



Administració Oberta  
de Catalunya

# PLIEGO DE PRESCRIPCIONES TÉCNICAS DEL CONTRATO DE LOS SERVICIOS DE LA PLATAFORMA DE DATOS DE LAS ADMINISTRACIONES LOCALES DE CATALUÑA

Se hace constar que se trata de una traducción automatizada y que, en caso de discrepancia, prevalece la versión catalana.

Realizado por: Consorci AOC

Versión: 2.0

Fecha: 03/06/2024

Archivo: 2023 Contrato Plataforma Datos Locales NGEU - Pliego Tècnic.docx

**Número de expediente: AOC 2024 47**

## Índice

1	Introducción.....	4
2	Contexto.....	5
2.1	Situación actual.....	5
2.2	Marco Normativo.....	5
2.3	Infraestructura de datos del Consorci AOC.....	7
3	Objetivos, metas y calendario.....	8
4	Objeto del contrato.....	9
4.1	LOTE 1: Servicios de la plataforma de datos.....	10
4.2	LOT2: Consultoría y explotación de negocio.....	11
5	Actividades y funciones de la empresa contratista.....	13
5.1	Actividades y funciones lote 1.....	13
5.2	Actividades y funciones lote 2.....	13
6	Finalidades y objetivos a alcanzar.....	15
6.1	Finalidades y objetivos del lote 1.....	15
6.2	Finalidades y objetivos del lote 2.....	16
7	Requisitos generales obligatorios.....	18
7.1	Requisitos generales obligatorios para todos los lotes.....	18
7.1.1	Normativa vigente de obligado cumplimiento.....	18
7.1.2	Requisitos referentes a protección de datos personales.....	18
7.2	Requisitos obligatorios LOT1.....	18
7.2.1	Generales y de arquitectura.....	19
7.2.2	Ingesta.....	20
7.2.3	Almacenamiento.....	21
7.2.4	Virtualización del dato.....	22
7.2.5	Inteligencia de negocio.....	23
7.2.6	Analítica avanzada.....	23
7.2.7	Monitorización y administración.....	24
7.2.8	Publicación e interoperabilidad.....	24
7.2.9	<i>Data warehouse</i> del Consorci AOC.....	24
7.2.10	Gobierno del dato.....	25
7.2.11	Descripción de medios personales.....	27
7.3	Requisitos obligatorios LOT2.....	28
7.3.1	Gobierno del dato.....	28
7.3.2	Marco de trabajo.....	28
7.3.3	Fases.....	28
7.3.4	Detección y estudio de casos de uso.....	28
7.3.5	Servicios de desarrollo e implementación de casos de uso.....	29

7.3.6 Servicio de Capacitación .....	33
7.3.7 Seguimiento .....	33
7.3.8 Casos de uso: prioridades, alcance y ámbitos .....	34
7.3.8.1 Priorización de los casos de uso .....	34
7.3.8.2 Alcance de los casos de uso .....	34
7.3.8.3 Ámbitos de los casos de uso .....	34
7.3.8.4 Ejemplo de caso de uso: Índice de Madurez Digital .....	35
7.3.9 Descripción de medios materiales .....	36
7.3.10 Descripción de medios personales .....	36
8 Formas de seguimiento y control de la ejecución de las condiciones .....	38
9 Aspectos de seguridad y ENS .....	39
Anexo I: Descripción Infraestructura de datos actual de la AOC .....	40
9.1 Descripción de las fuentes de datos de origen .....	41
9.1.1 Bases de datos gestionadas por el Consorci AOC .....	41
9.1.2 Bases de datos gestionadas por otras administraciones .....	41
9.1.3 Bases de datos gestionadas por proveedores externos del Consorci AOC ....	42
9.2 Instancias ETL del Consorci AOC .....	42
9.2.1 ETL SAP Dataservices para el DW estadístico de los servicios AOC .....	43
9.2.2 ETL SAP Dataservices para datos abiertos, y ciudadanos .....	44
9.3 Bases de datos y repositorios de destino .....	47
9.4 Aplicaciones que acceden a las bases de datos de destino .....	47
Anexo II: Datos disponibles del mundo local actualmente .....	55

# 1 Introducció

---

El contenido de estas prescripciones técnicas deriva del proyecto "Plataforma común de gestión y analítica de datos por las administraciones catalanas" con código C11. I03. P04. S14, aprobado en el marco del Mecanismo de Recuperación y Resiliencia (MRR), regulado por el reglamento 2021/241 del Parlamento Europeo y del Consejo de 12 de febrero de 2021.

Con la mera presentación de su oferta, la empresa licitadora acepta las prescripciones técnicas establecidas en este pliego.

Cualquier propuesta que no se ajuste a los requerimientos mínimos establecidos en este pliego quedará automáticamente excluida de la licitación.

## 2 Contexto

---

### 2.1 Situación actual

Las administraciones públicas todavía no tienen una cultura y una estrategia de gobierno del dato. Además, en Cataluña hay 2.250 organismos públicos que generan datos y que tienen la obligación de cumplir la Directiva (UE) 2019/1024, relativa a los datos abiertos y la reutilización de la información del sector público en formato reutilizable. De todos estos organismos, 200 son del sector público de la Generalitat de Catalunya, 1800 son entidades locales y el resto del sector universitario y otras.

### 2.2 Marco Normativo

La Inteligencia Artificial (IA) se ha convertido en un área de importancia estratégica con potencial para ser un motor clave del desarrollo económico, y tiene un amplio abanico de posibles implicaciones sociales. Como parte de su estrategia de mercado único digital, la Comisión Europea presentó en abril de 2018 una estrategia europea sobre IA en su Comunicación "Inteligencia Artificial per a Europa" COM(2018)237. Los objetivos de la estrategia europea anunciados en la comunicación son:

- Potenciar la capacidad tecnológica e industrial de la UE y la adopción de la IA en toda la economía, tanto en el sector público como privado
- Prepararse para los cambios socioeconómicos provocados por la IA
- Garantizar un marco ético y legal adecuado.

Posteriormente, en diciembre de 2018, la Comisión Europea y los Estados miembros publicaron un "Plan Coordinado sobre Inteligencia Artificial", COM (2018)795, sobre el desarrollo de la IA en la UE.

En el año 2020, el Joint Research Centre (JRC) de la Comisión Europea, en colaboración con la Dirección General para las Redes de Comunicación, Contenidos y Tecnología (DG CONNECT) publicó un informe que propone una definición operativa de inteligencia artificial, con el objetivo de supervisar el desarrollo, su absorción e impacto en Europa. El estudio define "los sistemas inteligencia artificial (IA) como los sistemas de software (y posiblemente también de hardware) diseñados por humanos que, dado un objetivo complejo, actúan en la dimensión física o digital percibiendo su entorno a través de la adquisición de datos, interpretando los datos estructurados o no estructurados recogidos, razonando sobre el conocimiento obtenido o procesar la información derivada de estos datos y decidir las mejores acciones a llevar a cabo para alcanzar el objetivo. Los sistemas de IA pueden utilizar reglas simbólicas o aprender un modelo numérico, y también pueden adaptarles comportamiento mediante el análisis de cómo el entorno se ve afectado por sus acciones anteriores".

Teniendo en cuenta las implicaciones que supone el uso de las tecnologías que se pueden meter en esta definición, la Comisión Europea está trabajando en dos propuestas de regulación para definir las reglas de su uso en el ámbito de la Unión. Son:

- Propuesta (COM/2021/2061) de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión. El objetivo de la propuesta es mejorar el funcionamiento del mercado interno mediante el establecimiento de un marco jurídico uniforme, en particular en cuanto al desarrollo, comercialización y utilización de la inteligencia artificial de conformidad con los valores de la Unión. Persigue diversas finalidades imperiosas de interés general, como asegurar un nivel elevado de protección de la salud, la seguridad y los derechos humanos, y garantiza la libre circulación transfronteriza de bienes y servicios basados en la IA, de manera que impide que los Estados miembros impongan restricciones al desarrollo, la comercialización y la utilización de sistemas de IA, salvo que este Reglamento lo autorice

expresamente. La propuesta de reglamento establece normas para regular la introducción en el mercado y la puesta en servicio de determinados sistemas de IA, lo que garantizará el buen funcionamiento del mercado interior y permitirá que estos sistemas se beneficien del principio de la libre circulación de bienes y servicios.

- Reglamento (UE) 2023/2854 del Parlamento Europeo y del Consejo, de 13 de diciembre, sobre normas armonizadas para un acceso justo a los datos y su utilización (Ley de datos). Sus objetivos son:
  - Facilitar el acceso a los datos y la utilización por parte de los consumidores y las empresas, preservando al mismo tiempo los incentivos para invertir en formas de generar valor a través de los datos.
  - En aquellas situaciones en las que exista una necesidad excepcional de obtener datos, prever el uso de los datos que estén en poder de las empresas por parte de los organismos del sector público y de las instituciones de la Unión.
  - Facilitar el cambio entre servicios en la nube. El acceso a servicios de tratamiento de datos competitivos e interoperables es una condición previa para una economía de datos floreciente, en la que los datos se puedan compartir fácilmente dentro de los ecosistemas sectoriales y entre ellos. El nivel de confianza en los servicios de tratamiento de datos determina la aceptación de los servicios mencionados por parte de los usuarios en todos los sectores de la economía.
  - Establecer salvaguardas contra la transferencia ilegal de datos sin notificación por parte de los proveedores de servicios en la nube.
  - Prever el desarrollo de normas de interoperabilidad para la reutilización de los datos entre sectores, con el fin de eliminar los obstáculos al intercambio de datos entre espacios comunes europeos de datos, de acuerdo con los requisitos de interoperabilidad sectoriales, y entre otros que no entren en el ámbito de aplicación de un espacio común europeo de datos específico.

En el estado español, y en el marco de la Estrategia Nacional de Inteligencia Artificial, la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación, en su artículo 23 anima a las administraciones públicas a utilizar esta tecnología. En concreto:

1. Las administraciones públicas favorecerán la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones públicas tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, siempre que sea factible técnicamente. En estos mecanismos se incluirán el diseño y los datos de entrenamiento y abordarán el potencial impacto discriminatorio. Para alcanzar este fin, se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio.
2. Las administraciones públicas, en el marco de sus competencias en el ámbito de los algoritmos involucrados en procesos de toma de decisiones, priorizarán la transparencia en el diseño y la implementación y la capacidad de interpretación de las decisiones adoptadas por los mismos.
3. Las administraciones públicas y las empresas deben promover el uso de una inteligencia artificial ética, fiable y respetuosa con los derechos fundamentales, siguiendo especialmente las recomendaciones de la Unión Europea en este sentido.
4. Se promoverá un sello de calidad de los algoritmos.

Finalmente cabe mencionar la Estrategia de Inteligencia Artificial de Cataluña, aprobada por Acuerdo de Gobierno de 18 de febrero de 2020, y que manifiesta entre sus objetivos:

- Impulsar un modelo de gobernanza transversal que apoye el desarrollo de un ecosistema de inteligencia artificial coordinado y conectado con el mundo.
- Disponer de las infraestructuras necesarias para el desarrollo de la inteligencia artificial y facilitar el acceso seguro a los datos públicos y privados.

- Impulsar la incorporació de la intel·ligència artificial com a tractor d'innovació en la administració i en sectors estratègics com el agroalimentari, la salut i el benestar, l'educació, el medi ambient, la mobilitat, el turisme, la cultura i la indústria, entre altres.

## 2.3 Infraestructura de datos del Consorci AOC

El Consorci AOC ja disposa d'un Data Warehouse, basat en tecnologia SAP, com a magatzem de dades que permet la publicació i anàlisi de dades relacionats amb els serveis que ofereix al conjunt d'Administracions Públiques Catalanes. La solució actual permet recollir les dades provinents dels diferents serveis i les deixa en un magatzem comú des del qual es poden publicar com a dades obertes i quadres de comandament. El sistema actual d'adquisició de dades i publicació inclou les següents eines:

- Microsoft Power BI com a sistema de generació de quadres de comandament
- CKAN com a sistema de publicació de les dades obertes que el Consorci AOC ofereix
- Socrata com a sistema de publicació de dades obertes de la Generalitat de Catalunya

La descripció detallada de la plataforma de dades del Consorci AOC es troba en l'Anexo d'aquest plec.

### 3 Objetivos, metas y calendario

---

El objetivo es crear una plataforma de datos para las administraciones públicas locales de Cataluña para poner en valor los datos abiertos a la sociedad e impulsar iniciativas basadas en el gobierno de los datos para desarrollar soluciones que ayuden a mejorar la planificación, la predicción y la toma de decisiones. Este espacio debe permitir desarrollar una nueva generación de servicios innovadores que faciliten la personalización y la proactividad, con una visión transversal e integrada de la administración pública.

Concretamente, habrá que alcanzar las siguientes metas y calendario relacionadas con los servicios de la plataforma:

1. Servicios de la plataforma de datos abiertos y del *data lake* del Consorci AOC y de las administraciones locales a las que presta servicio, federada con la de la Generalitat de Catalunya.
  - a. Inicio de los servicios de la plataforma de datos abiertos
    - i. Calendario: 4 meses desde la formalización del contrato
  - b. Inicio de los servicios del *data lake* del Consorci AOC
    - i. Calendario: 6 meses desde la formalización del contrato
  - c. Ingesta de nuevos conjuntos de datos en la *data lake* del Consorci AOC
    - i. ingesta de nuevos conjuntos de datos con una metodología iterativa y con una frecuencia de 3 meses, condicionado a su disponibilidad.
2. Inicio de los servicios de *data warehouse* para el Consorci AOC.
  - a. Calendario: 6 meses desde la formalización del contrato
3. Inicio de los servicios de gobernanza de la plataforma de datos del mundo local.
  - a. Calendario: 6 meses desde la formalización del contrato
4. Servicios de analítica avanzada de datos por el mundo local.
  - a. Puesta en marcha de las soluciones de analítica avanzada: 6 meses desde la formalización del contrato
  - b. Implantación de los casos de uso de analítica avanzada: entrega de nuevos casos de uso con una metodología iterativa y con una frecuencia máxima de 3 meses.

Estos hitos se detallan y desarrollan en detalle en el apartado 6 de estos pliegos. El licitador de cada lote debe hacer una propuesta de plazos para alcanzar estas metas y objetivos.

## 4 Objeto del contrato

Este pliego tiene por objeto establecer las prescripciones técnicas particulares que regirán los servicios de la Plataforma de Datos de las Administraciones Locales de Cataluña, federada con la Plataforma Transversal de Datos de la Generalitat de Catalunya, para crear una infraestructura y servicios que permita a todas las administraciones catalanas mejorar su gestión pública en base a los datos, en los ámbitos de la planificación, la toma de decisiones y la automatización de tareas.

Con la realización del objeto contractual mencionado, el órgano de contratación pretende cubrir las necesidades y/o funcionalidades descritas a continuación.

Actualmente, el Consorci AOC se encuentra en un proceso de transformación estratégico 2023-2028, basado en un nuevo modelo de gestión global de las plataformas y tecnologías TIC, siguiendo una estrategia manifiesta de evolución y migración de los servicios hacia el *cloud* público. Este proceso implica una clara hoja de ruta para mejorar la eficiencia y la eficacia de sus servicios gracias, entre otros, a la simplificación operacional, a la aplicación de economías de escala, a la obtención de sinergias entre los diferentes servicios y a la optimización de los recursos. El objetivo principal de este proceso de transformación es alinear el nuevo modelo TIC con las necesidades de la organización y aprovechar las mejoras tecnológicas presentes y futuras que ofrece el *cloud* público para conseguir soluciones más eficientes, flexibles, ágiles, robustas y sostenibles. En este sentido, el Consorci AOC ha apostado por los servicios *cloud* de Amazon Web Services (AWS).

El presente proyecto plantea el desarrollo, mantenimiento y gestión de una plataforma de datos sobre una infraestructura de *cloud* público en AWS, y que implemente un lago de datos (*data lake*) para las administraciones locales, permita el uso de herramientas de analítica avanzada y posibilite procesos de aprendizaje empleando grandes volúmenes de datos (*Big Data*).

