



Administració Oberta
de Catalunya

PLEC DE PRESCRIPCIONS TÈCNIQUES DEL CONTRACTE DELS SERVEIS DE LA PLATAFORMA DE DADES DE LES ADMINISTRACIONS LOCALS DE CATALUNYA

Realitzat per: Consorci AOC

Versió: 2.0

Data: 3/6/2024

Arxiu: 2023 Contracte Plataforma Dades Locals NGEU - Plec Tècnic.docx

Número d'expedient: AOC 2024 47

Índex

1	Introducció.....	4
2	Context.....	5
2.1	Situació actual	5
2.2	Marc Normatiu	5
2.3	Infraestructura de dades del Consorci AOC	7
3	Objectius, fites i calendari	8
4	Objecte del contracte	9
4.1	LOT 1: Serveis de la plataforma de dades.....	10
4.2	LOT2: Consultoria i explotació de negoci	11
5	Activitats i funcions de l'empresa contractista	13
5.1	Activitats i funcions lot 1	13
5.2	Activitats i funcions lot 2	13
6	Finalitats i objectius que s'han d'assolir.....	15
6.1	Finalitats i objectius del lot 1	15
6.2	Finalitats i objectius del lot 2	16
7	Requisits generals obligatoris	18
7.1	Requisits generals obligatoris per a tots els lots	18
7.1.1	Normativa vigent de compliment obligat.....	18
7.1.2	Requisits referents a protecció de dades personals	18
7.2	Requisits obligatoris LOT1	18
7.2.1	Generals i d'arquitectura	19
7.2.2	Ingesta	20
7.2.3	Emmagatzematge.....	21
7.2.4	Virtualització de la dada	22
7.2.5	Intel·ligència de negoci.....	22
7.2.6	Analítica avançada	23
7.2.7	Monitorització i administració	23
7.2.8	Publicació i interoperabilitat.....	24
7.2.9	Data warehouse del Consorci AOC.....	24
7.2.10	Govern de la dada	25
7.2.11	Descripció de mitjans personals.....	27
7.3	Requisits obligatoris LOT2	27
7.3.1	Govern de la dada	27
7.3.2	Marc de treball.....	28
7.3.3	Fases	28
7.3.4	Detecció i estudi de casos d'ús	28
7.3.5	Serveis de desenvolupament i implementació de casos d'ús	29

7.3.6 Servei de Capacitació.....	33
7.3.7 Seguiment.....	33
7.3.8 Casos d'ús: prioritats, abast i àmbits.....	33
7.3.8.1 Priorització dels casos d'ús	33
7.3.8.2 Abast dels casos d'ús.....	33
7.3.8.3 Àmbits dels casos d'ús.....	34
7.3.8.4 Exemple de cas d'ús: Índex de Maduresa Digital	35
7.3.9 Descripció de mitjans materials.....	35
7.3.10 Descripció de mitjans personals.....	35
8 Formes de seguiment i control de l'execució de les condicions.....	37
9 Aspectes de seguretat i ENS.....	38
Annex I: Descripció Infraestructura de dades actual de l'AOC.....	39
9.1 Descripció de les fonts de dades d'origen	40
9.1.1 Bases de dades gestionades pel consorci AOC	40
9.1.2 Bases de dades gestionades per altres administracions	40
9.1.3 Bases de dades gestionades per proveïdors externs del Consorci AOC	41
9.2 Instàncies ETL del Consorci AOC.....	41
9.2.1 ETL SAP Dataservices per al DW estadístic dels serveis AOC.....	42
9.2.2 ETL SAP Dataservices per a dades obertes, i ciutadans.....	43
9.3 Bases de dades i repositoris de destí	46
9.4 Aplicacions que accedeixen a les bases de dades de destí.....	46
Annex II: Dades disponibles del món local actualment.....	54

1 Introducció

El contingut d'aquestes prescripcions tècniques deriva del projecte "Plataforma comuna de gestió i analítica de dades per les administracions catalanes" amb codi C11.I03.P04.S14, aprovat en el marc del Mecanisme de Recuperació i Resiliència (MRR), regulat pel reglament 2021/241 del Parlament Europeu i del Consell de 12 de febrer de 2021.

Amb la mera presentació de la seva oferta, l'empresa licitadora accepta les prescripcions tècniques establertes en aquest plec.

Qualsevol proposta que no s'ajusti als requeriments mínims establerts en aquest plec quedarà automàticament exclosa de la licitació.

2 Context

2.1 Situació actual

Les administracions públiques encara no tenen una cultura i una estratègia de govern de la dada. A més, a Catalunya hi ha 2.250 organismes públics que generen dades i que tenen l'obligació de complir la Directiva (UE) 2019/1024, relativa a les dades obertes i la reutilització de la informació del sector públic en format reutilitzable. De tots aquests organismes, 200 són del sector públic de la Generalitat de Catalunya, 1800 són entitats locals i la resta del sector universitari i d'altres.

2.2 Marc Normatiu

La Intel·ligència Artificial (IA) s'ha convertit en una àrea d'importància estratègica amb potencial per ser un motor clau del desenvolupament econòmic, i té un ampli ventall de possibles implicacions socials. Com a part de la seva estratègia de mercat únic digital, la Comissió Europea va presentar l'abril de 2018 una estratègia europea sobre IA a la seva Comunicació "Intel·ligència Artificial per a Europa" COM(2018)237. Els objectius de l'estratègia europea anunciats al comunicació són:

- Potenciar la capacitat tecnològica i industrial de la UE i l'adopció de la IA a tota l'economia, tant mitjançant els sectors públic i privat
- Preparar-se per als canvis socioeconòmics provocats per la IA
- Garantir un marc ètic i legal adequat.

Posteriorment, el desembre de 2018, la Comissió Europea i els Estats membres van publicar un "Pla Coordinat sobre Intel·ligència Artificial", COM(2018)795, sobre el desenvolupament de la IA a la UE.

L'any 2020, el Joint Research Centre (JRC) de la Comissió Europea, en col·laboració amb la Direcció General per les Xarxes de Comunicació, Continguts i Tecnologia (DG CONNECT) va publicar un informe que proposa una definició operativa d'intel·ligència artificial, amb l'objectiu de supervisar el desenvolupament, la seva absorció i impacte a Europa. L'estudi defineix "els sistemes intel·ligència artificial (IA) com els sistemes de programari (i possiblement també de maquinari) dissenyats per humans que, donat un objectiu complex, actuen en la dimensió física o digital percebent el seu entorn a través de l'adquisició de dades, interpretant les dades estructurades o no estructurades recollides, raonant sobre el coneixement obtingut o processar la informació derivada d'aquestes dades i decidir les millors accions a dur a terme per assolir el donat objectiu. Els sistemes d'IA poden utilitzar regles simbòliques o aprendre un model numèric, i també poden adaptar-los comportament mitjançant l'anàlisi de com l'entorn es veu afectat per les seves accions anteriors".

Tenint en compte les implicacions que suposa l'ús de les tecnologies que es poden encabir en aquesta definició, la Comissió Europea està treballant en dues propostes de regulació per definir les regles del seu ús en l'àmbit de la Unió. Són:

- Proposta (COM/2021/206 final) de Reglament del Parlament Europeu i del Consell pel qual s'estableixen normes harmonitzades en matèria d'Intel·ligència Artificial (Llei d'Intel·ligència Artificial) i es modifiquen determinats actes legislatius de la Unió. L'objectiu de la proposta és millorar el funcionament del mercat intern mitjançant l'establiment d'un marc jurídic uniforme, en particular pel que fa al desenvolupament, comercialització i utilització de la intel·ligència artificial de conformitat amb els valors de la Unió. Persegueix diverses finalitats imperioses d'interès general, com ara assegurar un nivell elevat de protecció de la salut, la seguretat i els drets humans, i garanteix la lliure circulació transfronterera de béns i serveis basats en la IA, de manera que impedeix que els Estats membres imposin restriccions al desenvolupament, la comercialització i la utilització de sistemes d'IA, llevat que aquest Reglament ho autoritzi expressament. La

proposta de reglament estableix normes per regular la introducció al mercat i la posada en servei de determinats sistemes d'IA, cosa que garantirà el bon funcionament del mercat interior i permetrà que aquests sistemes es beneficiïn del principi de la lliure circulació de béns i serveis.

