos privados concretos de diferentes industrias. En el siguiente nivel, se aborda ya la creación de dataspaces, desde donde instanciar la compartición de datos (como elemento catalizador para la creación de valor en base a la explotación de volúmenes más extensos) y donde también abordar aspectos relacionados con la gestión y el gobierno del dato (vis à vis proveedores de servicios de datos). En esta capa podemos encontrar marcos de referencia para la compartición como el de los *International Data Spaces*, paradigmas en boga hoy en día como el *Data Mesh*, o soluciones tipo *marketplaces*.

La siguiente capa, de infraestructura software, es donde se concreta la instanciación práctica de los diversos paradigmas disponibles en torno a la gestión, compartición y explotación de datos. Se observan tanto aplicaciones generales como funcionalidades concretas, e incluso sistemas *middleware* tipo Gaia-X, que buscan habilitar un desacoplamiento lógico entre las funcionalidades de datos y sus infraestructuras tecnológicas, y con ello dotar de portabilidad, flexibilidad y capilaridad a dichos paradigmas. En última instancia, observamos también la capa de infraestructura hardware, más puramente física, donde se ubican los servicios de cómputo y almacenaje,

pero también incluso capacidades de HPC (*High Performance Computing*), sistemas ‘en el borde’ (*edge*) y el ansiado continuo computacional nube-a-borde (*cloud-to-edge continuum*), o incluso –más ambiciosamente– la computación cuántica.

En esencia, todas estas perspectivas tienden a complementarse de manera natural, ya que interactúan entre sí y se sustentan unas sobre otras. Esta complementariedad en realidad se consigue por diseño, en base a una abstracción de la **cadena de valor del dato**, que sirve para transformarlo desde su *estado natural* (una colección de registros bajo la perspectiva de un modelo representacional concreto), a *información* (al contextualizar los registros en base a baremos que ayudan a relacionarlos y relativizarlos entre sí), a *conocimiento* (vinculando diferentes contextos, y pudiendo así emplear la información para un propósito específico), y por último a *sabiduría*, generando una capacidad prescriptiva de alto valor añadido.

En resumen, las diferentes metodologías, vehículos y herramientas tecnológicas (correspondientes a cada una de las perspectivas descritas arriba) usadas en dicha transformación del dato, sirven para desarrollar modelos abstractos que después conjugan entre sí para resolver problemas complejos de relevancia lógica humana (donde reside el valor latente del dato). Este proceso de conjugación es por lo que comúnmente se habla del dato como el nuevo “petróleo del siglo XXI”, pues debe ser refinado y tratado para exprimirle provecho. Esta analogía es

apropiada en cuanto a que el dato pasa a considerarse un recurso estratégico para las organizaciones; un activo que –al igual que cualquier otro bien material– tiene un valor financiero, y cuya gestión genera también costes. Los datos creados, recogidos o utilizados en los procesos industriales/ organizacionales pueden también comercializarse con otras entidades como datos brutos o procesados, de modo que pasan a convertirse en productos en sí mismos, y a dar forma a un “mercado de datos” en los que apoyarse para desarrollar nuevos servicios y productos innovadores.



Building blocks para dataspaces

Al respecto de la abstracción de la **cadena de valor del dato**, ésta nos permite desplegar una modularización lógica dentro de la misma, y –así– encapsular funcionalidades básicas en torno a los llamados **building blocks** (o bloques de construcción). Esos son componentes comunes básicos, cohesionados entre sí para dar respuesta a los principios de diseño de los espacios de datos. Concretamente, los building blocks se desarrollan en torno a las siguientes características.