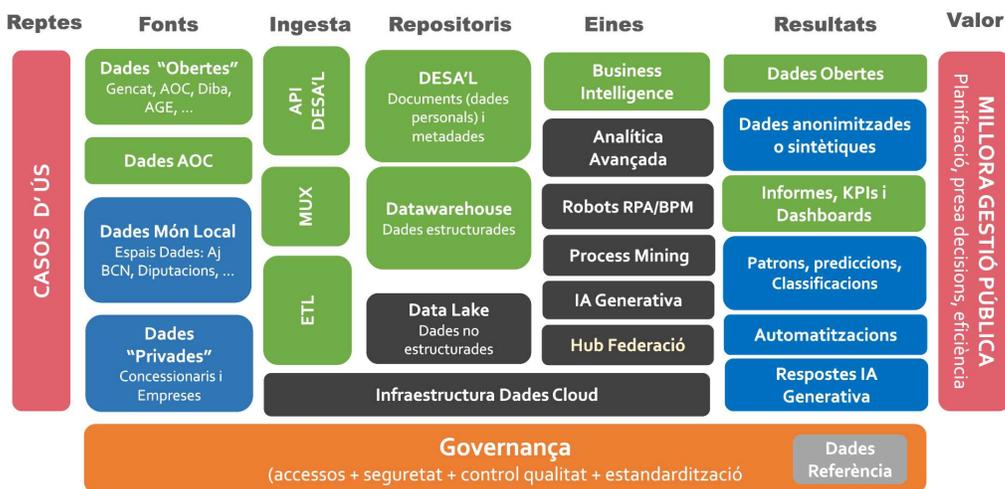


Figura 1. Plataforma de datos del Mundo Local

*Nota: en color verde se indican los módulos donde el Consorci AOC tiene soluciones ya desarrolladas parcial o totalmente*

La plataforma de datos deberá incluir:

- Mecanismos de ingesta que permitan la carga automatizada de datos provenientes de fuentes tanto internas como externas

- Un lago de datos como repositorio de grandes volúmenes de datos en bruto, tanto estructurados como no estructurados
- Entorno de *data warehouse* equivalente al descrito en el anexo I de estos pliegos
- Herramientas de transformación automática de los datos para su publicación y análisis
- Herramientas de indexación, catalogación y búsqueda de los datos
- Mecanismos de control de acceso a los datos, entornos de análisis y cuadros de mando
- Entorno de diseño y publicación de cuadros de mando
- Entorno de publicación de datos abiertos
- Entorno de analítica avanzada de datos con capacidad de tratar grandes volúmenes de datos (Big Data)

La plataforma de datos del mundo local de Cataluña deberá estar federada con la Plataforma Transversal de Datos de la Generalitat de Catalunya y la capacidad de estarlo con otros Espacios que puedan surgir o estar disponibles.

## — Model Federat de Dades

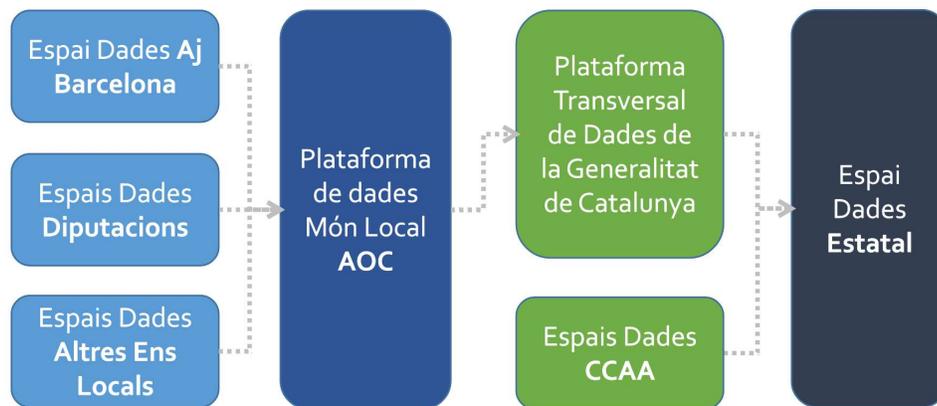


Figura 2. Modelo federado de datos

Teniendo en cuenta las características del proyecto, se propone la contratación de dos lotes, que contemplan por un lado los servicios de la plataforma de datos y por otro la gobernanza, consultoría y explotación de negocio.

### 4.1 LOTE 1: Servicios de la plataforma de datos

Los servicios de implantación, mantenimiento y gestión de la plataforma de datos por las administraciones locales de Cataluña es un proyecto complejo que requiere de una gran cantidad de recursos y conocimientos especializados. Es necesaria la contratación de un servicio especializado en la materia para diseñar, desarrollar, mantener y gestionar la plataforma de forma eficaz y eficiente, cumpliendo sus necesidades específicas.

En base a la plataforma tecnológica propuesta, habría que diseñar e implantar la plataforma de datos, configurar los procesos de recogida y tratamiento que el Consorci AOC actualmente ya tiene implantados y llevar a cabo un proceso continuo de carga de los datos disponibles en el *data warehouse*. En concreto, los servicios de la plataforma de datos deberán incluir:

- Plataforma de datos en base al servicio de *cloud* público Amazon Web Services (AWS).

- Configuración de la ingesta de datos de, como mínimo, las fuentes actualmente configuradas en el *data warehouse* del Consorci AOC.
- Migración de datos de los casos de uso que se vayan identificando y que aportan valor
- Establecimiento de los mecanismos de federación con el resto de plataformas de datos del entorno.
- Adaptación de los cuadros de mando actualmente existentes para integrarlos con el nuevo repositorio.
- Apoyo a metadatos: capacidades que admiten el uso extensivo de metadatos (p. ej., metadatos de uso, registros de transacciones, cargas de trabajo del sistema) para automatizar/mejorar las tareas de integración y explotación de datos.
- Apoyo de gobierno de datos: capacidades que ayudan a los mandatos de gobierno de datos (p. ej., calidad de los datos, linaje de datos) mientras se gestionan los datos para cumplir la integración de datos específicos casos de uso (p. ej., MDM, compartición de datos)

## 4.2 LOT2: Consultoría y explotación de negocio

La Plataforma de datos del mundo local requiere de la definición de un modelo de gobernanza que defina las políticas y protocolos necesarios para la adquisición, gestión y tratamiento de los datos para garantizar su calidad y seguridad. Por su explotación, aparte de la capacidad de disponer de cuadros de mando que permitan su análisis descriptivo, también requiere poder aplicar técnicas más avanzadas de Machine Learning y Deep Learning que permitan inferir conocimiento aplicable en la mejora de la eficiencia de las administraciones.

Un servicio de consultoría puede ayudar a definir e implementar los procesos necesarios para el almacenamiento, la gestión y el análisis de los datos, así como su explotación. La identificación las mejores herramientas y tecnologías para la gestión y tratamiento de los datos, así como diseñar e implementar los flujos de trabajo adecuados para garantizar la eficacia de su gestión es la base para poder disponer de datos de calidad que permitan a las administraciones una mejora continua de su funcionamiento y la automatización de sus procesos.

Por otra parte, se requiere de un conocimiento profundo del funcionamiento de las administraciones públicas y de gestión y analítica avanzada de datos para poder definir casos de uso, muchas veces comunes al conjunto de las administraciones locales, que permitan su análisis y posterior explotación. La publicación de informes y cuadros de mando con los resultados de estos análisis, de pequeñas APIs que permitan el uso de funcionalidades adquiridas durante los estudios, como puede ser el uso de redes neuronales entrenadas con los datos de la Plataforma, y los proyectos de automatización de procesos ayudarán al conjunto de las administraciones a transformar su funcionamiento.

En definitiva, este pliego plantea la contratación de un servicio de consultoría con los siguientes objetivos:

- Definición de la gobernanza del Servicio, ayudando a definir las políticas y los protocolos necesarios para garantizar la calidad y la seguridad de los datos. En concreto, la consultoría ayudará a definir las políticas de gestión de datos, las políticas de seguridad, las políticas de privacidad y las políticas de acceso a los datos. De esta manera se pretende evitar problemas como la pérdida de datos, las vulnerabilidades de seguridad y las infracciones de privacidad.
- Establecimiento de procedimientos para su explotación
- Detección y tratamiento de problemas comunes a todas las Administraciones Locales , Ayudar en la definición e implementación de casos de uso de analítica avanzada de datos y estableciendo mecanismos para su resolución.

En resumen, contratar un servicio de consultoría para la gobernanza y explotación de la plataforma de datos del mundo local debe servir para ayudar a detectar, estudiar y resolver

problemas comunes que pueda tener el conjunto de las administraciones públicas catalanas tanto individualmente como en conjunto. El estudio de casos de uso mediante técnicas de analítica avanzada de datos de Machine Learning y Deep Learning es la vía para extraer conocimiento de los datos que el Consorci AOC y las administraciones locales generan y disponen. Su explotación mediante la publicación de cuadros de mando, informes, pequeñas aplicaciones web o APIs contribuirá a que su funcionamiento sea más eficiente. Mediante la automatización de procesos robóticos, por ejemplo, las organizaciones pueden capitalizar los conocimientos de inteligencia artificial (IA) para agilizar muchas tareas que tradicionalmente se llevan a cabo de forma manual. En este sentido, y para facilitar la explotación del conocimiento adquirido, estará incluido en el alcance de este lote tanto el diseño, desarrollo e integración de la solución de automatización de tareas basada en tecnologías RPA con las funcionalidades que ofrecerá la Plataforma de Datos.

## 5 Actividades y funciones de la empresa contratista

---

Las funciones que debe asumir la empresa contratista son las siguientes:

### 5.1 Actividades y funciones lote 1

- **Diseño de la infraestructura tecnológica:** la empresa contratista debe diseñar la arquitectura y la plataforma de datos del mundo local utilizando la plataforma de *cloud* pública de Amazon Web Services (AWS) del Consorci AOC. Esto implica determinar los requisitos técnicos y funcionales, configurar el entorno de *cloud*, definir la estructura de las bases de datos y establecer los mecanismos de seguridad y protección de los datos.
- **Despliegue de la infraestructura:** la empresa contratista debe realizar todos los trabajos necesarios para el despliegue continuo de la plataforma de datos, incluyendo la instalación y configuración del software necesario, la realización de pruebas de funcionalidad y rendimiento, la implementación de medidas de seguridad, así como la formación del personal de los entes locales sobre el uso adecuado de la plataforma.
- **Servicios de *data warehouse*:** la empresa contratista debe garantizar que la plataforma de datos y, concretamente, el *data warehouse*, contenga, como mínimo, funcionalidades equivalentes a las descritas en el anexo I de estos pliegos. La plataforma debe disponer también de los mismos datos disponibles en el *data warehouse* del Consorci AOC, incluyendo los datos históricos. Los datos deben conservarse de manera precisa y segura durante todo el proceso.
- **Gestión de la infraestructura y explotación tecnológica:** la empresa contratista es la responsable de proporcionar y gestionar toda la infraestructura tecnológica necesaria para la creación y el funcionamiento de la plataforma de datos para el mundo local. Esto incluye toda la infraestructura en la nube relevante para el funcionamiento eficiente de la plataforma.
- **Colaboración con la definición de gobernanza de los datos:** la empresa contratista debe colaborar con la empresa adjudicataria del lote 2 de estos pliegos en la definición de las políticas de gobernanza de los datos. Esto incluye establecer las buenas prácticas para la adquisición, el almacenamiento, la seguridad, el acceso y el uso de los datos en la plataforma. La empresa contratista puede ofrecer orientación sobre los marcos de gobernanza, las políticas de privacidad, las regulaciones de protección de datos y otros aspectos relacionados con la gestión adecuada de los datos.

### 5.2 Actividades y funciones lote 2

- **Definición de la Gobernanza de la Plataforma de datos del mundo local,** estableciendo un catálogo de datos y fuentes de datos actualizable con el tiempo, así como las buenas prácticas para su adquisición, almacenamiento, acceso y uso, en colaboración con el adjudicatario del lote 1.
- **Definición y actualización de un Marco de Trabajo** para el estudio, aprobación, ejecución y explotación de casos de uso de analítica avanzada, empleando las funcionalidades y datos de la Plataforma.
- El adjudicatario del contrato podrá **proponer el uso de fuentes de datos no integradas en la Plataforma** siempre su análisis sirva para mejorar el funcionamiento de las administraciones locales, o del propio Consorci AOC, en el marco de un caso de uso objeto de estudio.
- **Llevar a cabo los casos de uso de analítica avanzada de datos,** ya sea aplicando técnicas de Machine Learning o Inteligencia Artificial.

- **Publicación** de los resultados obtenidos, ya sea en forma de cuadros de mando, informes o pequeñas APIs integrables con las aplicaciones del Consorci AOC y de las administraciones locales.
- Capitalizar los conocimientos obtenidos para agilizar muchas tareas que tradicionalmente se llevan a cabo de forma manual mediante **la automatización de procesos robóticos** (RPA)
- **Formación y capacitación** al personal del Consorci AOC y a los usuarios finales en el uso de los modelos obtenidos durante el proyecto

## 6 Finalidades y objetivos a alcanzar

Las finalidades y objetivos a alcanzar mediante la realización de este contrato son los siguientes:

### 6.1 Finalidades y objetivos del lote 1

Implantación de la plataforma de datos abiertos de las administraciones locales de Cataluña que, a fin de que:

- Integre los datos abiertos de todas las entidades locales y esté federada con la plataforma de la Generalitat de Catalunya, con el objetivo de impulsar la apertura de estos datos, así como de facilitar su reutilización para generar conocimiento y desarrollo en la sociedad.
- Impulse la apertura automatizada de los datos locales ya disponibles en los sistemas de información de gestión, priorizando los datos agregados disponibles por entidades supramunicipales.
- Fomente la reutilización de los datos por parte de administraciones, entidades, empresas, profesionales y ciudadanía.

Creación del *data lake* de las administraciones locales, federado con el de la Generalitat de Catalunya, impulsando una infraestructura en la nube colaborativa y compartida de datos relativos a la actividad de todas las administraciones catalanas, que disponga de:

- Un repositorio común de recogida de datos en bruto, tanto estructurados como no estructurados (*data lake*), con el fin de facilitar su recogida y posterior categorización, consulta y tratamiento.
- Un *data warehouse* como repositorio de datos limpios y estructurados diseñado para su consulta y análisis eficiente.
- Un servicio común de *business intelligence*, que permita el análisis descriptivo de los datos y ayuden a mejorar la planificación, la predicción y la toma de decisiones.
- Un servicio de analítica avanzada de datos, basado en técnicas de *machine learning* y *deep learning* haciendo uso de tecnología *Big Data*, para impulsar servicios personalizados y proactivos, con una visión transversal e integrada de la Administración.

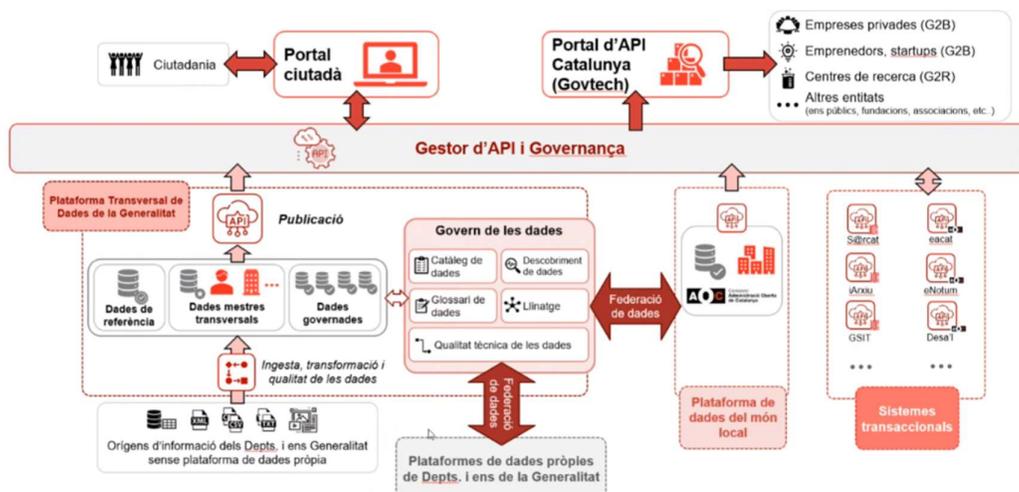


Figura 3. Modelo de federación con Plataforma Transversal de Datos de la Generalitat de Catalunya

Servicios de *data warehouse del Consorci AOC* con funcionalidades, como mínimo, equivalentes a las definidas en el Anexo I de estos pliegos.