- Reglament (UE) 2023/2854 del Parlament Europeu i del Consell, de 13 de desembre, sobre normes harmonitzades per a un accés just a les dades i la seva utilització (Llei de dades). Els seus objectius són:
 - Facilitar l'accés a les dades i la utilització per part dels consumidors i les empreses, preservant alhora els incentius per invertir en formes de generar valor a través de les dades.
 - En aquelles situacions en què hi hagi una necessitat excepcional d'obtenir dades, preveure l'ús de les dades que estiguin en poder de les empreses per part dels organismes del sector públic i de les institucions de la Unió.
 - Facilitar el canvi entre serveis al núvol. L'accés a serveis de tractament de dades competitives i interoperables és una condició prèvia per a una economia de dades florent, en què les dades es puguin compartir fàcilment dins dels ecosistemes sectorials i entre elles. El nivell de confiança en els serveis de tractament de dades determina l'acceptació dels serveis esmentats per part dels usuaris en tots els sectors de l'economia.
 - Establir salvaguardes contra la transferència il·legal de dades sense notificació per part dels proveïdors de serveis al núvol.
 - Preveure el desenvolupament de normes d'interoperabilitat per a la reutilització de les dades entre sectors, a fi d'eliminar els obstacles a l'intercanvi de dades entre espais comuns europeus de dades, d'acord amb els requisits d'interoperabilitat sectorials, i entre d'altres que no entrin a l'àmbit d'aplicació d'un espai comú europeu de dades específic.

A l'estat espanyol, i en el marc de l'Estratègia Nacional d'Intel·ligència Artificial, la Llei 15/2022, de 12 de juliol, integral per a la igualtat de tracte i la no-discriminació, al seu article 23 anima a les administracions públiques a emprar aquesta tecnologia. En concret:

1. Les administracions públiques afavoriran la posada en marxa de mecanismes perquè els algorismes involucrats en la presa de decisions que s'utilitzin a les administracions públiques tinguin en compte criteris de minimització de biaixos, transparència i rendició de comptes, sempre que sigui factible tècnicament. En aquests mecanismes s'inclouran el disseny i les dades d'entrenament i abordaran el potencial impacte discriminatori. Per assolir aquest fi, es promourà la realització d'avaluacions d'impacte que determinin el possible biaix discriminatori.
2. Les administracions públiques, en el marc de les seves competències en l'àmbit dels algorismes involucrats en processos de presa de decisions, prioritzaran la transparència en el disseny i la implementació i la capacitat d'interpretació de les decisions adoptades pels mateixos.
3. Les administracions públiques i les empreses han de promoure l'ús d'una intel·ligència artificial ètica, fiable i respectuosa amb els drets fonamentals, seguint especialment les recomanacions de la Unió Europea en aquest sentit.
4. Es promourà un segell de qualitat dels algorismes.

Finalment cal fer esment de l'Estratègia d'Intel·ligència Artificial de Catalunya, aprovada per Acord de Govern de 18 de febrer de 2020, i que manifesta entre els seus objectius:

- Impulsar un model de governança transversal que doni suport al desenvolupament d'un ecosistema de intel·ligència artificial coordinat i connectat amb el món.
- Disposar de les infraestructures necessàries per al desenvolupament de la intel·ligència artificial i facilitar l'accés segur a les dades públiques i privades.

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

2.3 Infraestructura de dades 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 relacionades amb els serveis que ofereix al conjunt d'Administracions Públiques Catalanes. La solució actual permet recollir les dades provinents dels diferents serveis i els desa en un magatzem comú des d'on es podran publicar com a dades obertes i es podran emprar construir quadres de comandament. L'actual sistema 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 l'actual plataforma de dades del Consorci AOC es troba a l'Annex d'aquest plec.

3 Objectius, fites i calendari

L'objectiu és crear una plataforma de dades per a les administracions públiques locals de Catalunya per posar en valor les dades obertes a la societat i impulsar iniciatives basades en el govern de les dades per desenvolupar solucions que ajudin a millorar la planificació, la predicció i la presa de decisions. Aquest espai ha de permetre desenvolupar una nova generació de serveis innovadors que facilitin la personalització i la proactivitat, amb una visió transversal i integrada de l'administració pública.

Concretament, caldrà assolir les següents fites i calendari relacionades amb els serveis de la plataforma:

1. Serveis de la plataforma de dades obertes i del *data lake* del Consorci AOC i de les administracions locals a les que presta servei, federada amb la de la Generalitat de Catalunya.
 - a. Inici dels serveis de la plataforma de dades obertes
 - i. Calendari: 4 mesos des de la formalització del contracte
 - b. Inici dels serveis del *data lake* del Consorci AOC
 - i. Calendari: 6 mesos des de la formalització del contracte
 - c. Ingesta de nous conjunts de dades al *data lake* del Consorci AOC
 - i. ingesta de nous conjunts de dades amb una metodologia iterativa i amb una freqüència de 3 mesos, condicionat a la seva disponibilitat.
2. Inici dels serveis de *data warehouse* per al Consorci AOC.
 - a. Calendari: 6 mesos des de la formalització del contracte
3. Inici dels serveis de governança de la plataforma de dades del món local.
 - a. Calendari: 6 mesos des de la formalització del contracte
4. Serveis d'anàlisi avançada de dades pel món local.
 - a. Posada en marxa de les solucions d'anàlisi avançada: 6 mesos des de la formalització del contracte
 - b. Implantació dels casos d'ús d'anàlisi avançada: lliurament de nous casos d'ús amb una metodologia iterativa i amb una freqüència màxima de 3 mesos.

Aquestes fites es detallen i desenvolupen en detall a l'apartat 6 d'aquests plecs. El licitador de cada lot ha de fer una proposta de terminis per assolir aquestes fites i objectius.

4 Objecte del contracte

Aquest plec té per objecte establir les prescripcions tècniques particulars que regiran la els serveis de la Plataforma de Dades de les Administracions Locals de Catalunya, federada amb la Plataforma Transversal de Dades de la Generalitat de Catalunya, per crear una infraestructura i serveis que permeti a totes les administracions catalanes millorar la seva gestió pública en base a les dades, en els àmbits de la planificació, la presa de decisions i l'automatització de tasques.

Amb la realització de l'objecte contractual esmentat, l'òrgan de contractació pretén cobrir les necessitats i/o funcionalitats descrites a continuació.

Actualment, el Consorci AOC es troba en un procés de transformació estratègic 2023-2028, basat en un nou model de gestió global de les plataformes i tecnologies TIC, seguint una estratègia manifesta d'evolució i migració dels serveis cap al *cloud* públic. Aquest procés implica un clar full de ruta per millorar l'eficiència i l'eficàcia dels seus serveis gràcies, entre d'altres, a la simplificació operacional, a l'aplicació d'economies d'escala, a l'obtenció de sinèrgies entre els diferents serveis i a l'optimització dels recursos. L'objectiu principal d'aquest procés de transformació és alinear el nou model TIC amb les necessitats de l'organització i aprofitar les millores tecnològiques presents i futures que ofereix el *cloud* públic per aconseguir solucions més eficients, flexibles, àgils, robustes i sostenibles. En aquest sentit, el Consorci AOC ha apostat pels serveis *cloud* d'Amazon Web Services (AWS).

El present projecte planteja el desenvolupament, manteniment i gestió d'una plataforma de dades sobre una infraestructura de *cloud* públic a AWS, i que implementi un llac de dades (*data lake*) per a les administracions locals, permeti l'ús d'eines d'anàlítica avançada i possibiliti processos d'aprenentatge emprant grans volums de dades (*Big Data*).