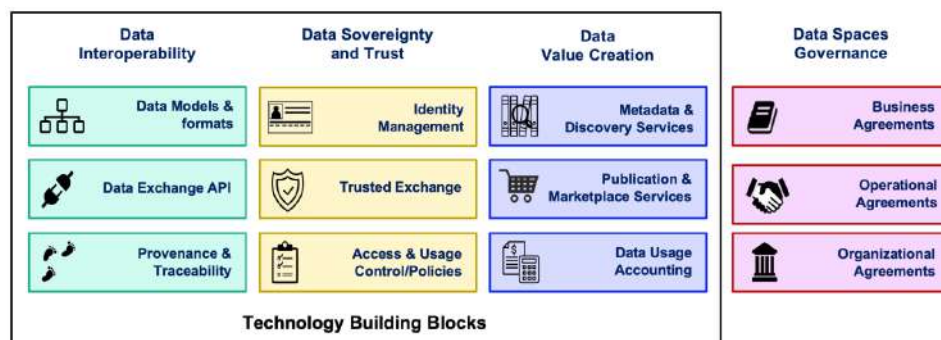


Figura 4. Familias de building blocks para desplegar espacios de datos

Tres familias con un sabor tecnológico:

- **Interoperabilidad.** Proveedores y consumidores que no se conocen deben hablar el mismo idioma para poder publicar, buscar y consumir datos. Esto se traduce en usar tanto **modelos de datos** y formatos (a la hora de publicar y consumir), como **APIs** (para realizar el intercambio) estandarizados e interoperables. Ejemplos de uso extendido son los *Smart Data Models*, las *Self-Descriptions*, modelos basados en el paradigma *Linked Data* como el NGSI-LD o el JSON-LD. Además, al proceso de provisión y consumo de datos se le añaden también componentes que

proporcionan información sobre la **trazabilidad y procedencia** de los datasets, relevantes para requerimientos específicos de transparencia y aseguramiento de sus características, por ejemplo la IDS Clearing house o el Gaia-X Registry.

- **Confianza y soberanía.** Dado que los participantes que se unen a un espacio de datos no se conocen, es necesario dotar de mecanismos que proporcionen confianza y garanticen la soberanía de los datos. Se debe por tanto realizar una **gestión de la identidad** (*identity management*) que garantice que todo participante es quien asegura ser, incluso a través de diferentes dominios (por ejemplo, utilizando gestores en base a PKI, DID, Verifiable Credentials o OpenID Connect), y cumpliendo en todo momento con la regulación europea (de identificación electrónica y de servicios de confianza) [eIDAS](#).

Para un **intercambio confiable** se debe además poder confirmar que las declaraciones de los participantes son verdaderas, y que el propio intercambio de mensajes y datos está dotado de integridad (por ejemplo, en base a JWTs y firma digital), así como el uso de privacy-enhancing technologies, que permiten operar sobre conjuntos de datos asegurando los niveles privacidad. Por último, deben implementarse **políticas de acceso, control y uso de los datos** a nivel de atributos, y complementado con derechos y licencias de uso, como punto clave para ejercer la soberanía (por ejemplo, con arquitecturas basadas en PDP, PAP y el lenguaje XACML).

- **Creación de valor.** Es necesario caracterizar los **recursos de datos como activos** con valor empresarial, de forma que puedan ser publicados y descubiertos y –eventualmente– negociados en el ámbito de un **marketplace** (o funcionalidad similar) entre las distintas y desconocidas partes. Para esta tarea resulta fundamental el uso de metadatos sobre dichos activos de datos, dotándoles de significado y contexto. Este intercambio implica además la necesidad de un componente que pueda llevar la **contabilidad/ auditabilidad** del traspaso y uso de dichos datasets.

Una familia no-tecnológica para aspectos de *Gobernanza*. Los participantes, y potenciales participantes, además de en torno a los componentes técnicos descritos anteriormente, necesitan también certeza sobre las consideraciones operativas, legales y comerciales que rigen los intercambios de datos. Los criterios alrededor de estos aspectos claramente podrán evolucionar con el tiempo, con el objetivo de no perder valor, y que en todo momento puedan reflejar adecuadamente las condiciones de mercado que deben mediar los procesos de compartición de datos.