La plataforma de datos debe cumplir además los siguientes objetivos generales:

- **Seguridad:** debe tener medidas de seguridad sólidas para proteger los datos y garantizar la privacidad de la ciudadanía.
- **Almacenamiento escalable:** debe ofrecer un almacenamiento escalable y rentable, y permitir a sus usuarios ampliar o reducir la capacidad de almacenamiento según sea necesario.
- **Integración e interoperabilidad:** debe ser compatible y fácilmente integrable con las aplicaciones y sistemas existentes de los entes locales.
- **Automatización:** las herramientas de automatización de las que debe disponer deben permitir reducir la carga de trabajo de los ente para permitirles enfocarse en otras prioridades.
- **Tecnologías avanzadas:** debe ofrecer tecnologías avanzadas para el procesamiento de datos y la analítica para permitir a los entes extraer información valiosa que puedan utilizar para tomar decisiones informadas.
- **Apoyo técnico:** debe proporcionar apoyo técnico y asistencia a los ente cuando surjan problemas o sea necesaria ayuda.
- **Capacidad de recuperación ante desastres:** debe tener medidas de recuperación ante desastres para garantizar que los datos sigan estando disponibles en caso de una interrupción del servicio.
- **Facilidad de uso:** debe ser fácil de usar para garantizar que los ente puedan aprovecharla de manera efectiva sin tener que enfrentar grandes barreras de aprendizaje.
- **Compatibilidad de formatos:** debe ser capaz de gestionar múltiples formatos de datos para alojar una amplia gama de conjuntos de datos.
- **Flexibilidad:** debe ser lo suficientemente flexible para adaptarse a diferentes necesidades y requerimientos de los ayuntamientos y ciudadanos.
- **Informes personalizados:** debe ser capaz de generar informes personalizados según los requisitos específicos de los ente.
- **Federación:** debe permitir el intercambio de información y conocimientos entre diferentes ente para fomentar el desarrollo colaborativo.
- **Movilidad:** debe ser usable desde dispositivos móviles para permitir el acceso a la información desde cualquier lugar y en cualquier momento.
- **Disponibilidad:** para garantizar que la información esté disponible en todo momento, debe tener una alta disponibilidad y capacidad de escalabilidad.
- **Análisis predictivo:** debe tener funciones de análisis predictivo para identificar tendencias y patrones futuros en la información.

## 6.2 Finalidades y objetivos del lote 2

Definir e implementar la gobernanza del espacio de datos, que deberá incluir:

- La definición de ontologías, diccionarios o datos maestros del proyecto, así como la elaboración de los procedimientos para gestionarlos.
- La definición de las políticas de custodia, calidad y perfilado y linaje de los datos.

- Modelo de gobierno analítico: deberá estandarizar y favorecer la repetibilidad y la automatización de todo el ciclo de vida del desarrollo y despliegue de los modelos analíticos, que dan respuesta a los casos de uso. Para construir este modelo, se deben definir y establecer los procedimientos operativos y elementos aceleradores, que permitan optimizar los tiempos de desarrollo y despliegue de modelos analíticos, así como su monitorización y mantenimiento, una vez desplegados.
- Medidas de seguridad y privacidad: se deben identificar y definir las medidas y políticas de seguridad y privacidad (autorización, acceso, auditoría de los datos, entre otros), a implementar en el lote 1 para dar cumplimiento a los requisitos de seguridad y protección de datos.
- El adjudicatario será responsable de las actualizaciones potenciales que se puedan producir en la planificación para el desarrollo y desarrollo de los casos de uso del proyecto, durante el periodo de duración del contrato. Las actualizaciones deberán ser aprobadas por el Consorci AOC.

## 7 Requisitos generales obligatorios

---

La empresa contratista debe disponer de los suficientes medios técnicos, materiales cualitativos y personales para desarrollar las tareas objeto de este contrato.

La prestación regulada en este pliego debe ajustarse, al menos, a los siguientes requisitos técnicos, sin perjuicio de los parámetros a valorar mediante los criterios de adjudicación establecidos.

### 7.1 Requisitos generales obligatorios para todos los lotes

#### 7.1.1 Normativa vigente de obligado cumplimiento

La empresa adjudicataria deberá tener en consideración los requisitos establecidos por la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación, en su artículo 23 y ya mencionados en este pliego. Por otra parte, siendo el Consorci AOC parte del sector público de Cataluña, se requerirá que la empresa adjudicataria esté certificada en el Esquema Nacional de Seguridad, regulado actualmente por el Real Decreto 311/2022, de 3 de mayo.

Finalmente, y teniendo en cuenta que son normas que aún no están vigentes, el adjudicatario deberá tener presente y considerar lo que dispongan tanto las Propuestas de Reglamento del Parlamento Europeo y del Consejo COM/2022/68 final sobre normas armonizadas para un acceso justo a los datos y su utilización y COM/2021/206 final por el que se establecen normas armonizadas en materia de Inteligencia Artificial.

#### 7.1.2 Requisitos referentes a protección de datos personales

Los tratamientos realizados en la plataforma de datos no implicarán en ningún caso datos de carácter personal. Así, se prevén únicamente dos tipos de fuentes de datos: conjuntos de datos procedentes de procesos de anonimización, y conjuntos de datos que no son personales en origen (como por ejemplo, es en principio la base de datos de geolocalización de antenas de telefonía móvil).

En relación a los conjuntos de datos procedentes de procesos de anonimización, la plataforma de datos deberá revisar la eficacia del tratamiento de anonimización antes de incorporar los datos de forma definitiva. En caso de que se detecte la posibilidad de reidentificación, los datos no se incorporarán a la plataforma.

### 7.2 Requisitos obligatorios LOT1

La especificación de requisitos se han agrupado en base a características o capacidades específicas requeridas por la plataforma de datos objeto de esta licitación. Estos requisitos han sido agrupados en los apartados siguientes, que la empresa adjudicataria garantizará en su propuesta cubrir en conjunto:

- Generales y de arquitectura
- Ingesta
- Almacenamiento
- Virtualización del Dato
- Inteligencia de negocio
- Analítica avanzada
- Monitorización y Administración
- Seguridad

- Publicación e interoperabilidad
- *Data warehouse* del Consorci AOC

### 7.2.1 Generales y de arquitectura

[R1-LOT1] La plataforma de datos debe diseñarse, implementarse, desarrollarse, mantenerse y operar en el *cloud* público de Amazon Web Services (AWS) de que dispone actualmente el Consorci AOC. En este sentido, es el Consorci AOC quien proporcionará y contratará los servicios de infraestructura AWS, no siendo estos objetos del presente contrato.

[R2-LOT1] La plataforma debe ser integral, estando todos sus componentes, productos y/o servicios integrados bajo un modelo de gobernanza común centralizada, y ejecutándose con la máxima eficiencia sobre la infraestructura propuesta. Todos los componentes que formen parte de la plataforma deben ser interoperables entre sí, permitiendo en su conjunto dar una respuesta extremo a extremo, desde la captura e ingesta de los datos hasta la publicación de los servicios analíticos.

[R3-LOT1] Si la solución propuesta por la empresa adjudicataria requiere de software, licencias y/o datos adicionales para cubrir cualesquiera de los requisitos establecidos en este pliego, ésta propondrá su contratación por parte del Consorci AOC, que podrá asumir los costes dentro del concepto "Servicios de aprovisionamiento" contemplado en el presupuesto base de licitación.

[R4-LOT1] Todos los componentes de la plataforma deberán tener soporte empresarial de su fabricante o distribuidor exclusivo, y deberán seguir estándares abiertos o de amplia adopción en el mercado.

[R5-LOT1] Ninguno de los componentes de la plataforma debe encontrarse en proceso de discontinuidad, descatalogación o fin de vida del fabricante. El adjudicatario deberá garantizar la vigencia y soporte (actualización y soporte de funcionalidades y protocolos del producto) de estos componentes durante, como mínimo, la vigencia del contrato.

[R6-LOT1] En el caso de ser necesario, durante la ejecución del contrato, de un cambio en algún componente de la plataforma por motivos de actualización tecnológica, el adjudicatario deberá notificarlo con una antelación mínima de 15 días al Consorci AOC, para poder evaluar el impacto en el proyecto y ser aceptado. En todo caso, el adjudicatario deberá mantener los acuerdos de nivel de servicio establecidos en estos pliegos en todo momento. Los eventuales cambios en los componentes por actualizaciones tecnológicas deberán implicar que éstos deberán sustituirse por componentes que cumplan todos los requisitos estipulados, sin detrimento de las características técnicas, funcionales o de cualquier tipo respecto a los componentes a sustituir.

[R7-LOT1] La plataforma debe ser resistente a fallos, siendo capaz de funcionar en alta disponibilidad y, por lo tanto, no teniendo un único punto de fallo.

[R8-LOT1] El adjudicatario pondrá a disposición dos entornos independientes para la implantación de la plataforma:

- Entorno de producción: entorno real donde se almacenarán los datos y a dónde tendrán acceso todos los usuarios de negocio y científicos de datos para consultar los elementos analíticos y trabajar sobre los mismos.
- Entorno de desarrollo: entorno preparado para que los equipos técnicos puedan desarrollar *scripts*, código, reglas, lanzar pruebas, informes, etc. antes de subirlo al entorno de producción.

[R9-LOT1] Debe ofrecer una gestión de permisos multiusuario y multi entidad versátil en la creación, lectura y modificación de los datos y de los espacios de trabajo. Esta funcionalidad permite a los diferentes usuarios de la plataforma tener diferentes niveles de acceso y controlar los conjuntos de datos disponibles para diferentes entidades. La gestión de permisos debe permitir diferenciar, como mínimo:

- Permisos para una entidad concreta: la plataforma debe permitir a los administradores de entes locales asignar permisos específicos. Esto significa que los usuarios de un ente

determinado tendrán acceso a los datos y funcionalidades relacionados con su propio ente, pero no a los de otro.

- Permisos para un conjunto de entidades: los administradores también deben poder gestionar los permisos para un conjunto de entidades. Esto significa que pueden asignar permisos a diversos grupos, permitiendo a los usuarios de estos grupos acceder y trabajar con los datos compartidos entre estos.
- Permisos para todas las entidades: la plataforma permitirá establecer permisos que afecten a todos los entes. Esto significa que algunos usuarios o grupos pueden tener acceso a todos los datos y funcionalidades de la plataforma, independientemente del ente al que pertenezcan.
- Permisos para todo el público: en algunos casos, puede ser necesario permitir el acceso a determinados datos o funcionalidades para todo el público. La plataforma debe tener la capacidad de gestionar este tipo de permisos, garantizando que los datos accesibles para el público estén adecuadamente protegidos y se limite el acceso no autorizado.
- Permisos para los servicios del Consorci AOC: la plataforma debe ser compatible con los servicios del Consorci AOC, como el servicio de Datos Abiertos. Esto significa que debe proporcionar una integración que permita a los servicios del Consorci AOC acceder y utilizar los datos específicos de la plataforma de manera adecuada, con los permisos y las restricciones establecidas.

[R10-LOT1] La plataforma debe permitir la conexión, como mínimo, con las siguientes fuentes de datos (tanto en origen como en destino), ya estén las mismas alojadas tanto en CPDs privados como en servicios en la nube públicos:

- Bases de datos relacionales: Oracle, MS SQL Server, PostgreSQL, MySQL
- Bases de datos NoSQL: MongoDB
- *Datalakes* de terceros, incluido aquellos que utilicen HDFS
- Ficheros de *log*, de imagen, hojas de cálculo, XML, JSON, CSV, ASCII, ficheros de compresión estándar.
- APIs y Web Services (SOAP, REST, OData)
- Sistemas de ficheros FTP, SFTP y FTPS.
- Repositorios de objetos en la nube: AWS-S3 y Azure Storage.
- Colas de eventos y servicios de *streaming* en tiempo real.
- LDAP / Active Directory

## 7.2.2 Ingesta

[R11-LOT1] La plataforma debe permitir capturar datos de diferentes fuentes de origen en formatos estructurados, semi-estructurados y no estructurados, pudiendo asumir grandes volúmenes de datos con flujos masivos y continuos. Debe permitir procesar la información capturada mediante capacidades de alto nivel de procesamiento, en baja latencia, con tolerancia a fallos, consistencia, gestión de eventos desordenados y con alta disponibilidad a nivel de operación y a nivel de persistencia de la información.

[R12-LOT1] La plataforma debe permitir la carga de datos en tiempo real, próxima a tiempo real o con procesos planificados, haciendo uso de técnicas de *streaming*, *micro-batch* y *batch*, a partir de múltiples fuentes de datos, como las especificadas en el requisito [R10-LOT1].

[R13-LOT1] Debe permitir tanto la carga de datos en bruto para su posterior transformación (ELT) como la realización de diferentes transformaciones previas (ETL). En este sentido, es necesario que disponga de funcionalidades de:

- Construcción de procesos de extracción, carga y transformación (ELT).

- Construcción de flujos de datos y orquestación de procesos, incorporando elementos de flujos de control, preparación de datos y evaluación de calidad.

### 7.2.3 Almacenamiento

[R14-LOT1] La plataforma debe proveer una infraestructura de almacenamiento a gran escala, con la capacidad de gestionar grandes volúmenes de datos, con un flujo masivo y continuo, proveniente de diferentes orígenes. Esta infraestructura de datos (*data lake*), basada en tecnologías *Big Data*, debe permitir el almacenamiento eficiente, en un repositorio centralizado, de datos estructurados, semi-estructurados y no estructurados con fuentes de datos heterogéneas. La infraestructura debe estar orientada a la persistencia de los datos, proporcionando mecanismos de recuperación en caso de quiebra, permitiendo su crecimiento de forma escalable y la conexión y procesamiento de datos, incluso desde otras *data lakes*.

[R15-LOT1] La infraestructura debe tener la capacidad de funcionar como *data warehouse*, sistema de almacenamiento de datos centralizado y orientado a las consultas. Debe permitir la recopilación, el almacenamiento y la gestión eficiente de datos estructurados, como datos tabulares o relacionales. Esto implica la organización de los datos en tablas y la posibilidad de realizar consultas complejas, así como la optimización del rendimiento para un acceso rápido y eficiente a los datos.

[R16-LOT1] En el momento de cargar los datos, se proporcionará a los mismos la mayor granularidad al proceso de captura y carga, con el objetivo de permitir un mayor y mejor análisis. Asimismo, la plataforma debe permitir que los datos se puedan almacenar cifrados.

[R17-LOT1] Los datos almacenados en *el data lake* deben poder ser normalizados y enriquecidos. En este sentido, la plataforma debe incluir funcionalidades de extracción de metadatos, conversión de formatos, aumento, extracción de entidades, agregación, desnormalización o indexación.

[R18-LOT1] El almacenamiento de datos masivos del *data lake* debe incluir, como mínimo, los siguientes sistemas (o de equivalentes):

- Sistemas de ficheros distribuidos (DFS).
- Sistemas de almacenamiento de objetos S3.
- Sistemas de bases de datos relacionales.
- Sistemas de almacenamiento NoSQL con estructuras columnares, basadas en documentos, clave-valor y orientada a indexación.
- Ficheros JSON, datos geoespaciales y grafos.
- Cualquier otro sistema que sea necesario para los componentes propuestos en la plataforma.

[R19-LOT1] Los datos que formen parte del *data lake* deben poder ser accesibles por sistemas externos, mediante interfaces de acceso basadas en estándares abiertos o de amplia aceptación, como APIs REST, JDBC, ODBC o tecnología equivalente.

[R20-LOT1] La plataforma debe soportar los formatos de compresión estándar (gzip, Avro, Parquet, ORC o tecnologías equivalentes) con el fin de conseguir la optimización del espacio utilizado en el almacenamiento de los datos.

[R21-LOT1] La plataforma debe proporcionar almacenamiento permanente para los datos. En este sentido, los datos almacenados no deben perderse con el tiempo y deben poder mantenerse de forma segura y accesible durante largos periodos de tiempo. El almacenamiento permanente es esencial para asegurar la consistencia y la disponibilidad de los datos a lo largo del tiempo, permitiendo el análisis histórico, la modificación dinámica de esquemas, y la generación de tendencias e informes a largo plazo.

[R22-LOT1] La plataforma debe disponer de características transaccionales que permitan el procesamiento y la gestión eficiente de transacciones de datos. Estas características deben garantizar la integridad, consistencia y durabilidad de las operaciones realizadas en torno al *data*

*lake*. En este sentido, es necesario que la plataforma disponga de transacciones atómicas, control de concurrencia, aislamiento de transacciones, recuperación hacia quiebras, registro de transacciones e integridad referencial.

[R23-LOT1] La plataforma debe tener la capacidad de cargar, almacenar y tratar datos de sistemas de sensorización, como los provenientes de las llamadas *smart cities* (ciudades inteligentes), teniendo en cuenta el gran volumen de datos y el flujo en tiempo real que las caracterizan.