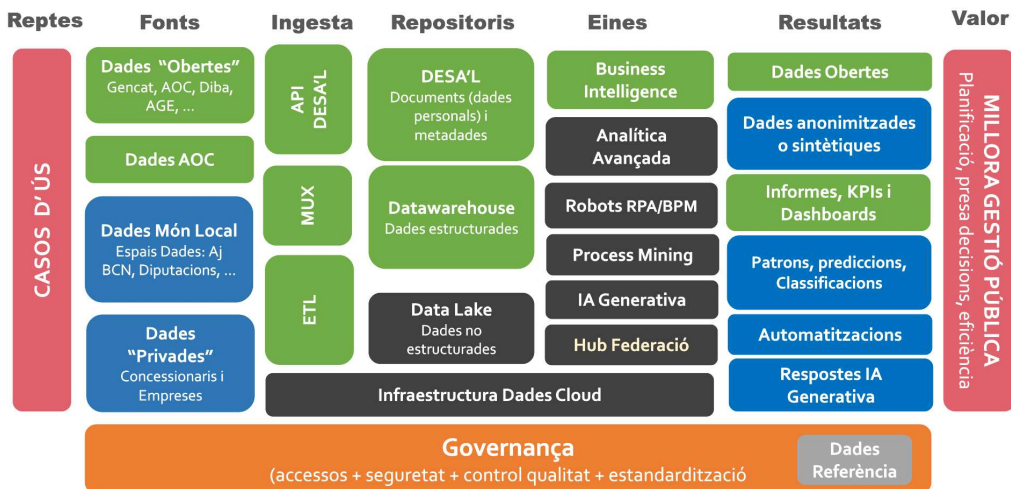


Figura 1. Plataforma de dades del Món Local

Nota: en color verd s'indiquen els mòduls on el Consorci AOC té solucions ja desenvolupades parcialment o totalment

La plataforma de dades haurà d'incloure:

- Mecanismes d'ingesta que permetin la càrrega automatitzada de dades provinents de fonts tant internes com externes
- Un llac de dades com a repositori de grans volums de dades en brut, tant estructurades com no estructurades
- Entorn de *data warehouse* equivalent al descrit a l'annex I d'aquests plecs

- Eines de transformació automàtica de les dades per la seva publicació i anàlisi
- Eines d'indexació, catalogació i cerca de les dades
- Mecanismes de control d'accés a les dades, entorns d'anàlisi i quadres de comandament
- Entorn de disseny i publicació de quadres de comandament
- Entorn de publicació de dades obertes
- Entorn d'anàlisi avançada de dades amb capacitat de tractar grans volums de dades (Big Data)

La plataforma de dades del món local de Catalunya haurà d'estar federada amb la Plataforma Transversal de Dades de la Generalitat de Catalunya i la capacitat d'estar-ho amb d'altres Espais que puguin sorgir o estar disponibles.

— Model Federat de Dades

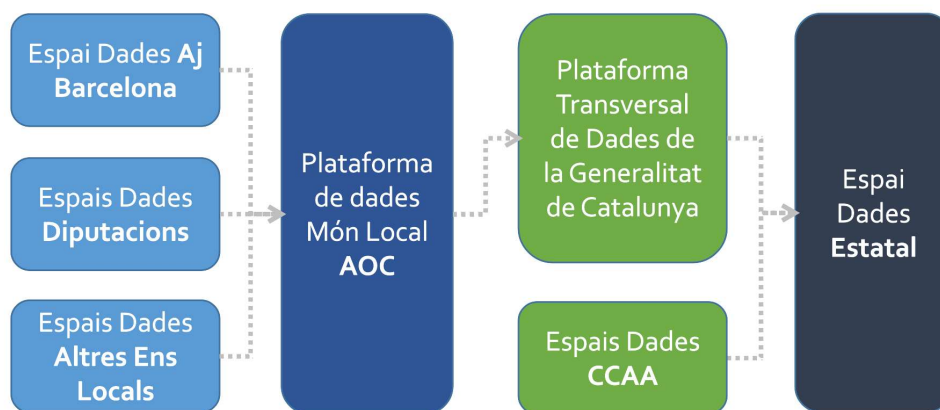


Figura 2. Model federat de dades

Tenint en compte les característiques del projecte, es proposa la contractació de dos lots, que contemplin per una banda els serveis de la plataforma de dades i per un altra la governança, consultoria i explotació de negoci.

4.1 LOT 1: Serveis de la plataforma de dades

Els serveis d'implantació, manteniment i gestió de la plataforma de dades per les administracions locals de Catalunya és un projecte complex que requereix d'una gran quantitat de recursos i coneixements especialitzats. És necessària la contractació d'un servei especialitzat en la matèria per dissenyar, desenvolupar, mantenir i gestionar la plataforma de forma eficaç i eficient, complint les seves necessitats específiques.

En base a la plataforma tecnològica proposada, caldria dissenyar i implantar la plataforma de dades, configurar el processos de recollida i tractament que el Consorci AOC actualment ja té implantats i portar a terme un procés continu de càrrega de les dades disponibles al *data warehouse*. En concret, els serveis de la plataforma de dades hauran d'incloure:

- Plataforma de dades en base al servei de *cloud* públic Amazon Web Services (AWS).
- Configuració de la ingesta de dades de, com a mínim, les fonts actualment configurades al *data warehouse* del Consorci AOC.
- Migració de dades dels casos d'ús que es vagin identificant i que aporten valor

- Establiment dels mecanismes de federació amb la resta de plataformes de dades de l'entorn.
- Adaptació dels quadres de comandament actualment existents per integrar-los amb el nou repositori.
- Suport a metadades: capacitats que admeten l'ús extensiu de metadades (p. ex., metadades d'ús, registres de transaccions, càrregues de treball del sistema) per automatitzar/millorar les tasques d'integració i explotació de dades.
- Suport de govern de dades: capacitats que ajuden als mandats de govern de dades (p. ex., qualitat de les dades, llinatge de dades) mentre es gestionen les dades per complir la integració de dades específiques casos d'ús (p. ex., MDM, compartició de dades)

4.2 LOT2: Consultoria i explotació de negoci

La Plataforma de dades del món local requereix de la definició d'un model de governança que defineixi les polítiques i protocols necessaris per l'adquisició, gestió i tractament de les dades per garantir la seva qualitat i seguretat. Per la seva explotació, a banda de la capacitat de disposar de quadres de comandament que permetin el seu anàlisi descriptiu, també requereix poder aplicar tècniques més avançades de Machine Learning i Deep Learning que permetin inferir coneixement aplicable en la millora de l'eficiència de les administracions.

Un servei de consultoria pot ajudar a definir i implementar els processos necessaris per a l'emmagatzematge, la gestió i l'anàlisi de les dades, així com la seva explotació. La identificació les millors eines i tecnologies per a la gestió i tractament de les dades, així com a dissenyar i implementar els fluxos de treball adequats per a garantir l'eficàcia de la seva gestió és la base per poder disposar de dades de qualitat que permetin a les administracions una millora continua del seu funcionament i l'automatització dels seus processos.

Per altra banda, es requereix d'un coneixement profund del funcionament de les administracions públiques i de gestió i analítica avançada de dades per poder definir casos d'ús, moltes vegades comuns al conjunt de les administracions locals, que permetin el seu anàlisi i posterior explotació. La publicació d'informes i quadres de comandament amb els resultats d'aquests anàlisis, de petites APIs que permetin l'ús de funcionalitats adquirides durant els estudis, com pot ser l'ús de xarxes neuronals entrenades amb les dades de la Plataforma, i els projectes d'automatització de processos ajudaran al conjunt de les administracions a transformar el seu funcionament.