- En cuanto a los **acuerdos comerciales**, se debe establecer un marco legal y encontrar consenso sobre aspectos como el modelo de costos, el propósito, y el uso de los datos en función de su criticidad, así como las responsabilidades asociadas. Además, los participantes de los casos de uso del espacio de datos pueden también identificar qué estándares de intercambio de datos, así como vocabularios y otras consideraciones técnicas, prefieren usar para interoperar entre sí.
- Así mismo, se hace necesario establecer **acuerdos operativos** que proporcionen confiabilidad en que algunas partes participantes del espacio de datos

(proveedores de servicios, de identidad, de autorización, etc.) estarán legalmente obligadas a mantener el tiempo de actividad y la disponibilidad de los servicios. Esto puede plasmarse mediante *Service Level Agreements* (SLAs) que se firmen a la hora de la entrada en vigor de dichos servicios, junto con otras consideraciones como la gestión del ciclo de vida de los participantes, de incidentes, del cambio y nuevas versiones. Incluso, se hace posible añadir acuerdos de comunicación con que proporcionar pautas a los participantes (en el espacio de datos), de cara a promocionar y hacer crecerlo.

- Por último, como decíamos, el espacio de datos debe ser capaz de evolucionar y adaptarse a las necesidades de sus participantes. Eso implica que los acuerdos arriba mencionados no deberían ser necesariamente inamovibles. Para ello, en base a **acuerdos organizativos** y por consenso de los participantes, se deben establecer órganos rectores que consoliden las directrices. Para espacios de datos abiertos, estos órganos deberían además ser neutrales e imparciales, de forma que aseguren la igualdad de oportunidades entre participantes.

Todos estos building blocks, tanto tecnológicos como de gobernanza, deben **confluir en un framework conceptual sobre el que desarrollar dataspace comunes e interoperables**, en base a unos criterios de diseño centrales. Debe por tanto adscribir una misión y visión común, con un plan de actuación para todas las partes involucradas, así como un *modelo tecnológico de referencia* con diferentes tecnologías y piezas de software modulares, en base a las que desplegar espacio de datos operativos.

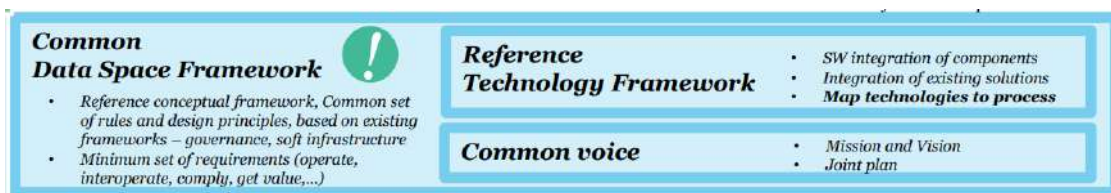


Figura 5. Framework común para los dataspaces

Además, esta convergencia tecnológica se ve también complementada con una confluencia en ámbitos operativos, como son el entendimiento y la adaptación, así como influencia, en materia de **Regulación** de espacios de datos; la catálisis de **innovación** a través del ecosistema –fuertemente ligada a la capacidad de **experimentación y la transferencia de conocimiento**– y, cercanamente, la **detección tecnológica temprana**, así como las estrategias y hojas de ruta para la evolución tecnológica. Se aspira a desarrollar ecosistemas de alcance sectorial (interconectados entre sí) que sean capaces de dar forma conjunta a espacios de datos abiertos, flexibles, y con alta capilaridad de mercado. Para tal efecto, se está trabajando también en crear una voz común respecto a los aspectos procedimentales, incluyendo la **identificación de oportunidades y match-making**, el peritaje en la **madurez de iniciativas**, el desarrollo de **manuales de recomendaciones** y mejores prácticas para su despliegue, así como aspectos relativos al talento, y otros tipos de **apoyo al crecimiento** de estos proyectos.

7

Conclusiones y siguientes pasos

A modo de conclusiones, se extraen las siguientes ideas y planteamientos:



Conclusiones

- Es fundamental empezar a trabajar en **definir las reglas del juego del espacio de datos de turismo**. No partiendo desde cero, sino de iniciativas que ya se hayan puesto en marcha, evitando solapamientos y asegurando que se trabaje conjuntamente.
- Dentro de dichas reglas del juego, conviene **definir claramente los roles de los diferentes actores**, así como la coordinación necesaria.
- No hay que perder de vista el nivel de **segmentación y fragmentación del sector turístico**.
- Es también importante **definir el dato, el nivel del dato, y las condiciones de uso** (con quizá el establecimiento de un modelo de recompensas).
- **Aprender de casos de uso existentes** y extraer de los mismos las lecciones pertinentes.
- Siendo el modelo de negocio clave, así como las implicaciones legales, hay que tener en cuenta que **la monetización tiene que consistir en el fin único, sino también mejorar los servicios, la eficiencia, etc.**
- La definición de la **gobernanza y tecnología** son fundamentales.
- No olvidar nunca la **transversalidad del sector turístico** y su fuerte vinculación a otros sectores de actividad.
- El nivel de partida en la Digitalización no es alto. Por ello, hay que dar una respuesta preventiva y **buscar soluciones de bajo coste**, teniendo en cuenta que la confianza no va a ser muy alta al principio.
- Es conveniente **determinar una estandarización muy alta dentro del espacio de datos y con otros espacios de datos**, e implementar tecnologías que permitan la interoperabilidad.
- Hay que ser pragmáticos y apostar por una dinámica práctica, para **que el participante pueda elegir** su modo de involucración y colaboración.
- Hay que **concienciar y convencer a las empresas del valor general latente** en la cesión de datos.
- Teniendo en cuenta que cada empresa tiene un lenguaje tecnológico diferente, se debería definir una **ontología común**.
- El espacio de datos debería ser también un **lugar desde el que crear oportunidades**.
- **Palabras clave: Consenso y unificación.** Unificar de manera lógica, en base una federación de sistemas bajo una misión común.

En líneas generales, se puede afirmar que en España existen numerosas iniciativas para fomentar la creación de espacios de datos en el sector turístico, aunque su característica principal es la atomización y diversidad de casos de uso debido a la idiosincrasia del sector, por lo que es importante remarcar la colaboración, generosidad de los participantes y sobre todo generar confianza para la compartición de los datos en pro de un bien común.

Por último, y de cara a avanzar en la construcción del espacio de datos del Turismo español, se consensuan los **siguientes pasos generales**, apoyados en la 'Metodología General para la Construcción de Espacios de Datos' [ver *Anexo*]:

01

Caracterización del entorno turístico, y los requisitos tecnológicos y no tecnológicos.

02

Definición y **aprobación de una gobernanza consensuada** entre los (las tipologías de) participantes de cada caso de uso.

03

Generación de una **red de colaboración**, con sinergias específicas con otros espacios de datos.

04

Debido a la falta de maduración de los espacios de datos, tanto desde su aspecto más tecnológico como desde su aspecto administrativo, se pone de relevancia la **importancia de la experimentación**, tanto para consolidación de los componentes tecnológicos, como para la validación de la infraestructura 'soft' que formen cada caso de uso (que conjuntamente conformarán el espacio de datos). Esta experimentación requiere de infraestructura, soporte, datos y formación, configurados a medida para cada caso de uso.



ANEXO – Metodología general para la construcción de espacios de datos

De cara a plasmar de forma fidedigna la radiografía del sector turístico, creemos que el siguiente paso es el levantamiento de distintos **casos de uso de compartición de datos**, que impulsen a los actores sectoriales a tomar acción práctica en temas de interés directo. La importancia de recoger esta información reside no solamente en conocer aquellos casos de uso de mayor urgencia, o valor añadido, sino también en la metodología del trabajo colaborativo, y la planificación del mismo, que deben abordar los diferentes grupos de interés, y sobre la que desplegar actuaciones concretas.

Estos casos de uso aspiran a ser recogidos en un **workshop**, en el que un grupo de organizaciones de la industria, lo más representativo posible, trabajará de forma conjunta en varias sesiones grupales. De cara a lograr una homogenización en dicho grupo, sobre la que luego desplegar un diseño homogéneo de casos de uso, se aspira a incluir primero una pequeña sesión de ‘baseline setting’, que sirva para fijar terminología y conceptos. Inmediatamente después, los participantes se enfrentarán a la identificación de oportunidades (o desafíos sectoriales) a abordar en base al intercambio de conjuntos de datos concretos, la adaptación de los mismos al alcance y visión específica de su organización, así como una **evaluación de la complejidad** de dichos casos de uso, en base a la interacción que se hace necesaria entre actores, y las medidas apropiadas para generar confianza e interoperabilidad entre las partes. Este trabajo debe incluir también un análisis de partida del estado de madurez de sus organismos y su grado de implicación, así como una **definición de un plan de acción** para cada caso de uso en base a:

- Los incentivos de participación y definición de factores de éxito
- El modelo colaborativo
- La gobernanza que se determine idónea
- El análisis arquitectónico de datos (incluyendo tecnologías y *roadmaps* de desarrollo)
- La identificación de grupos de interés y posibles partenariados estratégicos
- Monitorización y mejoras.