#### 7.2.4 Virtualización del dato

[R24-LOT1] La plataforma debe proporcionar una capa de virtualización del dato, como capa lógica entre las herramientas de análisis y explotación de los datos y las tecnologías de almacenamiento. Esta capa debe ofrecer una visión integrada de los datos a las herramientas de análisis y de explotación, con independencia de las fuentes y tecnologías utilizadas en el almacenamiento de los datos. Permitirá también la federación de diferentes bases de datos o *data lakes* sin la necesidad de integrar sus datos en el *data lake* propio de la plataforma, mostrándolas bajo unas vistas y accesos unificados.

[R25-LOT1] La plataforma debe admitir mecanismos de autorización y autenticación integrada de usuarios y de roles con permisos configurables y personalizables, permitiendo:

- Conexión con LDAP / Active Directory o tecnologías equivalentes para la autenticación de usuarios.
- Integración con los servicios de identidad digital orientados a los trabajadores públicos y gestionados por el Consorci AOC.

Adicionalmente, se deberá proporcionar apoyo para:

- Single sign-on (SSO)
- Autenticación de dos factores (2FA)
- Autorizaciones basadas en usuarios y/o roles (RBAC)

[R26-LOT1] La capa de virtualización de datos debe estar integrada con las herramientas de gobernanza de datos de la plataforma, de manera que las funcionalidades solicitadas se cumplan en su totalidad de forma coherente. En particular, la gestión de roles y permisos de la capa de virtualización de datos debe estar sincronizada con la capa de gestión de roles y permisos a la gobernanza del dato, para no duplicar su definición ni generar incoherencias a la hora de definir políticas de acceso a la plataforma. Esta integración debe ser compatible, como mínimo, con las herramientas de gobernanza de la Generalitat de Catalunya con el fin de garantizar la federación de datos. La Generalitat de Catalunya utiliza los productos Axon, EDC e IDQ del fabricante Informatica.

[R27-LOT1] La plataforma debe disponer de una interfaz gráfica de usuario (*GUI*) que permita efectuar de forma rápida y sencilla la conexión a los orígenes de datos y permita combinarlos y transformarlos, pudiendo crear vistas virtuales de los mismos y publicarlos en diferentes formatos para poder acceder a ellos mediante un conjunto gráfico de menús, asistentes, acciones de *drag-and-drop* y vistas de árbol.

[R28-LOT1] La plataforma debe permitir establecer vistas por dominio o contexto de datos que, además, puedan compartir datos entre ellos. El enmascaramiento de los datos se podrá definir a bajo nivel, llegando hasta el nivel de columna de una tabla y en función de reglas de negocio establecidas.

[R29-LOT1] Las herramientas de informes y de inteligencia de negocio, de gobernanza de datos y de publicación e interoperabilidad deben poder acceder a la capa de virtualización para poder obtener de forma ordenada la información federada, mediante interfaces de acceso basadas en estándares abiertos o de amplia aceptación, como API REST, JDBC, ODBC, etc. y poder aplicar políticas de acceso a los datos.

## 7.2.5 Inteligencia de negocio

[R30-LOT1] La plataforma debe permitir a los usuarios, de forma sencilla, autónoma y sin conocimientos de programación, analizar los datos en tiempo real, interactuando directamente con una representación visual de las mismas, que debe incluir:

- Objetos visuales sencillos: tablas, gráficos de barras y de áreas, circulares, etc.
- Objetos visuales complejos donde poder presentar los datos de una forma que ofrezca contexto e información detallada y visual, donde atributos como el color y las formas proporcionen información adicional al dato que se quiera representar.
- Gráficos preconfigurados: mapas de color, burbujas, geomapas, mapas de árbol, gráficos de dispersión, etc.
- Cuadros de mando que agrupen los elementos anteriores.

[R31-LOT1] La plataforma debe tener una integración compatible con Microsoft Power BI, la principal herramienta de informes externa utilizada por el Consorci AOC permitiendo mantener la compatibilidad con los informes actuales. La plataforma debe permitir exportar los datos y los indicadores clave de rendimiento (KPIs) hacia Power BI, asegurando la compatibilidad y la sincronización regular para la visualización y el análisis de los informes generados.

[R32-LOT1] Los informes y cuadros de mando generados deben poder ser distribuidos mediante APIs REST, JDBC, ODBC o tecnologías equivalentes, así como ser exportables en formatos de uso común como PDF, XML, XLS, CSV, HTML y otros.

[R33-LOT1] Los informes y cuadros de mando deben estar optimizados para su consumo en dispositivos móviles de forma interactiva, aprovechando las capacidades nativas de estos dispositivos. Asimismo, la plataforma debe permitir insertarlos de forma sencilla en portales web para su consulta pública y abierta, sin necesidad de identificación. Los informes y cuadros de mando que se publiquen en portales web deben permitir el filtrado y segmentación de datos por defecto de forma configurable, para permitir, por ejemplo, diferentes visualizaciones en función del ente público que los utilice.

[R34-LOT1] La plataforma debe proporcionar una herramienta con entorno gráfico de trabajo que permita:

- Modelar procesos de inteligencia de negocio, análisis y generación de informes basados en los datos almacenados.
- Facilitar el desarrollo rápido de procesos y flujos de datos sin necesidad de disponer de un conocimiento profundo del sistema; el análisis en modalidad autoservicio; el acceso, combinación y preparación de datos y la exploración visual, búsqueda y compartición de patrones.

[R35-LOT1] La plataforma debe hacer recomendaciones a los usuarios que exploren datos y diseñen informes, en forma de sugerencia de visualizaciones generadas de forma automática y dependiendo de la naturaleza de los datos explorados. Asimismo, debe permitir establecer de forma automática correlaciones y clústeres de predicciones sin requerir conocimientos de *machine learning* ni algorítmicos al usuario, y permitir búsquedas y consultas en lenguaje natural.

## 7.2.6 Analítica avanzada

[R36-LOT1] La plataforma debe permitir la toma de decisiones a partir de patrones, clasificaciones y aprendizaje. Para esta toma de decisiones se utilizarán modelos de *machine learning* o *deep learning* tanto en capacidades de aprendizaje supervisado (clasificación, regresión, etc.) como no supervisado (*clustering*, detección de anomalías, etc.). El procesamiento de los datos asociados debe poder ser en modos *streaming*, *batch* y *microbatch*.

[R37-LOT1] La plataforma debe permitir el desarrollo y ejecución de *scripts* y modelos de *machine learning* y *deep learning* con, como mínimo, los lenguajes de programación Python, R y Scala, y haciendo uso de Jupyter Notebooks y/o R Studio.

### 7.2.7 Monitorización y administración

[R38-LOT1] La plataforma debe disponer de una consola visual de administración, que permita realizar acciones de gestión, y de capacidades de monitorización, de forma que se pueda conocer en todo momento el estado de la plataforma, en relación a sus recursos y el estado de sus componentes, incorporando herramientas con interfaz de usuario que faciliten las diferentes actividades de monitorización.

[R39-LOT1] La plataforma dispondrá de, como mínimo, las siguientes funcionalidades de monitorización y administración:

- Monitorización a nivel de servicios, *tenant* y cargas de trabajo.
- Consulta de parámetros habituales como uso de procesador, memoria, parámetros de máquina virtual o contenedor, estado de almacenamiento, indicadores de entrada y salida y tasas de transferencia.
- Configuración de alertas personalizadas para notificar a las personas administradoras cuando determinados indicadores superen los umbrales configurados.
- Configuración del registro de eventos de seguridad para dar cumplimiento a los requisitos de seguridad.
- Acceso a los registros (*logs*) de la plataforma.
- Distribución, de forma óptima y eficiente, de las cargas de trabajo de larga ejecución (como las generadas por las herramientas de ciencia de datos) que requieran de alto tiempo de disponibilidad de recursos del clúster, tanto de forma manual como automática.

### 7.2.8 Publicación e interoperabilidad

[R40-LOT1] La plataforma debe ser interoperable con la plataforma de datos de la Generalitat de Catalunya, aspecto clave para asegurar una gestión eficiente y coordinada de los datos entre diferentes organismos públicos en Cataluña. Para conseguir esta interoperabilidad, la empresa contratista debe disponer de las habilidades y conocimientos necesarios para integrar la infraestructura de datos con la plataforma de datos de la Generalitat de Catalunya. Esto incluye la comprensión de los protocolos, formatos y estándares de comunicación, así como la capacidad para adaptar la infraestructura a las especificaciones de la plataforma. La empresa contratista también debe proporcionar la documentación necesaria y el apoyo técnico para asegurar una integración correcta y el mantenimiento de la interoperabilidad con la plataforma de datos de la Generalitat de Catalunya.

[R41-LOT1] Adicionalmente a las interfaces definidas a requisitos anteriores, la plataforma debe disponer de una herramienta de gestión de APIs que permita a los sistemas de negocio la integración con la plataforma, así como la publicación de conjuntos de datos (*datasets*) accesibles para terceros. Esta herramienta debe permitir la gestión y publicación controlada y segura de las APIs a terceros, y deberá estar integrada con el Gestor de APIs y Gobernanza de la Generalitat de Catalunya.

### 7.2.9 Data warehouse del Consorci AOC

[R42-LOT1] La plataforma debe disponer de las características técnicas y funcionales que permitan dar servicio al *data warehouse* del Consorci AOC (descritas en el anexo I de estos pliegos), incluyendo todas las partes del proceso: Ingesta de diferentes fuentes de origen, transformación y tratamiento de los datos, carga y almacenamiento a nivel estadístico en el *data warehouse*, generación y publicación de informes públicos, publicación en formato de datos abiertos y puesta a disposición de los datos por otros servicios del Consorci AOC.

[R43-LOT1] El adjudicatario será el responsable de configurar la infraestructura, cargar los datos, los procesos y las configuraciones actuales del Consorci AOC, descritas en el anexo I de estos

pliegos y en el requisito anterior, hacia la nueva plataforma de datos reduciendo al mínimo el impacto en los procesos actuales del *data warehouse* actual del Consorci AOC

[R44-LOT1] Teniendo en cuenta la complejidad interna de las aplicaciones y servicios del Consorci AOC y de las bases de datos MS SQL Server que forman parte del *data warehouse* actual, éstas deberán mantenerse, en una primera fase, en bases de datos MS SQL Server en la nueva plataforma de datos. Posteriormente, las bases de datos se podrán migrar a otro sistema de base de datos, si se considera oportuno.

## 7.2.10 Gobierno del dato

[R45-LOT1] La plataforma debe proveer una herramienta de gobernanza de datos, para responder a las necesidades del gobierno de los datos, incluyendo aspectos como las políticas de acceso, la calidad, el perfilado y el linaje de los datos, que deberá permitir:

- Llevar a cabo el perfilado de los datos, permitiendo la limpieza, filtrado y estandarización, así como la obtención de valores de tablas maestras, si fuera necesario.
- Realizar un linaje de los datos, de forma que se pueda conocer el flujo por el que han pasado, desde su origen, y las transformaciones sufridas, antes de ser puestos a disposición para el consumo. Así mismo deberá contemplar el análisis de impacto que permita descubrir dependencias entre objetos.
- Realizar el versionado de los datos, sobre los que se realizan las transformaciones, con el objetivo de poder reutilizar los datos para los diferentes casos de uso que se desplieguen.
- Clasificar y etiquetar los datos.
- Establecer ontologías, catálogo de datos, glosario de términos, diccionarios y datos maestros, que faciliten las transformaciones y validaciones para garantizar la calidad del dato y su estandarización.
- Descubrimiento automático de los datos, de modo que los usuarios puedan conocer qué datos tienen disponibles para su consumo.
- Autodescubrimiento de términos de negocio con los que agilizar la catalogación de activos.
- Facilitar el apoyo a las distintas auditorías, registrando las alertas y la resolución de los problemas detectados en los datos.

[R46-LOT1] Deberá ofrecer una visión completa del uso y la gestión de los datos, que permita configurar la estrategia, políticas, exigencias normativas y estándares que requiera el ecosistema analítico a la hora de poner a disposición de los diferentes usuarios la información de calidad de forma intuitiva, ágil, escalable, interoperable y segura.

[R47-LOT1] Permitirá una configuración y administración de la plataforma en base a los roles, responsabilidades y atributos (como etiquetas) definidos dentro del modelo de gobierno. Se requiere por tanto que la gestión de áreas de negocio, dominios y capas de información se pueda realizar mediante flujos de validación de forma gráfica y configurable tanto a nivel global como a nivel de dominio concretos.

[R48-LOT1] Se podrán definir múltiples dominios o espacios de datos, que contengan la información concreta de la plataforma que se requiera para este dominio, según las políticas de acceso definidas, enmascaramiento de la información necesaria, procesos de transformación, reglas de negocio etc. Asimismo, estos dominios podrán compartir información entre ellos, respetando las políticas de acceso que se definan.

[R49-LOT1] Se tendrá que cubrir las diferentes dimensiones relacionadas con la calidad del dato, tales como la completitud, consistencia, precisión, integridad y la pertinencia de los mismos. La solución deberá aportar funcionalidades que garanticen la calidad de los datos, permitiendo al menos:

- Detección de errores en información cargada.
- Duplicidad de datos y generación de Golden Record.
- Detección de incongruencias en los datos, datos vacíos o datos ilegibles.
- Disponer de reglas de calidad predefinidas (por ejemplo, en cuanto a completitud, formato e integridad, entre otras).
- Disponer de la capacidad de implementación de nuevas reglas de calidad ad-hoc, así como de capacidades relacionadas con la exportación y la reutilización a través de APIs.
- Definir reglas de negocio sin necesidad de disponer de conocimientos expertos en codificación y programación.
- Tratamiento de datos seguros durante el proceso de validación de la calidad (encriptación a curso y almacenamiento).

[R50-LOT1] Habrá que incluir mecanismos de seguridad que centralicen y faciliten la gestión de la seguridad perimetral y gestión de la autenticación, autorización y auditoría. La solución incorporará funcionalidades relacionadas con la seguridad y la gestión de usuarios, permisos y políticas de acceso:

- Herramientas y/o módulos que permitan la anonimización, pseudonimización, el enmascaramiento y el cifrado de los datos.
- Capacidades de integración con sistemas de gestión de la autenticación y autorización como LDAP, SAML, Kerberos, Oauth2 y OpenID Conect, o tecnologías equivalentes.
- Capacidades para realizar una gestión de políticas de autorización y auditoría de acceso usuarios basada en roles y con una alta granularidad en la asignación de permisos, permitiendo incluso asignar estos permisos a nivel de campos, o metadatos específicos.

[R51-LOT1] Deberá incluir un motor de búsqueda avanzada e inteligente (a partir de filtros dinámicos y los algoritmos de búsqueda correspondientes como, por ejemplo, búsquedas facetadas).

Esta búsqueda se podrá realizar con un lenguaje tanto técnico, como de negocio, con el objetivo de poner a disposición del usuario aquella información a la que pueda acceder por su rol y responsabilidades determinadas.

[R52-LOT1] Deberá incluir un repositorio centralizado de metadatos que será accesible vía API para un consumo autogestionado de terceros, entre los que se incluirá:

- Descubrimiento, extracción y asociación automatizada de metadatos técnicos y de negocio.
- Trazabilidad y linaje a nivel de campo.
- Análisis automático y gestión de impacto asociada a los cambios.
- Catálogo de datos y glosario de términos personalizado, organizado por orígenes de información.
- Capacidad de importación de términos a partir de plantillas externas con las extensiones más comúnmente utilizadas (csv, excel y equivalentes).
- Soporte de multitud de conectores a sistemas origen, de forma que permitan la conectividad de amplio espectro.

[R53-LOT1] Deberá permitir la configuración de flujos de trabajo de validación en función del modelo de gobierno que se defina en base a las siguientes actividades:

- Implementación de políticas, procesos y procedimientos del gobierno del dato en función de los roles y acciones que se ejecuten.

- Implementación de workflows de autorización/aprobación de contenidos.
- Sistema de alertas y notificaciones que facilite la personalización, colaboración y la interacción entre los usuarios de la herramienta.
- Trazabilidad y control de los pasos ejecutados para un riguroso seguimiento, control y auditoría.
- Integración con otras soluciones externas a la aplicación, tales como herramientas de ingesta, virtualización, explotación de datos y APIs.
- Incorporación de reglas lógicas de negocio para la automatización de los flujos de trabajo necesarios.
- Personalización y soporte de los roles del Modelo de Gobierno.