En definitiva, aquest plec planteja la contractació d'un servei de consultoria amb els següents objectius:

- Definició de la governança del Servei, ajudant a definir les polítiques i els protocols necessaris per a garantir la qualitat i la seguretat de les dades. En concret, la consultoria ajudarà a definir les polítiques de gestió de dades, les polítiques de seguretat, les polítiques de privacitat i les polítiques d'accés a les dades. D'aquesta manera es pretén evitar problemes com la pèrdua de dades, les vulnerabilitats de seguretat i les infraccions de privacitat.
- Establiment de procediments per la seva explotació
- Detecció i tractament de problemes comuns a totes les Administracions Locals, Ajudar en la definició i implementació de casos d'ús d'analítica avançada de dades i establint mecanismes per la seva resolució.

En resum, contractar un servei de consultoria per la governança i explotació de la plataforma de dades del món local ha de servir per ajudar a detectar, estudiar i resoldre problemes comuns que pugui tenir el conjunt de les administracions públiques catalanes tant individualment com en conjunt. L'estudi de casos d'ús mitjançant tècniques d'analítica avançada de dades de Machine Learning i Deep Learning és la via per extraure coneixement de les dades que el Consorci AOC i les administracions locals generen i disposen. La seva explotació mitjançant la publicació de quadres de comandament, informes, petites aplicacions web o APIs contribuirà a que el seu funcionament sigui més eficient. Mitjançant l'automatització de processos robòtics, per exemple,

les organitzacions poden capitalitzar els coneixements d'intel·ligència artificial (IA) per agilitzar moltes tasques que tradicionalment es porten a terme de forma manual. En aquest sentit, i per facilitar l'explotació del coneixement adquirit, estarà inclòs en l'abast d'aquest lot tant el disseny, desenvolupament i integració de la solució d'automatització de tasques basada en tecnologies RPA amb les funcionalitats que oferirà la Plataforma de Dades.

5 Activitats i funcions de l'empresa contractista

Les funcions que ha d'assumir l'empresa contractista són les següents:

5.1 Activitats i funcions lot 1

- **Disseny de la infraestructura tecnològica:** l'empresa contractista ha de dissenyar l'arquitectura i la plataforma de dades del món local utilitzant la plataforma de *cloud* pública d'Amazon Web Services (AWS) del Consorci AOC. Això implica determinar els requisits tècnics i funcionals, configurar l'entorn de *cloud*, definir l'estructura de les bases de dades i establir els mecanismes de seguretat i protecció de les dades.
- **Desplegament de la infraestructura:** l'empresa contractista ha de realitzar tots els treballs necessaris per al desplegament continu de la plataforma de dades, incloent la instal·lació i configuració del programari necessari, la realització de proves de funcionalitat i rendiment, la implementació de mesures de seguretat, així com la formació del personal dels ens locals sobre l'ús adequat de la plataforma.
- **Serveis de *data warehouse*:** l'empresa contractista ha de garantir que la plataforma de dades i, concretament, el *data warehouse*, contingui, com a mínim, funcionalitats equivalents a les descrites a l'annex I d'aquests plecs. La plataforma ha de disposar també de les mateixes dades disponibles al *data warehouse* del Consorci AOC, incloent les dades històriques. Les dades s'han de conservar de manera precisa i segura durant tot el procés.
- **Gestió de la infraestructura i explotació tecnològica:** l'empresa contractista és la responsable de proporcionar i gestionar tota la infraestructura tecnològica necessària per a la creació i el funcionament de la plataforma de dades per al món local. Això inclou tota la infraestructura al núvol rellevant per al funcionament eficient de la plataforma.
- **Col·laboració amb la definició de governança de les dades:** l'empresa contractista ha de col·laborar amb l'empresa adjudicatària del lot 2 d'aquests plecs en la definició de les polítiques de governança de les dades. Això inclou establir les bones pràctiques per a l'adquisició, l'emmagatzematge, la seguretat, l'accés i l'ús de les dades en la plataforma. L'empresa contractista pot oferir orientació sobre els marcs de governança, les polítiques de privacitat, les regulacions de protecció de dades i altres aspectes relacionats amb la gestió adequada de les dades.

5.2 Activitats i funcions lot 2

- **Definició de la Governança de la Plataforma de dades del món local,** establint un catàleg de dades i fonts de dades actualitzable amb el temps, així com les bones pràctiques per la seva adquisició, emmagatzematge, accés i ús, en col·laboració amb l'adjudicatari del lot 1.
- **Definició i actualització d'un Marc de Treball** per l'estudi, aprovació, execució i explotació de casos d'ús d'analítica avançada, emprant les funcionalitats i dades de la Plataforma.
- L'adjudicatari del contracte podrà **proposar l'ús de fonts de dades no integrades a la Plataforma** sempre el seu anàlisi serveixi per millorar el funcionament de les administracions locals, o del propi Consorci AOC, en el marc d'un cas d'ús objecte d'estudi.
- **Portar a terme els casos d'ús d'analítica avançada de dades,** ja sigui aplicant tècniques de Machine Learning o Intel·ligència Artificial.

- **Publicació** dels resultats obtinguts, ja sigui en forma de quadres de comandament, informes o petites APIs integrables amb les aplicacions del Consorci AOC i de les administracions locals.
- Capitalitzar els coneixements obtinguts per agilitzar moltes tasques que tradicionalment es porten a terme de forma manual mitjançant l'**automatització de processos robòtics** (RPA)
- **Formació i capacitat** al personal del Consorci AOC i als usuaris finals en l'ús dels models obtinguts durant el projecte

6 Finalitats i objectius que s'han d'assolir

Les finalitats i objectius que s'han d'assolir mitjançant la realització d'aquest contracte són els següents:

6.1 Finalitats i objectius del lot 1

Implantació de la plataforma de dades obertes de les administracions locals de Catalunya que, per tal que:

- Integri les dades obertes de totes les entitats locals i estigui federada amb la plataforma de la Generalitat de Catalunya, amb l'objectiu d'impulsar l'obertura d'aquestes dades, així com de facilitar-ne la reutilització per generar coneixement i desenvolupament a la societat.
- Impulsi l'obertura automatitzada de les dades locals ja disponibles als sistemes d'informació de gestió, prioritzant les dades agregades disponibles per entitats supramunicipals.
- Fomenti la reutilització de les dades per part d'administracions, entitats, empreses, professionals i ciutadania.

Creació del *data lake* de les administracions locals, federat amb el de la Generalitat de Catalunya, impulsant una infraestructura al núvol col·laborativa i compartida de dades relatives a l'activitat de totes les administracions catalanes, que disposi de:

- Un repositori comú de recollida de dades en brut, tant estructurades com no estructurades (*data lake*), per tal de facilitar la seva recollida i posterior categorització, consulta i tractament.
- Un *data warehouse* com a repositori de dades netes i estructurades dissenyat per la seva consulta i anàlisi eficient.
- Un servei comú de *business intelligence*, que permeti l'anàlisi descriptiu de les dades i ajudin a millorar la planificació, la predicció i la presa de decisions.
- Un servei d'anàlítica avançada de dades, basat en tècniques de *machine learning* i *deep learning* fent us de tecnologia *Big Data*, per impulsar serveis personalitzats i proactius, amb una visió transversal i integrada de l'Administració.