Un subproducto resultante de toda esta actividad será la identificación de lecciones aprendidas durante las dinámicas y siguientes pasos, que también permitirá la adaptación a futuro de los casos de uso que progresivamente aumenten el espacio de datos.

Una vez diseñados los casos de uso, dicha interoperabilidad y funcionalidad de negocio se empezará plasmar a través de un MVP (**Minimum Value Product**) del espacio de datos de turismo. La idea es que en éste se desplieguen casos de uso prominentes, que capturen particularidades del sector, tanto desde el punto de vista de creación de valor y de estandarización del consenso en la compartición, como de la diversidad de participantes involucrados, que sirva para encarnar el alto grado de atomización del sector y –por ende– que aborde satisfactoriamente la falta de interconexión nativa entre distintas fuentes de datos, y muestre el valor añadido de estos novedosos paradigmas. Se espera que dicho espacio de datos **crezca evolutivamente**, tanto en casos de uso,

como en participantes de casos de uso existentes. Esto implica que se hace necesaria una operativa mínima a tal efecto, en primer lugar, el alta y baja de usuarios; pero también servicios de identificación, autenticación y autorizaciones confiables; la capacidad de publicar y buscar datos en un catálogo dinámico de servicios; la habilidad de definir políticas de uso, incluyendo la negociación y aceptación/ rechazo entre las partes, junto a la creación de listas negras y blancas de usuarios con capacidad de acceso a los datos publicados; así como la facilidad de búsqueda, descarga y ejecución de aplicaciones desde una *App Store* común.

En vista de la diversidad de ámbitos a considerar de cara a llevar a buen puerto la generación de este MVP del espacio de datos de turismo, sugerimos una **metodología con pautas básicas** a seguir. Ésta busca facilitar tanto la creación de estos espacios de datos, como su evolución. La creación de espacios de datos operativos debe **arrancar desde una perspectiva de negocio**. En este sentido, se preferirán casos de uso que –por su fácil desarrollo– puedan servir como proyectos faro en el arranque, creando un “efecto llamada” a otros participantes, proyectos, e incluso otros sectores que quieran interaccionen y aporten a este espacio de datos de turismo. La metodología se basa en una serie de actuaciones o pasos:

1. **Identificación y oportunidades de negocio.** Para que la compartición a implantar sea efectiva y continuada, se deben dar unas condiciones de contexto sectoriales (económicas, de competencia, normativas, etc.). Se trata por tanto de unir a diferentes (clases de) participantes de la cadena de valor en torno a necesidades comunes (por ejemplo, de tipo regulatorio o de competencia) para crear oportunidades y escenarios *win-win*, sólo resolubles en base a la compartición de datos (o al menos de forma mucho más eficiente).
2. **Definición de los casos de uso** a implementar, que se basa en encontrar ejemplos de las citadas oportunidades. Se deben caracterizar y escoger casos de uso que demuestren de forma práctica el valor de los datos. Para dataspaces en una fase embrionaria, creemos recomendable escoger casos de uso con un número limitado de participantes. El objetivo de este escenario es que sirva para aprender y mejorar las dinámicas del ecosistema, y como proyecto faro. Estos pasos 1 y 2 permiten por tanto el nacimiento del ecosistema del espacio de datos.
3. Las **especificaciones de requisitos** para un caso de uso concreto, dentro de un espacio de datos, se pueden obtener a través de un cuestionario básico. Éste recorre las diferentes dimensiones generales que afectan a la compartición y explotación de datos, y sirve así para levantar una estructura básica para el mismo. Dichos ámbitos son:
 - ➔ **Misión y visión de la compartición y explotación de datos**, donde se define el propósito de tales estructuras y procesos, así como sus principios fundamentales, y cómo darles visibilidad y capilaridad, también de cara a que tengan una sostenibilidad a futuro. Así mismo, es importante abordar la dimensión de los cambios culturales que deben extenderse para que las organizaciones puedan aprovechar el valor latente en los conjuntos de datos, e incluir además la gestión del talento necesario.