[R54-LOT1] Deberá ofrecer la funcionalidad de Informes y/o cuadros de mando para monitorizar la operativa de la herramienta de Gobierno del Dato.

[R55-LOT1] La herramienta de gobierno del dato estará integrada, como mínimo, con la solución de gobernanza de la Generalidad de Cataluña para garantizar la federación de datos. La Generalitat utiliza los productos Axon, EDC e IDQ del fabricante Informatica.

### 7.2.11 Descripción de medios personales

La relación de perfiles del equipo de trabajo y las funciones a desarrollar por cada uno son las siguientes:

#### **Jefe de proyectos:**

- Dirigir, gestionar y coordinar el contrato, los servicios/proyectos incluidos y los equipos.
- Planificar y controlar la prestación y entrega de los trabajos contratados.
- Gestionar los riesgos y los planes de mitigación o contingencia.
- Gestionar la facturación y la financiación del servicio.
- Reuniones de coordinación con los responsables del Consorci AOC.

#### **Arquitecto / Consultor senior:**

- Establecer la visión técnica global de los equipos de desarrollo.
- Definición global de arquitecturas, procesos y modelos de apoyo a la gestión del ciclo del dato (ingesta, transformación, almacenamiento, procesamiento y explotación).
- Garantizar el funcionamiento correcto de la plataforma de extremo a extremo con sus correspondientes planes de pruebas.
- Determinar requisitos de usabilidad y protección de datos
- Análisis funcional.
- Determinar requisitos normativos, técnicos y jurídicos, a aplicar en tratamiento de la información y privacidad.

#### **Desarrollador / Consultor junior:**

- Apoyar al arquitecto de datos / consultor senior y al jefe de proyectos mediante los desarrollos correspondientes.
- Diseñar y desarrollar APIs y procesos ETL.
- Diseñar y desarrollar informes de inteligencia de negocio.

## 7.3 Requisitos obligatorios LOT2

### 7.3.1 Gobierno del dato

[R1-LOT2] El adjudicatario deberá definir la gobernanza del servicio, ayudando a definir las políticas y protocolos necesarios para garantizar la calidad y la seguridad de los datos. En concreto, la consultoría ayudará a definir las políticas de gestión de datos, las políticas de seguridad, las políticas de privacidad y las políticas de acceso a los datos. De esta forma se pretende evitar problemas como la pérdida de datos, las vulnerabilidades de seguridad y las infracciones de privacidad.

[R2-LOT2] El adjudicatario deberá implementar el modelo de gobernanza definido usando la herramienta de gobernanza del dato proveído en el lote 1 utilizando todas las funcionalidades descritas en el apartado 7.2.10, para responder a las necesidades del gobierno de los datos, incluyendo aspectos como las políticas de acceso, la calidad, el perfilado y el linaje de los datos.

### 7.3.2 Marco de trabajo

[R3-LOT2] Al comienzo de los trabajos, el adjudicatario definirá y elaborará un Marco de Trabajo común para abordar el proyecto i y desarrollo de casos de uso, que incluirá la definición de las medidas de seguridad y privacidad, así como los modelos de gobierno del dato y gobierno analítico a implantar y configurar en el Espacio de Datos, dando cumplimiento a la normativa aplicable. Se recogen, en este apartado, los principales aspectos que debe contemplar este Marco de Trabajo, pudiendo organizar o completar el adjudicatario esta relación como considere adecuado.

[R4-LOT2] Entregables del servicio:

- Modelo de gobierno de datos.
- Modelo de gobierno analítico.
- Medidas de seguridad y privacidad a implementar.
- Cualquier otro aspecto que el adjudicatario incluya en su propuesta.

### 7.3.3 Fases

El conjunto de actividades definidas en este apartado para los servicios de desarrollo e implementación de casos de uso se estructurarán en las siguientes fases:

- Fase de evaluación de la Viabilidad, asociada a la ejecución, de manera iterativa, de las actividades necesarias para evaluar la viabilidad de un caso de uso.
- Fase de desarrollo, asociada a la ejecución, de manera iterativa, de las actividades necesarias para el desarrollo, la implementación y el desarrollo de un caso de uso viable.

### 7.3.4 Detección y estudio de casos de uso

[R5-LOT2] Al comienzo de los trabajos, el adjudicatario realizará el conjunto de acciones que sean necesarias para poder tener una primera valoración de los datos gestionados por el Espacio de Datos así como una definición de casos de uso de analítica avanzada. En cualquier caso:

- Se basan en las fuentes de datos gestionadas por el Espacio de Datos, o bien se especifica y se desarrolla su integración.
- Se basan en las capacidades de la solución que han sido proporcionadas.
- Se dispone una certeza razonable, sobre la base del conocimiento ya disponible de los datos, respecto a su viabilidad.

- Para realizar la identificación de nuevos casos de uso, El Consorci AOC podrá pedir que se hagan dinámicas de trabajo con diferentes interlocutores tanto del Consorci AOC como de las Administraciones Públicas Catalanas.

El Consorci AOC deberá aprobar los casos de uso presentados y establecerá su priorización, con el objetivo de optimizar la obtención de resultados, en el menor tiempo posible, contribuyendo a los objetivos de negocio.

[R6-LOT2] El adjudicatario deberá, al menos:

- Definir el conjunto de criterios que empleará para el análisis inicial de casos de uso.
- Realizar un análisis inicial de los casos de uso, en base a los criterios definidos, y obtener una primera valoración respecto a su viabilidad.
- Priorizar los casos de uso, de acuerdo con lo que el Consorci AOC especifique.

En caso de que, como resultados del análisis inicial, se evidenciaran dificultades en la consecución de algún caso de uso, el adjudicatario deberá proponer las reorientaciones que se consideren necesarias, para optimizar la obtención de resultados, en el menor tiempo posible, siempre contribuyendo a los objetivos de negocio.

[R7-LOT2] Como fruto de la primera fase del proyecto, el adjudicatario entregará al Consorci AOC un documento especificando:

- Criterios de análisis y priorización de casos de uso.
- Análisis inicial y priorización de los casos de uso.
- Planificación para el desarrollo y el despliegue de los casos de uso.
- Cualquier otro aspecto que el adjudicatario incluya en su propuesta.

### 7.3.5 Servicios de desarrollo e implementación de casos de uso

[R8-LOT2] El adjudicatario deberá llevar a cabo las tareas asociadas al desarrollo y la implementación de casos de uso de analítica avanzada, a la monitorización y mantenimiento de los modelos analíticos (algoritmos software), que dan respuesta a los casos de uso, y a la capacitación a los usuarios finales, consumidores de los servicios analíticos, derivados de los casos de uso.

[R9-LOT2] Como resultado de cada caso desarrollado, el adjudicatario entregará al Consorci AOC:

- Informes con el conocimiento adquirido de los datos, con información como el listado y volúmenes de datos adquiridos (número de registros y campos por registro).
- Scripts desarrollados para la planificación y sistematización de procesos de análisis.

[R10-LOT2] El adjudicatario deberá garantizar la explicabilidad y la reducción de sesgo en todos los modelos analíticos que se desarrollen en el marco del proyecto, para dar respuesta a los casos de uso, apoyándose en las capacidades que se requieren, a tal efecto, en la capa analítica de la Solución.

[R11-LOT2] El licitador deberá identificar aquellos algoritmos sobre los que el Consorci AOC puede exigir su propiedad intelectual, por no haber sido desarrollados en el marco del contrato. Los modelos analíticos, una vez implementados sobre la Solución y desplegados para utilizarlos, podrán ser consumidos por los entes usuarios de los servicios del Consorci AOC, sin más limitaciones que las capacidades de recursos hardware que pueda tener la Solución.

[R12-LOT2] La consultoría objeto de este lote podrá abordar un máximo de dos casos de uso en paralelo.

[R13-LOT2] En caso de que el adjudicatario proponga la incorporación de fuentes de datos externas a las disponibles desde el Espacio de datos, que sean necesarios o de utilidad para el desarrollo y despliegue de algún caso de uso, se incorporarán al data lake de la Plataforma de Datos objeto del lote 1 de este pliego. Cualquier fuente externa que el adjudicatario proponga incorporar al data lake de la solución, deberá cumplir las condiciones de seguridad y protección de datos con acuerdo a la normativa aplicable.

### **Actividades asociadas al desarrollo y desarrollo de cada caso de uso**

Las actividades que se recogen a continuación se basan en el modelo estándar CRISP-DM (Cross Industrial Standard Process for Data Mining), que establece las fases necesarias para cubrir el ciclo de vida propio de las técnicas de minería de datos, y que son comúnmente aplicadas al desarrollo de modelos analíticos.

Asimismo, la prestación de este servicio deberá basarse en:

[R14-LOT2] El conjunto de mejores prácticas, definidas por MLOps (Machine Learning Ops), a la hora de desarrollar modelos de machine Learning, con el objetivo de la puesta en producción.

[R15-LOT2] Se emplearán metodologías de desarrollo ágil, en concreto se hará uso de SCRUM. Se llevarán a cabo entregas continuas, a lo largo del transcurso del servicio, los cuales deberán ser operativos, iterativos e incrementales, es decir, entre una iteración (SPRINT) y la siguiente deberán introducirse nuevas funcionalidades al desarrollo, hasta completar el alcance del caso de uso.

### **Conocimiento del negocio**

La definición de cada caso de uso objeto del presente lote contendrá como mínimo:

[R16-LOT2] Los objetivos de negocio, los criterios de negocio para la evaluación del caso de uso, los objetivos analíticos, la complejidad, los pasos a seguir y las técnicas analíticas a emplear para el desarrollo y la implementación de cada caso.

[R17-LOT2] Los indicadores y criterios que permitan valorar si se responde con éxito al problema que se quiere resolver. Para ello, se contará con la participación de perfiles técnicos y de negocio del Consorci AOC y, en su caso, de los entes del mundo local.

[R18-LOT2] Elaboración del plan de trabajo y su calendario. Asimismo, se revisará y actualizará la planificación general del proyecto, realizada en el servicio de análisis inicial y priorización.

[R19-LOT2] La definición de cada caso de uso así como del plan de trabajo que lo implementará deberá ser aprobado por el Consorci AOC.

### **Conocimiento, recopilación y descripción de los datos**

[R20-LOT2] Habrá que identificar claramente los datos que serán objeto del caso de uso y su adecuación al procesamiento futuro. Esta tarea implica la elaboración de informes, con la lista de los datos adquiridos y la localización. Si los datos no se encuentran en el Espacio de Datos objeto de esta licitación habrá que especificar también las técnicas utilizadas en la recolección, los volúmenes de datos (número de registros y campos por registro), el significado y el formato de cada campo, etc. En el caso de que sea necesario, y la AOC lo determine, el licitador deberá proveer los datos a cargo del presupuesto definido para esta finalidad.

[R21-LOT2] Se podrá llevar a cabo un análisis exploratorio de los datos objeto de cada caso de uso revelar sus propiedades.

[R22-LOT2] Se efectuarán verificaciones sobre la calidad de los datos para determinar la consistencia de los valores individuales de los campos, la cantidad y la distribución de

los valores nulos, la localización de valores fuera de rango, etc., con el objetivo de asegurar la completa y corrección de los datos.

[R23-LOT2] Se seleccionará el subconjunto de los datos adquiridos en la fase anterior, que esté relacionado con las técnicas de modelado analítico que se aplicarán. En su caso, los datos seleccionados se podrán pre-procesar usando técnicas como:

- Optimizará la calidad de los datos, con el fin de prepararlos para la fase de modelado analítico, mediante técnicas como la normalización de los datos, la discretización de campos numéricos, el tratamiento de valores ausentes o la reducción del volumen de datos, entre otros.
- Se podrán realizar operaciones de preparación de los datos, como la generación de nuevos atributos a partir de atributos ya existentes o la transformación de valores para atributos existentes.
- Se podrán realizar transformaciones sintácticas de los datos, sin modificar su significado, con la idea de permitir o facilitar el uso de alguna técnica de modelado analítico en particular.

### **Modelado analítico**

[R24-LOT2] Se seleccionará/n la/las técnica/es de modelado analítico más apropiada/es al tipo de problema a resolver en cada caso de uso. Previamente a la construcción del modelo analítico, se definirá el método de evaluación a utilizar para determinar la calidad y la validez (en base a parámetros como el rendimiento, fiabilidad, robustencia o explicabilidad, entre otros) del modelo analítico.

[R25-LOT2] Una vez seleccionada/as la/las técnica/es de modelado analítico, se construirá el modelo y se ejecutará sobre los datos previamente preparados. Se pueden generar, en esta fase, uno o más modelos analíticos.

[R26-LOT2] Se interpretarán los resultados del análisis, teniendo en cuenta el conocimiento preexistente y los criterios preestablecidos. Se valorarán factores como la exactitud y la Generalitat de Catalunya del modelo. e) Evaluación de los resultados

### **Evaluación de los resultados**

[R27-LOT2] De forma conjunta con los responsables del servicio del Consorci AOC, se realizará la evaluación del modelo analítico, en relación con los indicadores establecidos, y se determinará si hay alguna razón para la que el modelo sea deficiente, o si se considera óptimo para pasar a la fase de desarrollo. En concreto:

- Se evaluará el proceso completo de modelado, con el fin de identificar elementos que se puedan mejorar.
- Se establecerán los siguientes pasos a la luz de la evaluación de los resultados ante los objetivos. Se evaluará la viabilidad del caso de uso y, previa aprobación del Consorci AOC se decidirá si se han generado resultados satisfactorios y se puede pasar a la fase de desarrollo.
- Si los resultados no son satisfactorios o se detecta alguna oportunidad de mejora, se aborda otra iteración desde la fase de preparación de datos o desde la fase de modelado, teniendo en cuenta los resultados obtenidos.
- En caso de que se detecte la inviabilidad del caso de uso, se propondrá su sustitución por uno de complejidad equivalente y se comenzará con un nuevo proyecto de modelado analítico. Para proponer la sustitución de un caso de uso, el adjudicatario elaborará un informe donde se detallen los trabajos realizados, se determinen las causas por las que el caso de uso no es viable, justificando el número de iteraciones realizadas e identifiquen las condiciones que deberían darse para que el caso de uso pudiera ser viable. El informe deberá ser validado por parte del Consorci AOC.

- Si los resultados son satisfactorios se diseñará y elaborará un plan de implantación, que incluirá las tareas a realizar para la publicación o consumo del modelo analítico (despliegue del caso de uso), que deberá ser aprobado por el Consorci AOC previamente al desarrollo. Los resultados de los modelos analíticos se almacenarán o desplegados en el Espacio de datos, desde donde podrán ser consultados por los usuarios finales, a través de las capacidades de explotación de la solución, o ser consumidos por otros sistemas/servicios.

### **Desarrollo**

[R28-LOT2] El adjudicatario abordará las tareas asociadas al consumo de los modelos analíticos mediante:

- El desarrollo de APIs, que permitan la publicación del modelo analítico, así como el consumo de los servicios analíticos expuestos por otros sistemas de información de la organización.
- La construcción de informes y visualizaciones ad-hoc, que faciliten la presentación de resultados y consumo por parte de los usuarios.

### **Plan de monitorización y mantenimiento del modelo**

[R29-LOT2] Se diseñarán y elaborarán las estrategias de monitorización y mantenimiento del modelo, una vez desplegado, las cuales permitirán valorar si el modelo está funcionando y siendo utilizado apropiadamente, emprender las tareas de mantenimiento que sean necesarias y controlar las diferentes versiones del modelo.

### **Revisión del proyecto analítico**

[R30-LOT2] Se evaluará todo el proyecto de desarrollo y despliegue del caso de uso y se identificarán lecciones aprendidas y acciones de mejora. Se realizará un informe final de conclusiones que cubrirá, para cada caso de uso, al menos, los siguientes aspectos:

- Informes asociados al conocimiento de los datos.
- La construcción y el uso del modelo analítico: proceso de elaboración, fundamentos, resultados, uso, consumo y conclusiones del modelo, entre otros aspectos.
- El plan de implantación.
- El plan de monitorización y mantenimiento del modelo.
- Lecciones aprendidas y acciones de mejora, asociadas a la fase de revisión del proyecto.