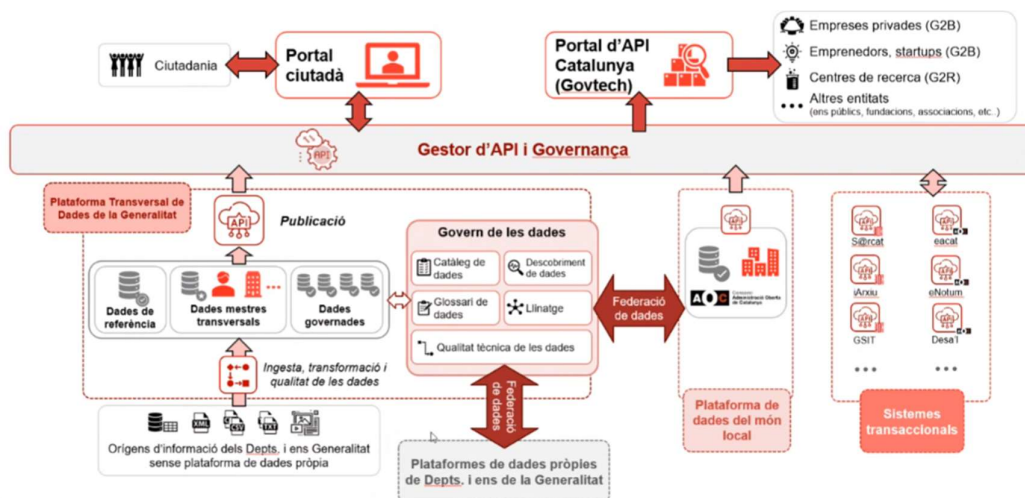


Figura 3. Model de federació amb Plataforma Transversal de Dades de la Generalitat de Catalunya

Serveis de *data warehouse* del Consorci AOC amb funcionalitats, com a mínim, equivalents a les definides a l'Annex I d'aquests plecs.

La plataforma de dades ha de complir a més els següents objectius generals:

- **Seguretat:** ha de tenir mesures de seguretat sòlides per protegir les dades i garantir la privadesa de la ciutadania.
- **Emmagatzematge escalable:** ha d'oferir un emmagatzematge escalable i rendible, i permetre als seus usuaris ampliar o reduir la capacitat d'emmagatzematge segons sigui necessari.
- **Integració i interoperabilitat:** ha de ser compatible i fàcilment integrable amb les aplicacions i sistemes existents dels ens locals.
- **Automatització:** les eines d'automatització de què ha de disposar han de permetre reduir la càrrega de treball dels ens per permetre'ls enfocar-se en altres prioritats.
- **Tecnologies avançades:** ha d'oferir tecnologies avançades per al processament de dades i l'anàlisi per permetre als ens extreure'n informació valuosa que puguin utilitzar per prendre decisions informades.
- **Suport tècnic:** ha de proporcionar suport tècnic i assistència als ens quan sorgeixin problemes o calgui ajuda.
- **Capacitat de recuperació davant de desastres:** ha de tenir mesures de recuperació davant de desastres per garantir que les dades segueixin estant disponibles en cas d'una interrupció del servei.
- **Facilitat d'ús:** ha de ser fàcil de fer servir per garantir que els ens puguin aprofitar-la de manera efectiva sense haver d'enfrontar grans barreres d'aprenentatge.
- **Compatibilitat de formats:** ha de ser capaç de gestionar múltiples formats de dades per allotjar una àmplia gamma de conjunts de dades.
- **Flexibilitat:** ha de ser prou flexible per adaptar-se a diferents necessitats i requeriments dels ajuntaments i ciutadans.
- **Informes personalitzats:** ha de ser capaç de generar informes personalitzats segons els requisits específics dels ens.
- **Federació:** ha de permetre l'intercanvi d'informació i coneixements entre diferents ens per fomentar el desenvolupament col·laboratiu.
- **Mobilitat:** ha de ser usable des de dispositius mòbils per permetre l'accés a la informació des de qualsevol lloc i en qualsevol moment.
- **Disponibilitat:** per garantir que la informació estigui disponible a tota hora, ha de tenir una alta disponibilitat i capacitat d'escalabilitat.
- **Anàlisi predictiva:** ha de tenir funcions d'anàlisi predictiva per identificar tendències i patrons futurs en la informació.

6.2 Finalitats i objectius del lot 2

Definir i implementar la governança de l'espai de dades, que haurà d'incloure:

- La definició d'ontologies, diccionaris o dades mestres del projecte, així com l'elaboració dels procediments per gestionar-los.
- La definició de les polítiques de custòdia, qualitat i perfilat i llinatge de les dades.
- Model de govern analític: haurà d'estandarditzar i afavorir la repetibilitat i l'automatització de tot el cicle de vida del desenvolupament i desplegament dels models analítics, que

donen resposta als casos d'ús. Per construir aquest model, s'han de definir i establir els procediments operatius i elements acceleradors, que permetin optimitzar els temps de desenvolupament i desplegament de models analítics, així com la seva monitorització i manteniment, una vegada desplegats.

- Mesures de seguretat i privadesa: s'han d'identificar i definir les mesures i polítiques de seguretat i privadesa (autorització, accés, auditoria de les dades, entre d'altres), a implementar en el lot 1 per donar compliment als requisits de seguretat i protecció de dades.
- L'adjudicatari serà responsable de les actualitzacions potencials que es puguin produir en la planificació per al desenvolupament i desplegament dels casos d'ús del projecte, durant el període de durada del contracte. Les actualitzacions hauran de ser aprovades pel Consorci AOC.

7 Requisits generals obligatoris

L'empresa contractista ha de disposar dels suficients mitjans tècnics, materials qualitius i personals per a desenvolupar les tasques objecte d'aquest contracte.

La prestació regulada en aquest plec ha d'ajustar-se, almenys, als requisits tècnics següents, sens perjudici dels paràmetres que s'han de valorar mitjançant els criteris d'adjudicació establerts.

7.1 Requisits generals obligatoris per a tots els lots

7.1.1 Normativa vigent de compliment obligat

L'empresa adjudicatària haurà de tenir en consideració els requisits establerts per la Llei 15/2022, de 12 de juliol, integral per a la igualtat de tracte i la no-discriminació, al seu article 23 i ja esmentats en aquest plec. Per altra banda, sent el Consorci AOC part del sector públic de Catalunya, es requerirà que l'empresa adjudicatària estigui certificada en l'Esquema Nacional de Seguretat, regulat actualment pel Real Decret 311/2022, de 3 de maig.

Finalment, i tenint en compte que són normes que encara no estan vigents, l'adjudicatari haurà de tenir present i considerar el que disposin tant les Propostes de Reglament del Parlament Europeu i del Consell COM/2022/68 final sobre normes harmonitzades per a un accés just a les dades i la seva utilització i COM/2021/206 final pel qual s'estableixen normes harmonitzades en matèria d'Intel·ligència Artificial.

7.1.2 Requisits referents a protecció de dades personals

Els tractaments realitzats en la plataforma de dades no implicaran en cap cas dades de caràcter personal. Així, es preveuen únicament dos tipus de fonts de dades: conjunts de dades procedents de processos d'anonimització, i conjunts de dades que no són personals en origen (com per exemple, és en principi la base de dades de geolocalització d'antenes de telefonia mòbil).

En relació als conjunts de dades procedents de processos d'anonimització, la plataforma de dades haurà de revisar l'eficàcia del tractament d'anonimització abans d'incorporar les dades de forma definitiva. En cas que es detecti la possibilitat de reidentificació, les dades no s'incorporaran a la plataforma.

7.2 Requisits obligatoris LOT1

L'especificació de requisits s'han agrupat en base a característiques o capacitats específiques requerides per la plataforma de dades objecte d'aquesta licitació. Aquests requisits han estat agrupats en els apartats següents, que l'empresa adjudicatària garantirà en la seva proposta cobrir en conjunt:

- Generals i d'arquitectura
- Ingesta
- Emmagatzematge
- Virtualització de la Dada
- Intel·ligència de negoci
- Analítica avançada
- Monitorització i Administració
- Seguretat

- Publicació i interoperabilitat
- *Data warehouse* del Consorci AOC

7.2.1 Generals i d'arquitectura

[R1-LOT1] La plataforma de dades s'ha de dissenyar, implementar, desplegar, mantenir i operar al *cloud* públic d'Amazon Web Services (AWS) de què disposa actualment el Consorci AOC. En aquest sentit, és el Consorci AOC qui proporcionarà i contractarà els serveis d'infraestructura AWS, no sent aquests objectes del present contracte.

[R2-LOT1] La plataforma ha de ser integral, estant tots els seus components, productes i/o serveis integrats sota un model de governança comuna centralitzada, i executant-se amb la màxima eficiència sobre la infraestructura proposada. Tots els components que formin part de la plataforma han de ser interoperables entre sí, permetent en el seu conjunt donar una resposta extrem a extrem, des de la captura i ingesta de les dades fins a la publicació dels serveis analítics.