- ➔ **Negocio** es el ámbito desde el que se caracteriza el valor a extraer por medio de la explotación de los datos, y se identifican derechos sobre el uso de los mismos. En este sentido, las organizaciones pueden querer desarrollar modelos de evaluación de riesgos, con que evaluar hasta dónde profundizar en el uso de los datos, y qué oportunidades están dispuestas a explorar en profundidad. Este ámbito explora por tanto el contexto y objeto de negocio en base a la compartición de datos, detalla las principales partes interesadas y sus funciones, alcance y normas de negocio aplicables.
- ➔ Desde **Normativa** se abordan temas como los principios contractuales que regulan los casos de uso del espacio de datos, incluyendo los derechos sobre datos, pero también las responsabilidades y los mecanismos de gobernanza de la compartición y explotación de los conjuntos de datos. Los espacios de datos son entornos que aspiran a desarrollar un cumplimiento (de reglas) fundamentalmente en base a tecnología, no sólo en base a contratos, pero –más generalmente– este ámbito considera también el acuerdo constitutivo del propio espacio de datos, la adhesión y cese del mismo. También se especificarán todas aquellas condiciones de contorno legislativas de aplicación a los casos de uso del espacio, así como cualquier condición o normas particular relevante de los mismos, incluyendo los marcos de certificación y las potenciales sanciones.
- ➔ El ámbito de **Funcionalidades** agrupa todas las capacidades que deben existir en un espacio de datos, de cara a cumplir con las expectativas de sus participantes. Esto incluye tanto los servicios de datos que deben ofrecerse (clave para asegurar la liquidez del dataspace, y por ende su utilidad), como los aspectos de seguridad y soberanía, o las opciones con que asegurar la privacidad que se demanda. También es importante definir la capilaridad del ecosistema de datos que se hace necesario (que irá ligado a las herramientas y los mecanismos de marketing, en el ámbito de *Operaciones*), o el modelo de interacción y experiencia de usuario que se da en el espacio federado.
- ➔ El **ámbito de datos** es uno de los dos tecnológicos. En éste se abordan dinámicas y características propias de los conjuntos de datos a transformar y explotar en los casos de uso, e incluye políticas de gobierno y gestión del dato, de calidad de datos, modelos de datos, procesos y frameworks de Ciencia de Datos y Machine Learning, y también aspectos de consumo de la información (reportes, cuadros de mandos), y su sostenibilidad y evaluación continua (à la *DataOps*). Estas dimensiones abordan tanto el formato y la estructura óptima para los datos y los metadatos asociados, qué tipos de datos se van a compartir (e.g. abiertos, propietarios, confidenciales) y por ende el tipo de permisos o restricciones a aplicarles, y el acceso a los mismos, y sus transformación y distribución.
- ➔ **Tecnología** es el ámbito de IT, donde definimos la arquitectura de los sistemas, la infraestructura a emplear y las soluciones comunes, el posible desarrollo de módulos y sus respectivos roadmaps y recursos, así como la administración y el soporte para cualquier tecnología que deba implementarse como parte del ciclo de vida de los datasets datos.

También incluye los aspectos técnicos de los interfaces de comunicación y APIs, las soluciones de gestión de identidad y acceso e identidad (en base a los criterios de seguridad y privacidad establecidos en el ámbito de *Negocio*) y cualquier otro aspecto técnico relativo a la escalabilidad del sistema, el almacenamiento y disponibilidad de herramientas informáticas, control de versiones, y la monitorización.

➔ **Operaciones** es el ámbito que da coherencia práctica al resto de dimensiones. Se incluye tanto el proceso operativo de certificación de componentes y participantes (según el modelo acordado) como la descripción de los Service Level Agreements (SLAs: acuerdos de niveles de servicio) por parte de los proveedores, o la gestión de incidencias y de cambios en la operativa de los casos de uso. También es importante contar con mecanismos para la resolución de posibles disputas, así como de elementos de marketing para la ampliación del ecosistema y la búsqueda de financiación de iniciativas. Por último, se dispone también de un kit de herramientas para la operativa diaria del espacio de datos, como puede ser aplicaciones informáticas de diverso tipo.