### **Capacitación asociada a los casos de uso**

[R31-LOT2] El Adjudicatario ofrecerá un servicio de formación sobre aquellos casos de uso en los que sea necesario impartir capacitación sobre la publicación o consumo del modelo analítico asociado. En estos casos, se llevará a cabo la capacitación dirigida a los usuarios finales o consumidores de los servicios analíticos derivados de cada caso de uso. Esta capacitación se basará principalmente en la elaboración de manuales, guías de usuario o vídeos autoexplicativos, aunque podrá ser necesario recurrir puntualmente a sesiones presenciales.

### **Entregables**

En resumen, el conjunto de entregables asociados a un caso de uso viable serán:

- Plan de trabajo de cada caso de uso.

- Método de evaluación del modelo analítico para determinar la calidad y la validez.
- Datasets empleados en el entrenamiento y la validación de los modelos analíticos.
- Código fuente, desarrollado en el marco del contrato, de los procesos ETL, los algoritmos (modelos analíticos) o las API para consumo de los modelos, entre otros.
- Informe final o de conclusiones, que debe incluir todos los entregables asociados a sus versiones finales.
- Documentación y elementos audiovisuales asociados a la capacitación.

Si se determina la inviabilidad de un caso de uso, los entregables serán:

- Plan de trabajo de cada caso de uso. Informes asociados al conocimiento de los datos.
- Método de evaluación del modelo analítico para determinar la calidad y la validez.
- Datasets empleados en el entrenamiento y la validación de los modelos analíticos.
- Código fuente, desarrollado en el marco del contrato, de los procesos ETL, los algoritmos (modelos analíticos).
- Informe de necesidad de sustitución del caso de uso.

### 7.3.6 Servicio de Capacitación

Este servicio incluye las acciones de capacitación dirigidas a los profesionales del Consorci AOC y usuarios finales de los modelos analíticos desarrollados. Las acciones de capacitación, incluidas en este servicio, deben ser graduales y continuadas durante todo el periodo de ejecución del contrato, y no se debe focalizar sólo en los momentos de desarrollo de los casos de uso.

[R32-LOT2] Las acciones de capacitación seguirán las siguientes directrices:

- Se deben utilizar modelos de capacitación presencial y en línea.
- Para cada grupo de usuarios se definirán actividades formativas específicas, indicando las propuestas de temarios y los perfiles del personal docente que las impartirá, sin perjuicio de que parte del temario que contenga cada actividad formativa pueda ser coincidente.
- La duración y el número de las actividades formativas, la forma de impartición y el número de asistentes mínimos y máximos a las mismas, deben adecuarse a las necesidades del temario de cada una de las propuestas diseñadas por el adjudicatario.
- El adjudicatario realizará una propuesta de cada una de las actividades formativas específicas (Plan de capacitación), que deberá ser aprobada por el Consorci AOC.

#### **Entregables**

Los entregables relacionados con el servicio de capacitación incluirán:

- Materiales de formación elaborados para las sesiones presenciales y de capacitación en línea: presentaciones, casos prácticos o vídeos, entre otros.
- Manuales de usuario final, en su caso.

### 7.3.7 Seguimiento

[R33-LOT2] El adjudicatario deberá llevar a cabo la monitorización y evaluación del rendimiento de los modelos y su mantenimiento durante el periodo de duración del contrato. Se entiende por mantenimiento, la reestimación, el recalibrado o el reentrenamiento de los modelos, aplicando una u otra opción, en función del resultado de la evaluación del rendimiento.

### 7.3.8 Casos de uso: prioridades, alcance y ámbitos

Uno de los objetivos que busca la aplicación de técnicas de aprendizaje automático y analítica de datos en las organizaciones es el de obtener información para mejorar la gestión y su eficiencia, que en última instancia supone un ahorro en tiempo y dinero. En el ámbito de las administraciones públicas, la aplicación del conocimiento obtenido como ayuda para la toma de decisiones y la automatización de procesos puede ser un acelerador en el proceso de transformación digital, en beneficio de toda la ciudadanía.

#### 7.3.8.1 Priorización de los casos de uso

Se priorizarán los casos de uso que aporten resultados a corto plazo, con valor público e impacto para las administraciones locales, a nivel municipal y supramunicipal. En este sentido, se priorizarán los ámbitos donde ya se disponga de datos de interés público para todos los municipios de Cataluña, como por ejemplo el datawarehouse de la AOC y las plataformas de datos abiertos.

#### 7.3.8.2 Alcance de los casos de uso

Los casos de uso que son objeto de la presente licitación deberán aplicar, con criterio general, las siguientes técnicas:

- Estandarización del formato de los datos de las administraciones locales y definición de un catálogo de datos con la finalidad de facilitar su consulta, acceso y estudio.
- Identificar buenas prácticas e iniciativas innovadoras en el ámbito de las administraciones locales
- Descubrimiento de conocimiento (Insights) para mejorar la toma de decisiones y la planificación
  - Descubrimiento de relaciones no conocidas entre variables
  - Identificación de patrones de comportamiento
  - Obtención de modelos predictivos tanto univariantes como multivariantes
  - Detección de anomalías y estudio de su efecto en el comportamiento esperado de los datos
  - Cálculo de indicadores que permitan agrupar el conocimiento contenido en los datos
  - Realizar estudios de clasificación, tanto jerárquicos como no jerárquicos, con el fin de descubrir agrupaciones de individuos y variables según la semejanza en su comportamiento
  - Diseño y entrenamiento de redes neuronales para el desarrollo de sistemas de generación de lenguaje natural especializado en el ámbito de las administraciones locales.
- Desarrollo de pequeñas soluciones que implementen, en forma de aplicación web o de API, el conocimiento obtenido. Por ejemplo en forma de sistemas de clasificación, generación de previsiones, de lenguaje natural o automatización de tareas.
- Automatización de procesos (si procede)

#### 7.3.8.3 Ámbitos de los casos de uso

El proyecto tratará casos de uso de los ámbitos que sean de interés para la mejora de la gestión pública local. Sin hacer un análisis exhaustivo, algunos de los ámbitos de potencial interés son:

- **Servicios dirigidos a la ciudadanía**

- Servicios digitales
- Inclusión digital
- Servicios sociales
- Movilidad
- Gestión del espacio público
- Turismo
- Comercio
- Medio ambiente y clima
- Gestión residuos y selectiva
- Agua y energía
- Seguridad
- Redes sociales
- Atención ciudadana
- **Gestión interna de la Administración**
  - Transformación digital interna
  - Presupuesto
  - Recursos humanos
  - Planificación de servicios, infraestructuras y equipamientos
  - Tramitación: plazos y eficiencia

El análisis de los datos de estos ámbitos se realizará con visión municipal y también supramunicipal: comarcal, provincia y Cataluña.

#### **7.3.8.4 Ejemplo de caso de uso: Índice de Madurez Digital**

Por poner un ejemplo, y en la línea de encontrar indicadores que faciliten la toma de decisiones, el Consorci AOC ya ofrece al conjunto de las Administraciones Públicas Catalanas el Índice de Madurez Digital (IMD). es un índice compuesto que evalúa el grado de transformación digital de las administraciones locales catalanas. Se calcula mediante la suma ponderada de más de 30 indicadores organizados en tres dimensiones: derechos digitales, actividad de los servicios digitales y gobierno abierto.

Se trata de una iniciativa única e innovadora, tanto a nivel estatal e internacional. Hay una gran cantidad de índices que miden el desarrollo del gobierno digital, como el Índice de Economía y Sociedad Digital (DESI) de la CE, el Índice de Desarrollo de Gobierno Electrónico (EGDI) de la ONU o el e-Government Benchmark de la CE, pero todos están definidos a nivel de país y no se pueden aplicar a nivel regional o local. Además, se basan en encuestas o datos estadísticos, mientras que el IMD utiliza datos públicos y objetivables.

El conjunto de datos del IMD contiene una serie histórica de más de 30 indicadores relacionados con el gobierno digital de 988 entidades locales catalanas. Los datos del IMD son permanentemente actualizados y disponen de un amplio número de metadatos, incluyendo la descripción de cada uno de los indicadores utilizados.

Un caso de uso a cubrir durante el contrato sería el de mejorar, con los datos disponibles y empleando técnicas de analítica avanzada de datos este índice con el fin de garantizar su ajuste a lo que precisamente se quiere medir.

### 7.3.9 Descripción de medios materiales.

Las empresas licitadoras podrán proponer la consulta a fuentes de datos que impliquen un coste, siempre que estas consultas aporten valor a los objetivos que establece el proyecto y su importe esté dentro del presupuesto definido a esta finalidad.

### 7.3.10 Descripción de medios personales

La relación de perfiles del equipo de trabajo y las funciones a desarrollar por cada uno son las siguientes:

#### **Jefe de proyecto:**

- Dirigir, gestionar y coordinar el contrato, los servicios/proyectos incluidos y los equipos.
- Planificar y controlar la prestación y entrega de los trabajos contratados.
- Gestionar los riesgos y los planes de mitigación o contingencia.
- Gestionar la facturación y la financiación del servicio.
- Reuniones de coordinación con los responsables del Consorci AOC servicio.
- Establecer la visión técnica global de los equipos de desarrollo.
- Definición del modelo de gobernanza y definición del catálogo de datos, en colaboración con la empresa adjudicataria del Lote 1 y el Consorci AOC
- Análisis funcional.
- Identificación de KPIs a través de las diferentes oportunidades que puedan surgir para los casos de uso.
- Procedimentado de usuarios de la plataforma, configurando diferentes roles para el módulo del gobierno del dato.
- Garantizar la fiabilidad de los datos, realizando cruces e integración de datos con las diferentes herramientas corporativas.

#### **Data Scientist / Consultor de datos:**

- Definición global de arquitecturas, procesos y modelos de apoyo a la gestión del ciclo del dato (ingesta, transformación, almacenamiento, procesamiento y explotación).
- Determinar requisitos de usabilidad y protección de datos
- Determinar requisitos normativos, técnicos y jurídicos, a aplicar en tratamiento de la información y privacidad.
- Extraer conocimiento de los datos para los diferentes casos de uso, tanto para personal técnico cualificado, como para usuarios sin conocimientos avanzados.
- Desarrollo de analítica avanzada, modelos con Machine Learning y Deep Learning.
- Crear modelos de datos, consulta de datos, gestión de datos, limpieza de datos y vinculación de datos.
- Diseño de cuadros de mando.
- Configurar módulo BI y preparar entornos para análisis y reporting.
- Utilizar las diferentes herramientas para extraer y analizar los datos generados por la actividad de los usuarios.

**Consultor junior / Programador:**

- Desarrollo de APIs y aplicaciones web para utilizar los resultados y modelos obtenidos de los análisis de datos llevados a cabo.
- Apoyo a los perfiles de Jefe de proyectos y Data Scientist.

## 8 Formas de seguimiento y control de la ejecución de las condiciones

---

El adjudicatario deberá explicar en su propuesta cuál es el modelo de relación que propone para garantizar el éxito del proyecto.

Como mínimo, sin embargo, será necesario que se establezcan los siguientes niveles de interlocución:

### **Reuniones de estrategia y dirección con las siguientes características:**

- Interlocutores: responsable del servicio por parte del adjudicatario. Jefe de Servicio del servicio y/o Jefe de Proyecto por parte del Consorci AOC y personal directivo de la AOC en su caso.
- Nocturnidad: trimestral.
- Objetivo: hacer el seguimiento del contrato, analizando diversos aspectos: productividad, control de horas, temas de facturación, seguimiento de metas (a alto nivel), etc.
- Entregables:
  - Informe resumen de las actuaciones ya resueltas, en curso y pendientes, con su grado de priorización.
  - Escandallo de horas totales realizadas en el periodo con carácter mensual.

### **Reuniones de seguimiento con las siguientes características:**

- Interlocutores: las personas asignadas por el adjudicatario para llevar a cabo el servicio. Por parte del Consorci AOC será el Jefe de Servicio y/o el Jefe de Proyecto, o alguno de los técnicos asignados al proyecto.
- Atribuciones: semanal.
- Objetivo: seguimiento del cumplimiento de la ANS, rendimiento de la plataforma, incidencias más destacables, y seguimiento de los evolutivos y correctivos a priorizar.

La comunicación, gestión y dedicación de las tareas, incidencias y propuestas de mejora se realizará mediante alguna de las siguientes herramientas: JIRA (como herramienta de negocio y seguimiento), Teams y Zendesk del Consocio AOC.

## 9 Aspectos de seguridad y ENS

Los sistemas que integren la Plataforma de datos de las Administraciones Locales de Cataluña deberán estar integrados con el landing zone de seguridad de Amazon Web Services del Consorci AOC y estar certificados con un nivel medio de acuerdo con lo establecido en el Esquema Nacional de Seguridad, regulado por el Real Decreto 311/2022, de 3 de mayo. De acuerdo con el análisis de riesgos efectuado, la categorización de los servicios objeto de esta licitación es la siguiente:

			Metodología ENS					
Activo esencial	Descripción	Subsistema	D	C	I	A	T	DP
Data Lake	Datos en bruto cargados al sistema antes de ser tratados.	AWS	Baix	Medio	Medio	Medio	Medio	NA
Datawarehouse	Datos ya procesados con estructura y formato especialmente pensado para facilitar su análisis	AWS	Baix	Medio	Medio	Medio	Medio	NA

C – Confidencialidad I – Integridad T – Trazabilidad A – Autenticidad D - Disponibilidad

DP – Protección de datos

Aunque la adquisición de la infraestructura AWS sobre la que se montará la plataforma no está contemplada en esta licitación, las empresas adjudicatarias de ambos lotes deberán garantizar que cualquier implantación o servicio que se monte cumplirá con los requisitos de seguridad definidos en el Esquema para este nivel.

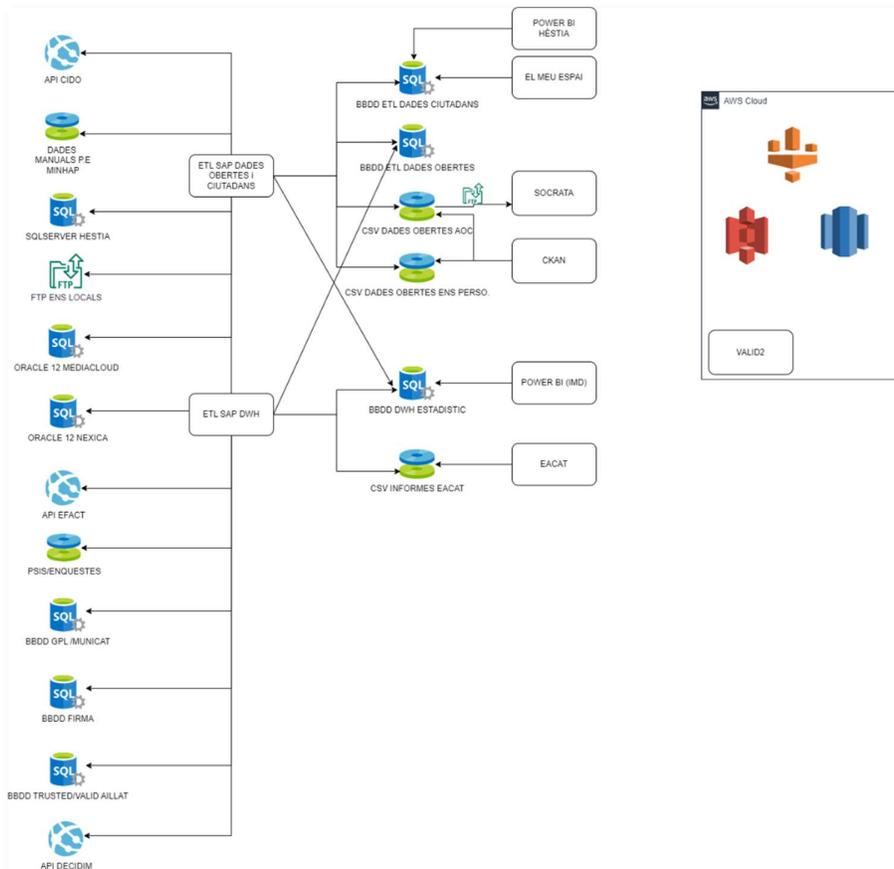
# Anexo I: Descripción Infraestructura de datos actual de la AOC

A continuación se describe la infraestructura de datos actual del Consorci AOC acotada al contexto de:

- La generación de los datos estadísticos de los servicios gestionados por el Consorci AOC, su almacenamiento en el *data warehouse* del Consorci AOC, y su publicación vía informes públicos, por ejemplo en: <https://www.aoc.cat/indicadors>
- La recopilación de múltiples datos de las administraciones catalanas para su almacenamiento y publicación en formato de datos abiertos a: <https://dadesobertes.seu-e.cat>
- La recopilación de múltiples datos de los ciudadanos para ponérselos a su disposición en "Mi Espacio" : <https://www.aoc.cat/serveis-aoc/el-meu-espai>

Se empezará describiendo las diferentes fuentes origen de datos, se continuará con el tratamiento que se realiza a estos datos mediante procesos ETL, y por último se detallarán las bases de datos de destino y las aplicaciones de negocio que acceden a las mismas.