[R3-LOT1] Si la solució proposada per l'empresa adjudicatària requereix de programari, llicències i/o dades addicionals per cobrir qualssevol dels requisits establerts en aquest plec, aquesta en proposarà la seva contractació per part del Consorci AOC, que podrà assumir els costos dins del concepte "Serveis d'aprovisionament" contemplat al pressupost base de licitació.

[R4-LOT1] Tots els components de la plataforma hauran de tenir suport empresarial del seu fabricant o distribuïdor exclusiu, i hauran de seguir estàndards oberts o d'àmplia adopció al mercat.

[R5-LOT1] Cap dels components de la plataforma s'ha de trobar en procés de discontinuïtat, descatalogació o fi de vida del fabricant. L'adjudicatari haurà de garantir la vigència i suport (actualització i suport de funcionalitats i protocols del producte) d'aquests components durant, com a mínim, la vigència del contracte.

[R6-LOT1] En el cas de ser necessari, durant l'execució del contracte, d'un canvi en algun component de la plataforma per motius d'actualització tecnològica, l'adjudicatari haurà de notificar-ho amb una antelació mínima de 15 dies al Consorci AOC, per tal de poder avaluar l'impacte al projecte i ser acceptat. En tot cas, l'adjudicatari haurà de mantenir els acords de nivell de servei establerts en aquests plecs en tot moment. Els eventuais canvis als components per actualitzacions tecnològiques hauran d'implicar que aquests s'hauran de substituir per components que compleixin tots els requisits estipulats, sense detriment de les característiques tècniques, funcionals o de qualsevol tipus respecte els components a substituir.

[R7-LOT1] La plataforma ha de ser resistent a fallides, sent capaç de funcionar en alta disponibilitat i, per tant, no tenint un únic punt de fallida.

[R8-LOT1] L'adjudicatari posarà a disposició dos entorns independents per la implantació de la plataforma:

- Entorn de producció: entorn real on s'emmagatzemaran les dades i a on tindran accés tots els usuaris de negoci i científics de dades per consultar els elements analítics i treballar sobre els mateixos.
- Entorn de desenvolupament: entorn preparat per tal que els equips tècnics puguin desenvolupar *scripts*, codi, regles, llançar proves, informes, etc. abans de pujar-ho a l'entorn de producció.

[R9-LOT1] Ha d'oferir una gestió de permisos multiusuari i multiens versàtil en la creació, lectura i modificació de les dades i dels espais de treball. Aquesta funcionalitat permet als diferents usuaris de la plataforma tenir diferents nivells d'accés i controlar els conjunts de dades disponibles per a diferents ens. La gestió de permisos ha de permetre diferenciar, com a mínim:

- Permisos per a un ens concret: la plataforma ha de permetre als administradors d'ens locals assignar permisos específics per a cada ens. Això significa que els usuaris d'un ens determinat tindran accés a les dades i funcionalitats relacionades amb el seu propi ens, però no tindran accés a les dades d'altres ens.

- Permisos per a un conjunt d'ens: els administradors també han de poder gestionar els permisos per a un conjunt d'ens. Això significa que poden assignar permisos a diversos grups d'ens, permetent als usuaris d'aquests grups accedir i treballar amb les dades compartides entre aquests ens.
- Permisos per a tots els ens: la plataforma ha de permetre establir permisos que afectin a tots els ens. Això significa que alguns usuaris o grups poden tenir accés a totes les dades i funcionalitats de la plataforma, independentment de l'ens al qual pertanyen.
- Permisos per a tot el públic: en alguns casos, pot ser necessari permetre l'accés a determinades dades o funcionalitats per a tot el públic. La plataforma ha de tenir la capacitat de gestionar aquest tipus de permisos, garantint que les dades accessibles per al públic estiguin adequadament protegides i es limiti l'accés no autoritzat.
- Permisos per als serveis del Consorci AOC: la plataforma ha de ser compatible amb els serveis del Consorci AOC, com ara el servei de Dades Obertes. Això significa que ha de proporcionar una integració que permeti als serveis del Consorci AOC accedir i utilitzar les dades específiques de la plataforma de manera adequada, amb els permisos i les restriccions establertes.

[R10-LOT1] La plataforma ha de permetre la connexió, com a mínim, amb les següents fonts de dades (tant en origen com en destí), estiguin aquestes allotjades tant en CPDs privats com en serveis al núvol públics:

- Bases de dades relacionals: Oracle, MS SQL Server, PostgreSQL, MySQL
- Bases de dades NoSQL: MongoDB
- *Datalakes* de tercers, inclòs aquells que facin servir HDFS
- Fitxers de *log*, d'imatge, fulles de càlcul, XML, JSON, CSV, ASCII, fitxers de compressió estàndard.
- APIs i Web Services (SOAP, REST, OData)
- Sistemes de fitxers FTP, SFTP i FTPS.
- Repositoris d'objectes al núvol: AWS-S3 i Azure Storage.
- Cues d'events i serveis de *streaming* en temps real.
- LDAP / Active Directory

7.2.2 Ingesta

[R11-LOT1] La plataforma ha de permetre capturar dades de diferents fonts d'origen en formats estructurats, semi-estructurats i no estructurats, podent assumir grans volums de dades amb fluxos massius i continus. Ha de permetre processar la informació capturada mitjançant capacitats d'alt cabdal de processament, en baixa latència, amb tolerància a fallides, consistència, gestió d'esdeveniments desordenats i amb alta disponibilitat a nivell d'operació i a nivell de persistència de la informació.

[R12-LOT1] La plataforma ha de permetre la càrrega de dades en temps real, propera a temps real o amb processos planificats, fent ús de tècniques de *streaming*, *micro-batch* i *batch*, a partir de múltiples fonts de dades, com les especificades al requisit [R10-LOT1].

[R13-LOT1] Ha de permetre tant la càrrega de dades en brut per la seva posterior transformació (ELT) com la realització de diferents transformacions prèvies (ETL). En aquest sentit, cal que disposi de funcionalitats de:

- Construcció de processos d'extracció, transformació i càrrega (ETL).
- Construcció de processos d'extracció, càrrega i transformació (ELT).
- Construcció de fluxos de dades i orquestració de processos, incorporant elements de fluxos de control, preparació de dades i avaluació de qualitat.

7.2.3 Emmagatzematge

[R14-LOT1] La plataforma ha de proveir una infraestructura d'emmagatzematge a gran escala, amb la capacitat de gestionar grans volums de dades, amb un flux massiu i continu, provinent de diferents orígens. Aquesta infraestructura de dades (*data lake*), basada en tecnologies *Big Data*, ha de permetre l'emmagatzematge eficient, en un repositori centralitzat, de dades estructurades, semi-estructurades i no estructurades amb fonts de dades heterogènies. La infraestructura ha d'estar orientada a la persistència de les dades, proporcionant mecanismes de recuperació en cas de fallida, permetent el seu creixement de forma escalable i la connexió i processament de dades, inclús des d'altres *data lakes*.

[R15-LOT1] La infraestructura ha de tenir la capacitat de funcionar com a *data warehouse*, sistema d'emmagatzematge de dades centralitzat i orientat a les consultes. Ha de permetre la recopilació, l'emmagatzematge i la gestió eficient de dades estructurades, com ara dades tabulars o relacionals. Això implica l'organització de les dades en taules i la possibilitat de realitzar consultes complexos, així com l'optimització del rendiment per a un accés ràpid i eficient a les dades.

[R16-LOT1] En el moment de carregar les dades, es proporcionarà a les mateixes la major granularitat al procés de captura i càrrega, amb l'objectiu de permetre un major i millor anàlisi. Així mateix, la plataforma ha de permetre que les dades es puguin emmagatzemar xifrades.