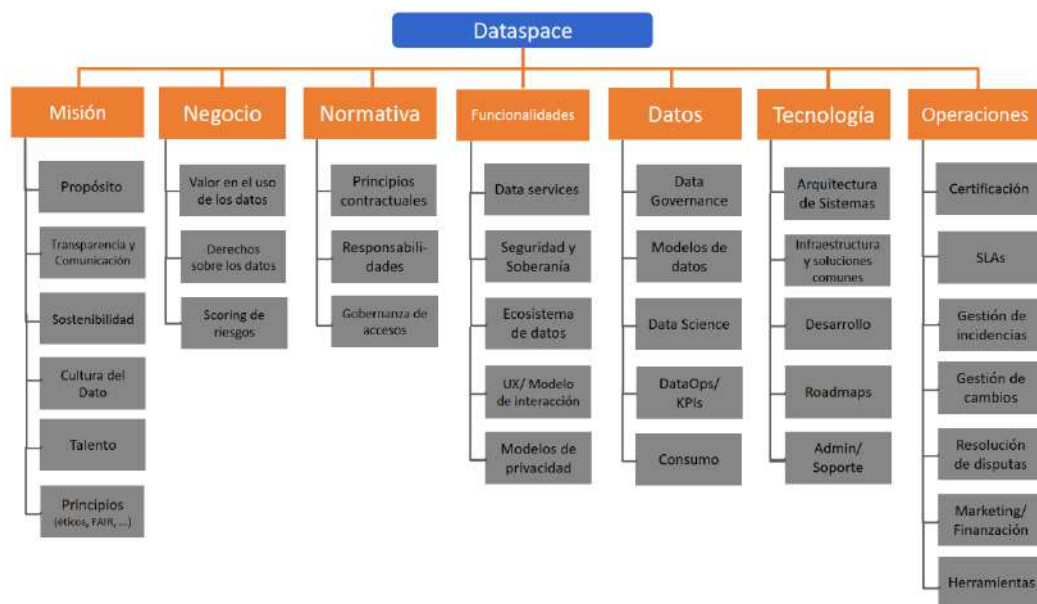


Figura 6. Ámbitos generales en la gestión, compartición y explotación de datos

De esta manera, se busca que los participantes de cada caso de uso consensuen una serie de patrones para gobernar y asegurar una involucración mantenida en el tiempo, garantizando así la sostenibilidad del espacio de datos. Estos requisitos sirven de base para la definición e instanciación de roles, y también influenciarán en los requisitos técnicos que siguen.

4. **Instanciación de la arquitectura** de referencia. Es el proceso práctico que da forma a una arquitectura concreta para el espacio de datos, en base al mapeo entre los requisitos del paso 3 (los criterios en torno a la compartición y explotación de datos a la que se aspira), y el *framework conceptual* con que desplegar espacios de datos (sección 6). Puesto que dicho instrumento incluye también un modelo tecnológico, éste se emplea en el paso 5.
5. La selección y/ o **programación de elementos técnicos** es el paso con que desarrollar e integrar los componentes de software que dan forma a la arquitectura escogida, así como otras posibles actuaciones en materia de preparación y gestión del dato.
6. Por último, con la **integración de los participantes** al espacio de datos, se validan las funcionalidades del caso de uso, y se **arranca la operativa** del mismo. Se dispone entonces de un espacio de datos en funcionamiento, gestionado por un operador que se rige por las reglas definidas, y cuyas gestiones habituales se enmarcan en acciones tipo como:
 - ↗ Alta/ baja de participantes en el ecosistema, así como atención de incidencia.
 - ↗ Certificación de componentes con el sello de garantía arquitectónico.
 - ↗ Confirmación de identidades.
 - ↗ Desarrollo e integración de *data apps*.
 - ↗ Creación de nuevos *brokers* o catálogos de datos.
 - ↗ Creación de nuevos vocabularios o semánticas de datos.
 - ↗ Definición y aceptación o rechazo de pólizas bilaterales o multilaterales.