Un esquema resumido es el siguiente, donde a la izquierda están las fuentes de datos, en el medio están los dos ETL que procesan los datos, seguidos a continuación por las bases de datos destino; y por último las diferentes aplicaciones/servicios que acceden:



## 9.1 Descripción de las fuentes de datos de origen

### 9.1.1 Bases de datos gestionadas por el Consorci AOC

A continuación se enumeran las bases de datos del Consorci AOC, éstas se encuentran divididas en cuatro centros de datos diferenciados (Mediacloud, Nexica, Azure y AWS). La descripción de cada servicio mencionado se puede encontrar en [www.aoc.cat/serveis](http://www.aoc.cat/serveis).

- Oracle 12i multiservicio en Mediacloud: esta base de datos incluye los datos de negocio de los siguientes servicios del consorcio: VÍA ABIERTA, ENOTUM, ETAULER, ETRAM2, OVER, REPRESENTA, VALISA, SIR, MUX2, PADRO, IDCAT MÓVIL, SIRI, SEDE ELECTRONICA y TRANSPARENCIA.
- SQLServer a Mediacloud: para la plataforma EACAT
- SQLServer a Mediacloud: para la plataforma ETRAM1
- Oracle 12i en Mediacloud: por el servicio de firmador centralizado (trustedX)
- Oracle 12i en Mediacloud: por el servicio VALID
- Oracle 12i en Nexica: por el servicio VALID DR
- Oracle 12i de Nexica: que da servicio a ERES e iARXIU
- Mongo DB en Mediacloud: que da servicio al VALIDADOR
- Mongo DB en Mediacloud: que da servicio al HUB de carpetas ciudadanas
- PostgreSQL a mediacloud: para el servicio de datos abiertos CKAN
- SQLServer a Azure: para el servicio HÈSTIA
- Mysql a Azure: para el servicio Fallamos
- Mysql a AWS: para el servicio VALID2
- Elasticsearch a AWS: para el servicio MUX3
- Elasticsearch a AWS: para el servicio DESAL2

### 9.1.2 Bases de datos gestionadas por otras administraciones

El consorcio en la prestación de sus servicios también accede a bases de datos de otras administraciones mediante VPN, API keys o SFTP, entre ellas:

1. Oracle GAH: en esta base de datos se encuentra información sobre la gestión y presupuestos GPL de los entes locales, los cargos electos, información institucional de los entes del mundo local, relaciones entre organismos, etc.
  - API JSON del organismo CIDO: de donde se obtiene información de diferentes AAPP como por ejemplo: ayudas becas cursos premios subvenciones, calendarios y padrones fiscales, careos, convenios colectivos municipios, estatutos organismos, ofertas públicas de convocatorias personal y sus resultados, ordenanzas fiscales, ordenanzas reguladoras y reglamentos, pliegos de cláusulas generales, presupuestos y plantillas, todo ello con el objetivo de mostrarlas en la plataforma de datos abiertos del Consorcio. AOC.
  - API JSON de la DIBA: para acceder a datos de SAIP y del registro de actividades.
  - API JSON de la plataforma de datos abiertos de la Generalitat (Socrata): se accede entre otros a los datos del registro de contratos de toda Cataluña y a los de la plataforma de

licitación PSCP con la intención de mostrarlos como datos abiertos dentro de la plataforma de datos abiertos del Consorci AOC.

- API JSON global de descarga del servicio Fallamos: se obtienen los datos de actividad de todos los Decidim de Catalunya (aunque no sean hospedados por el Consorci AOC).
- Descarga SFTP/FTP/FTPS de otras administraciones: en la prestación del servicio de datos abiertos el consorcio accede a unos 60 de organismos locales para sincronizar sus ficheros de datos abiertos con la plataforma de datos abiertos del Consorci AOC, actualmente son unos 600 ficheros diferentes.

Se puede obtener la lista completa de todos los organismos y de sus datasets en el siguiente enlace: <https://dadesobertes.seu-e.cat/organization>

- Subida SFTP otras administraciones: en el marco de colaboración con la plataforma de datos de la Generalitat el Consorci AOC también sube información al SFTP de la plataforma Socrata para su publicación en datos abiertos. Concretamente se suben diariamente unos 25 ficheros entre ellos: cargos del mundo local, estatutos, presupuestos, careos, convocatorias de personal, ordenanzas fiscales, ordenanzas reguladoras y reglamentos, convocatorias de subvenciones y ayudas, organismos dependientes, presupuestos y liquidaciones, convenios funcionariales laborales y sindicales, calendarios y padrones fiscales, liquidaciones por principales ingresos tributarios, padrón municipal de habitantes por municipio año, endeudamiento, datos de e-administración, coste efectivo de los servicios, y objetivos de estabilidad presupuestaria, etc.
- Descarga manual de otros ficheros: en los casos en que no es posible automatizar la descarga de datos de otras administraciones ésta se realiza de forma manual periódicamente, como por ejemplo con los datos del Ministerio de administración pública y Hacienda en el caso de: endeudamiento, plazo medio de pago, convocatoria y concesiones de subvenciones, etc.

### 9.1.3 Bases de datos gestionadas por proveedores externos del Consorci AOC

En algunos servicios del Consorci AOC las bases de datos se encuentran en centros de datos contratados directamente por proveedores externos, sería el caso de:

- Base de datos mysql en el CPD de Firma profesional: utilizada por el servicio de certificación digital tanto a nivel de generación de certificados para Administraciones Públicas (T-CAT) como para el certificado idCAT destinado a la ciudadanía.
- API de acceso a las facturas de las Administraciones Públicas catalanas de CERES: es un API que permite descargar las facturas y los estados de las mismas para un periodo de tiempo determinado.
- Descarga manual de datos de Google analytics: para todos los servicios del Consorci AOC, como por ejemplo las visitas a los más de 1200 portales de transparencia.
- Descarga manual de los datos de las estadísticas en survey monkey: con las encuestas de todos los servicios prestados por el Consorci AOC.

## 9.2 Instancias ETL del Consorci AOC

El Consorci AOC dispone de dos instancias ETL para la extracción, transformación y carga de los datos en las bases de datos finales. Las dos instancias están basadas en SAP data services

versión 9.2 que ya no tiene soporte del fabricante, y que por lo tanto hay que sustituir de forma incluida en esta licitación...

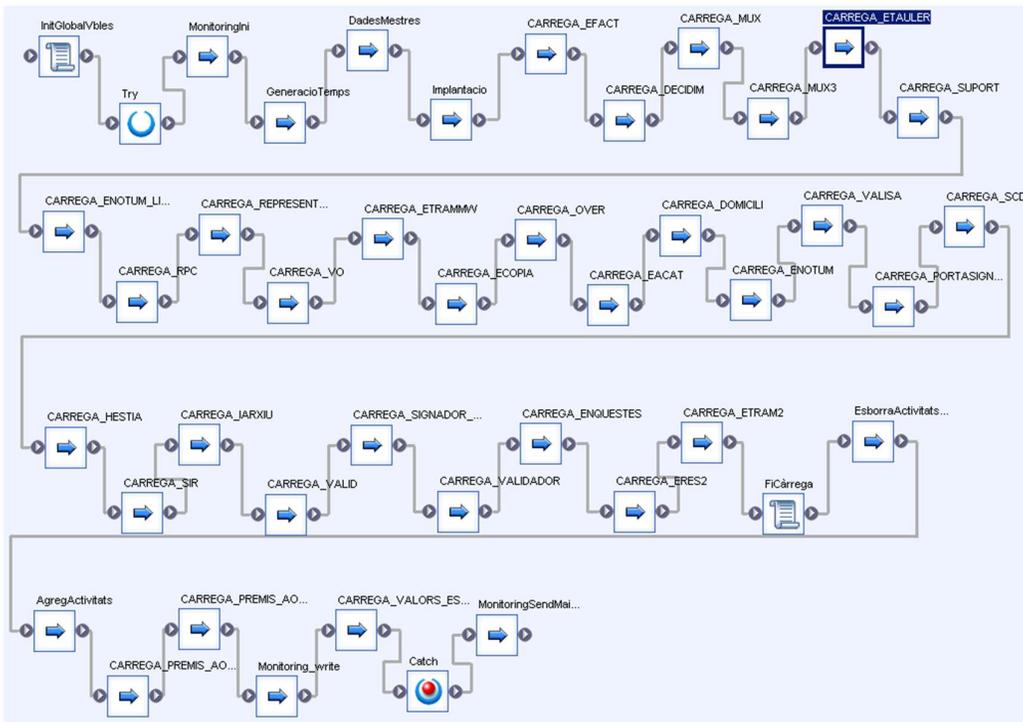
Ambos ETLs tienen capacidad de ejecutar ficheros Python para aquellos casos en que la ETL no tenga las funcionalidades requeridas para el tratamiento de datos.

Los dos ETL son los siguientes:

### 9.2.1 ETL SAP Dataservices para el DW estadístico de los servicios AOC

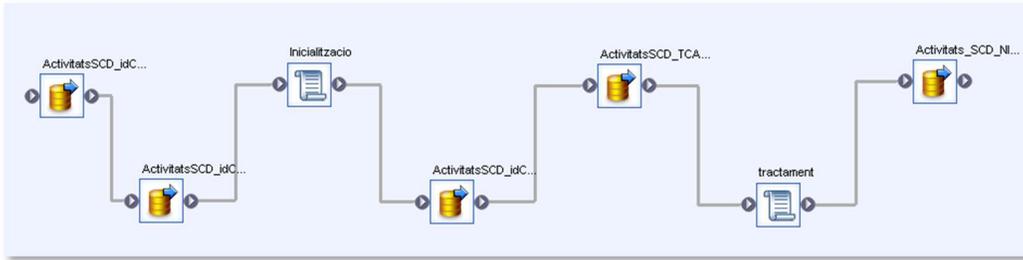
Este primer ETL obtiene datos de actividad de todos los servicios del consorcio a nivel estadístico, lo hace a partir de las fuentes de datos detalladas anteriormente, y lo desata principalmente en la base de datos estadística del data warehouse. De la misma forma este ETL también ayuda a la realización del estudio de los reconocimientos AOC, que cada año se entregan a los organismos del mundo local, para hacer esto último incluye datos adicionales más allá de los propios servicios del Consorci AOC.

- Ejemplo flujo principal del ETL con todos los subflujos:

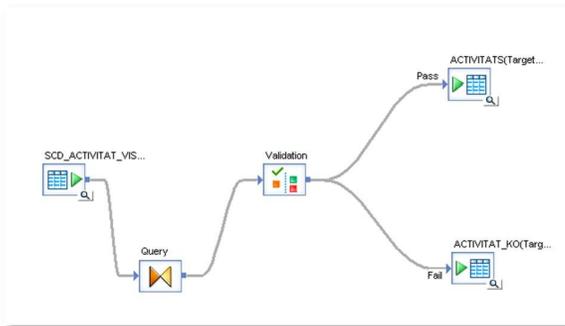


- Ejemplo subflujo de carga del servicio SCD:

Cada flujo de carga se suele estructurar en dos niveles, un primer nivel que obtiene y desata los datos en bruto y un segundo que agrega y transforma los datos hacia su destino final



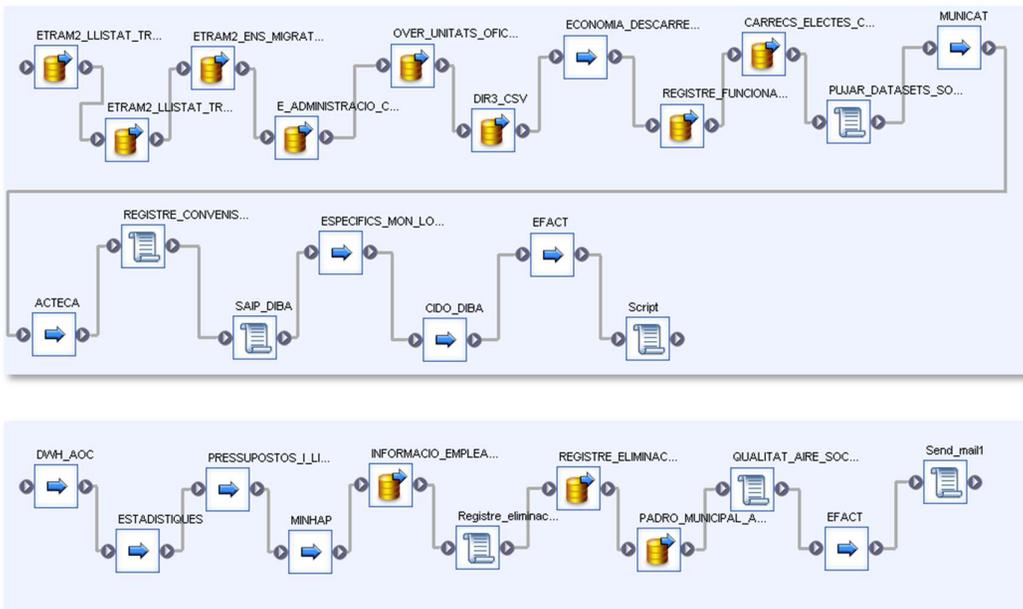
A nivel de data warehouse los datos estadísticos se acaban guardando principalmente en la tabla ACTIVIDADES, que contiene las actividades que realiza cada ente del mundo local en los servicios del Consorci AOC, junto con otras dimensiones como la temporalidad de esta actividad.



### 9.2.2 ETL SAP Dataservices para datos abiertos, y ciudadanos

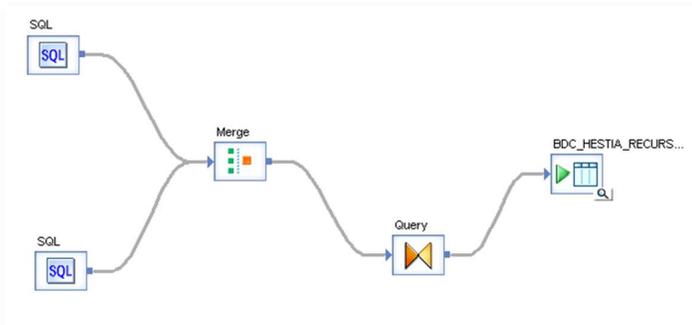
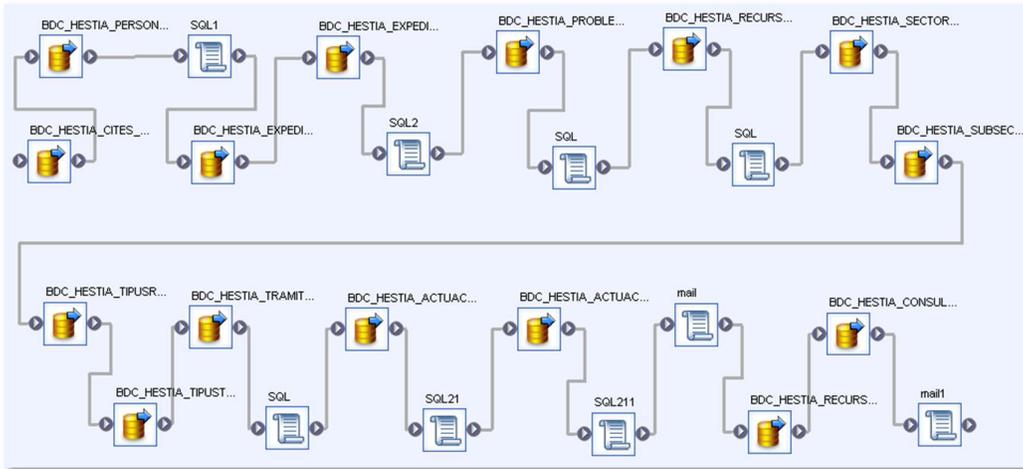
Este segundo ETL obtiene los datos abiertos de las diferentes Administraciones Públicas que serán publicados en los portales de transparencia; además tiene la responsabilidad de alimentar parte de los datos de "Mi Espacio".