[R17-LOT1] Les dades emmagatzemades al *data lake* han de poder ser normalitzades i enriquides. En aquest sentit, la plataforma ha d'incloure funcionalitats d'extracció de metadades, conversió de formats, augment, extracció d'entitats, agregació, desnormalització o indexació.

[R18-LOT1] L'emmagatzematge de dades massives del *data lake* ha d'incloure, com a mínim, els següents sistemes (o d'equivalents):

- Sistemes de fitxes distribuïts (DFS).
- Sistemes d'emmagatzematge d'objectes S3.
- Sistemes de bases de dades relacionals.
- Sistemes d'emmagatzematge NoSQL amb estructures columnars, basades en documents, clau-valor i orientada a indexació.
- Fitxers JSON, dades geoespacionals i grafs.
- Qualsevol altre sistema que sigui necessari per als components proposats a la plataforma.

[R19-LOT1] Les dades que formin part del *data lake* han de poder ser accessibles per sistemes externs, mitjançant interfícies d'accés basades en estàndards oberts o d'àmplia acceptació, com APIs REST, JDBC, ODBC o tecnologia equivalent.

[R20-LOT1] La plataforma ha de suportar els formats de compressió estàndard (gzip, Avro, Parquet, ORC o tecnologies equivalents) per tal d'aconseguir l'optimització de l'espai utilitzat a l'emmagatzematge de les dades.

[R21-LOT1] La plataforma ha de proporcionar emmagatzematge permanent per a les dades. En aquest sentit, les dades emmagatzemades no s'han de perdre amb el temps i s'han de poder mantenir de forma segura i accessible durant llargs períodes de temps. L'emmagatzematge permanent és essencial per assegurar la consistència i la disponibilitat de les dades al llarg del temps, permetent l'anàlisi històrica, la modificació dinàmica d'esquemes, i la generació de tendències i informes a llarg termini.

[R22-LOT1] La plataforma ha de disposar de característiques transaccionals que permetin el processament i la gestió eficient de transaccions de dades. Aquestes característiques han de garantir la integritat, consistència i durabilitat de les operacions realitzades a l'entorn del *data lake*. En aquest sentit, cal que la plataforma disposi de transaccions atòmiques, control de concurrència, aïllament de transaccions, recuperació envers fallides, registre de transaccions i integritat referencial.

[R23-LOT1] La plataforma ha de tenir la capacitat de carregar, emmagatzemar i tractar dades de sistemes de sensorització, com els provinents de les anomenades *smart cities* (ciutats intel·ligents), tenint en compte el gran volum de dades i el flux en temps real que les caracteritzen.

7.2.4 Virtualització de la dada

[R24-LOT1] La plataforma ha de proporcionar una capa de virtualització de la dada, com a capa lògica entre les eines d'anàlisi i explotació de les dades i les tecnologies d'emmagatzematge. Aquesta capa ha d'oferir una visió integrada de les dades a les eines d'anàlisi i d'explotació, amb independència de les fonts i tecnologies utilitzades a l'emmagatzematge de les dades. Permetrà també la federació de diferents bases de dades o *data lakes* sense la necessitat d'integrar les seves dades al *data lake* propi de la plataforma, mostrant-les sota unes vistes i accessos unificats.

[R25-LOT1] La plataforma ha d'admetre mecanismes d'autorització i autenticació integrada d'usuaris i de rols amb permisos configurables i personalitzables, permetent:

- Connexió amb LDAP / Active Directory o tecnologies equivalents per a l'autenticació d'usuaris.
- Integració amb els serveis d'identitat digital orientats als treballadors públics i gestionats pel Consorci AOC.

Adicionalment, s'haurà de proporcionar suport per a:

- Single sign-on (SSO)
- Autenticació de dos factors (2FA)
- Autoritzacions basades en usuaris i/o rols (RBAC)

[R26-LOT1] La capa de virtualització de la dada ha d'estar integrada amb les eines de governança de la dada de la plataforma, de manera que les funcionalitats sol·licitades es compleixin en la seva totalitat de forma coherent. En particular, la gestió de rols i permisos de la capa de virtualització de la dada ha d'estar sincronitzada amb la capa de gestió de rols i permisos a la governança de la dada, per no duplicar la seva definició ni generar incoherències a l'hora de definir polítiques d'accés a la plataforma. Aquesta integració ha de ser compatible, com a mínim, amb les eines de governança de la Generalitat de Catalunya per tal de garantir la federació de dades. La Generalitat fa servir els productes Axon, EDC i IDQ del fabricant Informatica.

[R27-LOT1] La plataforma ha de disposar d'una interfície gràfica d'usuari (*GUI*) que permeti efectuar de forma ràpida i senzilla la connexió als orígens de dades i permeti combinar-los i transformar-los, podent crear vistes virtuals dels mateixos i publicar-los en diferents formats per poder accedir-hi mitjançant un conjunt gràfic de menús, assistents, accions de *drag-and-drop* i vistes d'arbre.

[R28-LOT1] La plataforma ha de permetre establir vistes per domini o context de dades que, a més, puguin compartir dades entre ells. L'emascarament de les dades es podrà definir a baix nivell, arribant fins el nivell de columna d'una taula i en funció de regles de negoci establertes.

[R29-LOT1] Les eines d'informes i d'intel·ligència de negoci, de governança de la dada i de publicació i interoperabilitat han de poder accedir a la capa de virtualització de la dada per tal de poder obtenir de forma ordenada la informació federada, mitjançant interfícies d'accés basades en estàndards oberts o d'àmplia acceptació, com API REST, JDBC, ODBC, etc. i poder aplicar polítiques d'accés a les dades.

7.2.5 Intel·ligència de negoci

[R30-LOT1] La plataforma ha de permetre als usuaris, de forma senzilla, autònoma i sense coneixements de programació, analitzar les dades en temps real, interactuant directament amb una representació visual de les mateixes, que ha d'incloure:

- Objectes visuals senzills: taules, gràfics de barres i d'àrees, circulars, etc.

- Objectes visuals complexes a on poder presentar les dades d'una forma que ofereixi context i informació detallada i visual, a on atributs com el color i les formes proporcionin informació addicional a la dada que es vulgui representar.
- Gràfics preconfigurats: mapes de color, bombolles, geomapes, mapes d'arbre, gràfics de dispersió, etc.
- Quadres de comandament que agrupin els elements anteriors.

[R31-LOT1] La plataforma ha de tenir una integració compatible amb Microsoft Power BI, la principal eina d'informes externa utilitzada pel Consorci AOC tot permetent mantenir la compatibilitat amb els informes actuals. La plataforma ha de permetre exportar les dades i els indicadors clau de rendiment (KPIs) cap a Power BI, assegurant la compatibilitat i la sincronització regular per a la visualització i l'anàlisi dels informes generats.

[R32-LOT1] Els informes i quadres de comandament generats han de poder ser distribuïts mitjançant APIs REST, JDBC, ODBC o tecnologies equivalents, així com ser exportables en formats d'ús comú com PDF, XML, XLS, CSV, HTML i d'altres.

[R33-LOT1] Els informes i quadres de comandament han d'estar optimitzats per al seu consum en dispositius mòbils de forma interactiva, aprofitant les capacitats natives d'aquests dispositius. Així mateix, la plataforma ha de permetre inserir-los de forma senzilla a portals web per la seva consulta pública i oberta, sense necessitat d'identificació. Els informes i quadres de comandament que es publiquin a portals web han de permetre el filtratge i segmentació de dades per defecte de forma configurable, per permetre, per exemple, diferents visualitzacions en funció de l'ens públic que els utilitzi.

[R34-LOT1] La plataforma ha de proporcionar una eina amb entorn gràfic de treball que permeti:

- Modelar processos d'intel·ligència de negoci, anàlisi i generació d'informes basats en les dades emmagatzemades.
- Facilitar el desenvolupament ràpid de processos i fluxos de dades sense necessitat de disposar d'un coneixement profund del sistema; l'anàlisi en modalitat autoservei; l'accés, combinació i preparació de dades i l'exploració visual, cerca i compartició de patrons.