Llegados a este punto, se inicia un proceso cíclico de cara a incorporar tanto a participantes adicionales (a casos de uso existentes), como actualizaciones de estos, o casos de uso de nuevo diseño y desarrollo. Este proceso evolutivo da forma a un espacio de datos que aspira a gozar de utilidad para el sector y sus agentes, y por ende sostenibilidad.

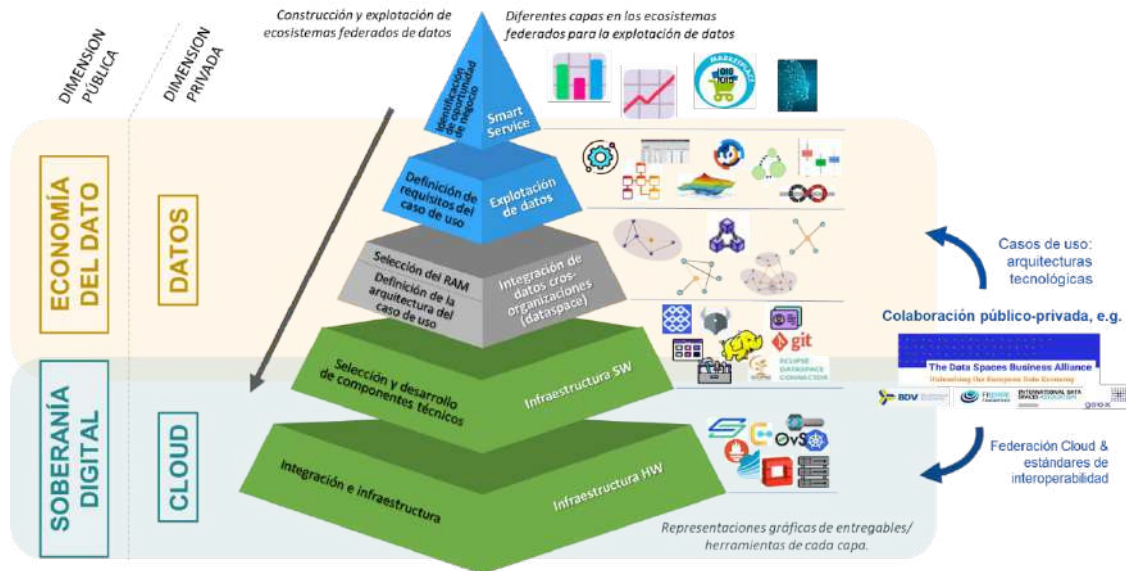


Figura 7. Metodología para desarrollar espacios de datos, con la cobertura tecnológica asociada a cada paso de la misma

En la Fig. 7 observamos una representación gráfica de la metodología explicada arriba. Concretamente, la cara Oeste de la pirámide describe los pasos de ésta, desde la “Identificación de una oportunidad de negocio” en el ápice, a “Integración de participantes e infraestructura del espacio de datos” en la base. La cara Este relaciona dichos pasos con la cobertura tecnológica que le aplica a cada uno de ellos (en una relación 1-a-1), y cuyos elementos técnicos coinciden con el eje vertical de la gráfica en la Fig. 3 de la participación de iniciativas europeas a lo largo de la cadena de valor de compartición y explotación del dato. Cada paso metodológico tiene como *output* un entregable de cobertura tecnológica. Por ejemplo, la identificación de una oportunidad de negocio tiene como resultado eventual (tras abordar los diferentes pasos de la metodología) un *Smart Service* con un impacto directo en las estrategias de dicho Negocio impulsor.

Así mismo, en la figura se observa también una superposición de dichos entregables sobre los ejes de la [Estrategia Europea de los Datos](#). Ésta tiene como focos principales tanto el fomento de la *Economía del Dato*, como vehículo para impulsar la innovación en base a la analítica de datos y la Cultura Data-Driven, como incrementar la *Soberanía Digital*, con la que asegurar que dicha innovación no cree dependencias económico-políticas de estados extranjeros. Por último, el gráfico muestra también el área de influencia de la *Data Space Business Alliance*, cuyas organizaciones miembro se desenvuelven dentro de ámbitos que abarcan desde las capas de datos a las de servicios y tecnologías cloud.