- Ejemplo flujo principal de carga de datos abiertos:





- *Ejemplo de flujo de carga de HÉSTIA:*



### 9.3 Bases de datos y repositorios de destino

El destino de los datos de todos los flujos de carga ETL anteriores son los siguientes:

- Base de datos del ETL de los ciudadanos: se trata de un servidor SQL Server que da servicio específico a "Mi Espacio" a la hora de mostrar sus datos a los ciudadanos.
- Base de datos SQLServer del ETL de datos abiertos: se trata de un servidor SQL Server que da servicio a la generación de ficheros CSV de datos abiertos.

Estos ficheros CSV de datos abiertos se encuentran en un filesystem de la ETL de datos abiertos, que contiene:

- Los ficheros CSV generados por Consorci AOC en la prestación de sus servicios en las AAPP.
- Los ficheros CSV subidos directamente por los organismos del mundo local. Estos son sincronizados por la ETL de datos abiertos cada noche accediendo a los múltiples FTPs de los entes del mundo local.
- Base de datos SQLServer de la ETL del DWH del Consorci AOC: se trata de un servidor SQL Server donde se almacena:
  - un vuelco inicial de primer nivel de los datos de actividad de los diferentes servicios del Consorci AOC como se ha comentado anteriormente, así como los datos necesarios para el índice de madurez digital.
  - el tratamiento posterior de segundo nivel de los datos para que sean analizables fácilmente a nivel estadístico y se puedan mostrar.

Dada la complejidad interna de las aplicaciones, estas bases de datos deberán mantenerse, cuando menos temporalmente, en la nueva plataforma de datos del mundo local, moviendo estos SQLServers al cloud correspondiente de forma completa e incluyendo estas tareas en el marco de esta licitación. Y luego, en su caso, en una segunda fase pasando estas bases de datos al nuevo sistema de base de datos que se decida.

### 9.4 Aplicaciones que acceden a las bases de datos de destino

A continuación se detallan las aplicaciones de negocio que acceden a los repositorios destino de datos que se han comentado anteriormente.

Estas aplicaciones deberán tenerse en cuenta en los movimientos de los datos, incluidos en este pliego, hacia la nueva plataforma de datos locales del mundo local, , con el objetivo de que la migración tenga el mínimo impacto posible hacia los usuarios finales, ya sean ciudadanos o empleados públicos.

- Plataforma de datos abiertos de la AOC (CKAN): una de las principales aplicaciones que accede al repositorio de ficheros CSV del ETL de datos abiertos es el propio servicio de datos abiertos del Consorci AOC: <http://dadesobertes.seu-e.cat> con más de 600 datasets provenientes de diversas fuentes de datos (a nivel de mundo local, autonómico y estatal), y que son procesados por el ELT de datos abiertos como se ha comentado anteriormente.

The screenshot shows the 'Dades obertes' (Open Data) portal. At the top, there is a navigation bar with the AOC logo, the text 'Consorci Administració Oberta de Catalunya', and 'Dades obertes'. Below this is a search bar with the text 'Cercar conjunts de dades: Cercar conjunts de dades...'. Underneath the search bar is a section titled 'O filtra per categoria' (Filter by category) which contains a grid of 24 icons representing different data categories, each with a count of datasets. The categories and their counts are:

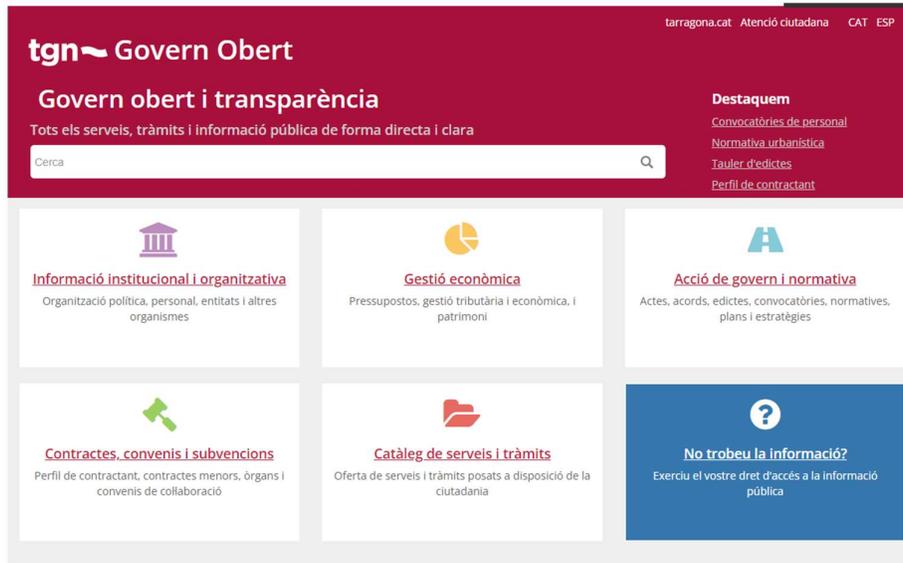
- Ciència i Tecnologia (6)
- Comerç (31)
- Cultura i oci (52)
- Demografia (47)
- Economia (60)
- Educació (10)
- Energia (8)
- Esport (7)
- Habitatge (2)
- Hisenda (5)
- Indústria (13)
- Legislació i justícia (12)
- Medi Ambient (24)
- Medi Rural i Pesca (1)
- Salut (20)
- Sector Públic (136)
- Seguretat (6)
- Societat i benestar (37)
- Transport (21)
- Treball (9)
- Turisme (9)
- Urbanisme i infraestructures (46)

Hay que comentar también que la plataforma de datos abiertos del consorcio da también servicio a la creación de portales de datos abiertos propios generados por los propios entes locales, accediendo al repositorio de la ETL de datos abiertos para mostrar los datasets correspondientes, siendo ahora mismo una veintena de organizaciones las que disponen de este tipo de portal:

The screenshot shows a grid of local government portals for open data. Each tile contains a logo, the name of the municipality, and the number of datasets available. The municipalities and their dataset counts are:

- Ajuntament de Gava: 57 Conjunts de dades
- Ajuntament de Ginestar: 0 Conjunts de dades
- Global: Organització per a conjunts de dades globals, 50 Conjunts de dades
- Ajuntament de Granollers: 18 Conjunts de dades
- Ajuntament de la Nou de Gaià: 1 Conjunt de dades
- Ajuntament de les Preses: 10 Conjunts de dades
- Ajuntament de Llagostera
- Ajuntament de Lleida

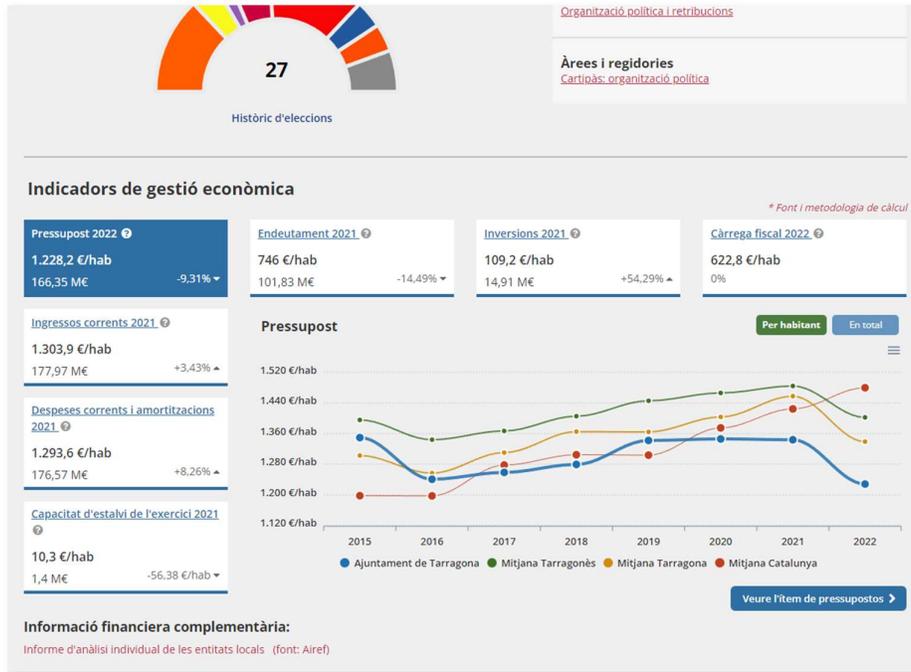
A su vez diferentes servicios acceden a la plataforma de datos abiertos del Consorci AOC para mostrar su información, como por ejemplo los más de 1200 portales de sede electrónica y/o transparencia del mundo local provistos también por el Consorci AOC (por ejemplo [www.seu-e.cat/web/tarragona](http://www.seu-e.cat/web/tarragona)), destacando por este motivo la importancia de tenerlo en cuenta en la migración hacia la nueva plataforma de datos del mundo local.



En estos portales de transparencia también se muestran incrustados los datos abiertos de la personalizados para cada ente:



También se muestran diferentes gráficos alimentados por estos datos abiertos de la plataforma CKAN, como por ejemplo los indicadores de gestión económica, la estructuración por programas del presupuesto y la liquidación del mismo, etc.



- Mi Espacio: permite a la ciudadanía acceder a sus datos en su relación con las administraciones públicas catalanas y de una forma interadministrativa. Se puede consultar la ficha de "mi Espacio" en <https://www.aoc.cat/serveis-aoc/el-meu-espai/>, y se puede acceder al entorno productivo aquí: <https://seu-e.cat/mygov/9821920002>

La teva activitat els darrers 6 mesos amb les administracions:

**15**

**Tràmits**

Les meves sol·licituds, gestions i factures.

**0**

**Notificacions**

Les comunicacions adreçades a mi.

**5**

**Consultes a les meves dades**

Qui, quan i per quin motiu.

**66**

**Inicis de sessió**

Quan m'he connectat, amb qui i com.

**0**

**Expedients**

L'estat dels meus expedients.

[Veure tota la meva activitat](#)

[Administracions que mostren informació a El meu espai](#)

Cercador d'Informació i Documentació Oficials (CIDO)

**Selecció de personal i ofertes de treball (en proves)**

Cercador de processos selectius, borses de treball i altres convocatòries

#### **Recomanador social (en proves)**

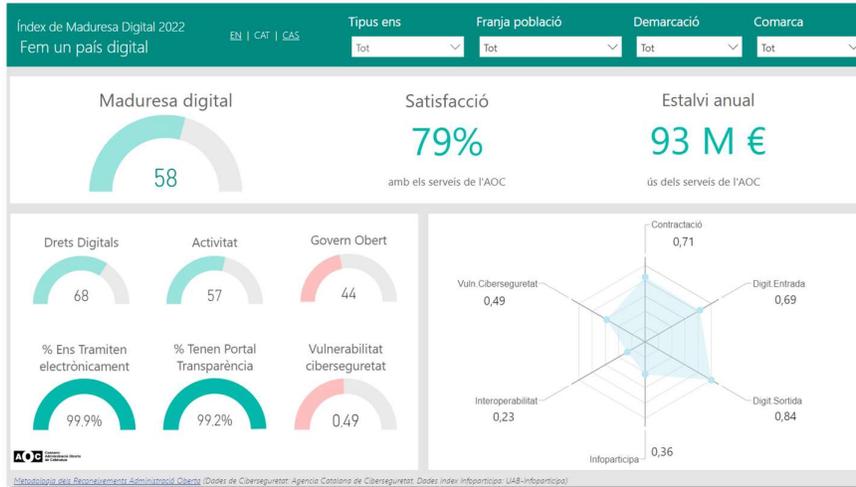
Obté informació sobre els ajuts socials que t'ofereixen les diferents administracions públiques en funció de la teva situació personal i familiar.

[En la versió beta del servei només es contemplen els ajuts promoguts pel Govern d'Espanya i la Generalitat de Catalunya. Pròximament s'inclouran els ajuts municipals i de la resta d'entitats del sector públic]

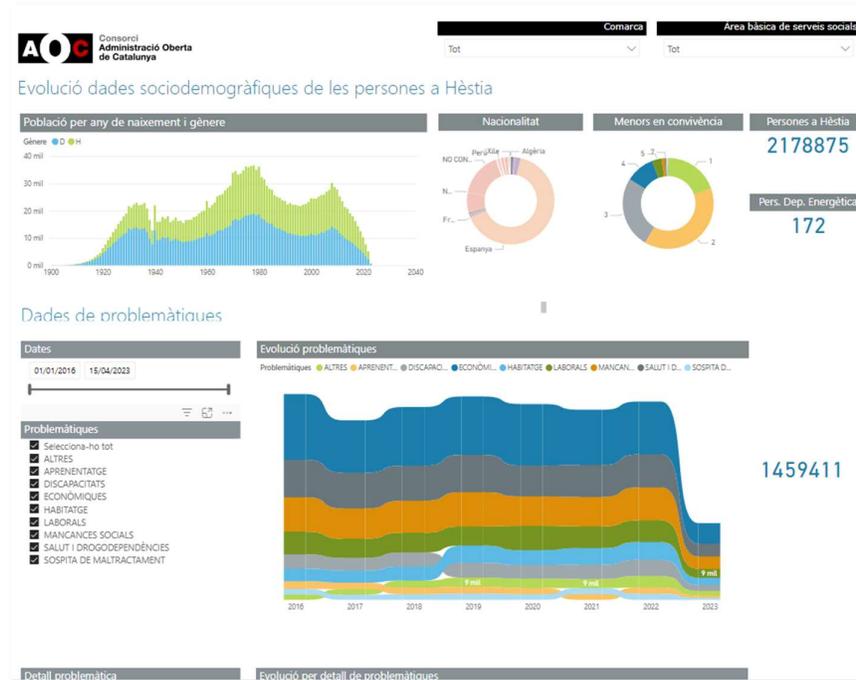
[ACCEDIR AL RECOMANADOR](#)

Ens ajudes a millorar?





- **Power bi de cuadros de mando específicos:** que incluyen ratios del servicio concreto:



- **Plataforma de datos abiertos de la Generalitat (Socrata):** es otra aplicación que accede a los datos abiertos generados por el Consorci AOC es la plataforma Socrata de la Generalitat <https://analisi.transparenciacatalunya.cat/ca/> . En este caso concreto desde el repositorio de la ETL de datos abiertos se suben a Socrata vía SFTP unos 20 datasets con datos del mundo local de forma diaria.

**dades obertes**  
catalunya

Cercar

1151 Resultats Filtre >

---

**#català947. La llengua catalana als municipis de Catalunya** + Dossier o document

Societat i benestar

Conjunt de dades demogràfiques (origens de la població), sociolingüístiques (coneixements i usos lingüístics) i d'aprenentatge de català per adults i d'oferta de català als comerços dels municipis de Més

Actualitzat: 28 de febrer de 2023  
Vistes: 479.295

Temes: municipis, situació català, situació sociolingüística, coneixements i usos lingüístics, territori, i 1 més

---

**Registre de casos de COVID-19 a Catalunya per municipi i sexe** Conjunt de dades

Salut

El Departament de Salut ha deixat d'actualitzar aquest conjunt de dades en data 26-7-22. Podeu trobar tota la informació relacionada amb la COVID-19 i la resta d'infeccions respiratòries en els Més

Actualitzat: 17 de març de 2023  
Vistes: 255.301

Temes: covid-19, coronavirus, sars-cov-2, covid19 Documents de l'API

- La plataforma EACAT: es la extranet de las administraciones públicas catalanas (<https://www.aoc.cat/serveis-aoc/eacat-tramits/>), y también accede al repositorio del ETL del Data warehouse para la generación de informes CSV de actividad de los servicios Via Oberta y Padró, para los diferentes entes que soliciten su descarga.

## Anexo II: Datos disponibles del mundo local actualmente

---

Las principales fuentes de datos del mundo local son:

- *Data warehouse* de l'AOC
  - o La información de actividad de los servicios del Consorci AOC se puede consultar, de forma agregada, en el siguiente informe: [Indicadores – Consorci AOC](#)
- Plataforma compartida de datos abiertos de la AOC, con conjuntos de datos globales (publicados y gestionados por el Consorci AOC) y específicos de entes locales:
  - o [Datos Abiertos – Consorci AOC](#)
- Plataforma de datos abiertos de la Generalitat de Catalunya:
  - o [Datos Abiertos de Cataluña](#)
- Plataforma de datos abiertos de la Administración General del Estado
  - o [Datos Abiertos del Gobierno de España](#)
- Plataformas de datos abiertas de otras administraciones locales con plataforma propia (Ayuntamiento de Barcelona, Diputación de Barcelona, etc).
  - o [Open Data BCN](#)
  - o [Portal de Datos Abiertos de la Diputación de Barcelona](#)