[R35-LOT1] La plataforma ha de fer recomanacions als usuaris que explorin dades i dissenyin informes, en forma de suggeriment de visualitzacions generades de forma automàtica i depenent de la naturalesa de les dades explorades. Així mateix, ha de permetre establir de forma automàtica correlacions i clústers de prediccions sense requerir coneixements de *machine learning* ni algorítmics a l'usuari, i permetre cerques i consultes en llenguatge natural.

7.2.6 Anàlisi avançada

[R36-LOT1] La plataforma ha de permetre la presa de decisions a partir de patrons, classificacions i aprenentatge. Per aquesta presa de decisions es faran servir models de *machine learning* o *deep learning* tant en capacitats d'aprenentatge supervisat (classificació, regressió, etc.) com no supervisat (*clustering*, detecció d'anomalies, etc.). El processament de les dades associades ha de poder ser en modes *streaming*, *batch* i *microbatch*.

[R37-LOT1] La plataforma ha de permetre el desenvolupament i execució de *scripts* i models de *machine learning* i *deep learning* amb, com a mínim, els llenguatges de programació Python, R i Scala, i fent ús de Jupyter Notebooks i/o R Studio.

7.2.7 Monitorització i administració

[R38-LOT1] La plataforma ha de disposar d'una consola visual d'administració, que permeti realitzar accions de gestió, i de capacitats de monitorització, de forma que es pugui conèixer en tot moment l'estat de la plataforma, en relació als seus recursos i l'estat dels seus components, incorporant eines amb interfície d'usuari que facilitin les diferents activitats de monitorització.

[R39-LOT1] La plataforma disposarà de, com a mínim, les següents funcionalitats de monitorització i administració:

- Monitorització a nivell de serveis, *tenant* i càrregues de treball.
- Consulta de paràmetres habituals com ús de processador, memòria, paràmetres de màquina virtual o contenidor, estat d'emmagatzematge, indicadors d'entrada i sortida i taxes de transferència.
- Configuració d'alertes personalitzades per notificar a les persones administradores quan determinats indicadors superin els llindars configurats.
- Configuració del registre d'esdeveniments de seguretat per donar compliment als requisits de seguretat.
- Accés als registres (*logs*) de la plataforma.
- Distribució, de forma òptima i eficient, de les càrregues de treball de llarga execució (com les generades per les eines de ciència de dades) que requereixin d'alt temps de disponibilitat de recursos del clúster, tant de forma manual com automàtica.

7.2.8 Publicació i interoperabilitat

[R40-LOT1] La plataforma ha de ser interoperable amb la plataforma de dades de la Generalitat de Catalunya, aspecte clau per assegurar una gestió eficient i coordinada de les dades entre diferents organismes públics a Catalunya. Per aconseguir aquesta interoperabilitat, l'empresa contractista ha de disposar de les habilitats i coneixements necessaris per integrar la infraestructura de dades amb la plataforma de dades de la Generalitat de Catalunya. Això inclou la comprensió dels protocols, formats i estàndards de comunicació, així com la capacitat per adaptar la infraestructura a les especificacions de la plataforma. L'empresa contractista també ha de proporcionar la documentació necessària i el suport tècnic per assegurar una integració correcta i el manteniment de la interoperabilitat amb la plataforma de dades de la Generalitat de Catalunya.

[R41-LOT1] Addicionalment a les interfícies definides a requisits anteriors, la plataforma ha de disposar d'una eina de gestió d'APIs que permeti als sistemes de negoci la integració amb la plataforma, així com la publicació de conjunts de dades (*datasets*) accessibles per a tercers. Aquesta eina ha de permetre la gestió i publicació controlada i segura de les APIs a tercers, i haurà d'estar integrada amb el Gestor d'APIs i Governança de la Generalitat de Catalunya.

7.2.9 Data warehouse del Consorci AOC

[R42-LOT1] La plataforma ha de disposar de les característiques tècniques i funcionals que permetin donar servei al *data warehouse* del Consorci AOC (descrites a l'annex I d'aquests plec), incloent totes les parts del procés: Ingesta de diferents fonts d'origen, transformació i tractament de les dades, càrrega i emmagatzematge a nivell estadístic al *data warehouse*, generació i publicació d'informes públics, publicació en format de dades obertes i posada a disposició de les dades per altres serveis del Consorci AOC.

[R43-LOT1] L'adjudicatari serà el responsable de configurar la infraestructura, carregar les dades, els processos i les configuracions actuals del Consorci AOC, descrites a l'annex I d'aquests plec i al requisit anterior, cap a la nova plataforma de dades reduint al mínim l'impacte en els processos actuals del *data warehouse* actual del Consorci AOC

[R44-LOT1] Tenint en compte la complexitat interna de les aplicacions i serveis del Consorci AOC i de les bases de dades MS SQL Server que formen part del *data warehouse* actual, aquestes s'hauran de mantenir, en una primera fase, en bases de dades MS SQL Server a la nova plataforma de dades. Posteriorment, les bases de dades es podran migrar a un altre sistema de base de dades, si es considera oportú.

7.2.10 Govern de la dada

[R45-LOT1] La plataforma ha de proveir una eina de governança de la dada , per respondre a les necessitats del govern de les dades, incloent-hi aspectes com les polítiques d'accés, la qualitat, el perfilat i el llinatge de les dades, que haurà de permetre:

- Dur a terme el perfilat de les dades, permetent la neteja, filtrat i estandardització, així com l'obtenció de valors de taules mestres, si cal.
- Realitzar un llinatge de les dades, de manera que es pugui conèixer el flux pel qual han passat, des del seu origen, i les transformacions sofertes, abans de ser posats a disposició per al consum. Així mateix, haurà de contemplar l'anàlisi d'impacte que permeti descobrir dependències entre objectes.
- Dur a terme el versionat de les dades, sobre les quals es realitzen les transformacions, amb el objectiu de poder reutilitzar les dades per als diferents casos d'ús que es despleguin.
- Classificar i etiquetar les dades.
- Establir ontologies, catàleg de dades, glossari de termes, diccionaris i dades mestres, que facilitin les transformacions i validacions per garantir la qualitat de la dada i la seva estandardització.
- Descobriments automàtics de les dades, de manera que els usuaris puguin conèixer quines dades tenen disponibles per al consum.
- Autodescobriments de termes de negoci amb què agilitzar la catalogació d'actius.
- Facilitar el suport a les diverses auditories, registrant les alertes i la resolució dels problemes detectats a les dades.

[R46-LOT1] Haurà d'oferir una visió completa de l'ús i la gestió de les dades, que permeti configurar la estratègia, polítiques, exigències normatives i estàndards que requereixi l'ecosistema analític a l'hora de posar a disposició dels diferents usuaris la informació de qualitat d'una manera intuïtiva, àgil, escalable, interoperable i segura.

[R47-LOT1] Permetrà una configuració i administració de la plataforma en base als rols, responsabilitats i atributs (com ara etiquetes) definits dins del model de govern. Es requereix per tant que la gestió d'àrees de negoci, dominis i capes d'informació es pugui fer mitjançant fluxos de validació de manera gràfica i configurable tant a nivell global com a nivell de domini concrets.

[R48-LOT1] Es podran definir múltiples dominis o espais de dades, que continguin la informació concreta de la plataforma que es requereixi per a aquest domini, segons les polítiques d'accés definides, emmascarament de la informació necessària, processos de transformació, regles de negoci etc. Alhora, aquests dominis podran compartir informació entre ells, respectant les polítiques de accés que es defineixin.

[R49-LOT1] S'haurà de cobrir les diferents dimensions relacionades amb la qualitat de la dada, com ara la completesa, consistència, precisió, integritat i la pertinència dels mateixos. La solució haurà d'aportar funcionalitats que garanteixin la qualitat de les dades, permetent al menys:

- Detecció d'errors en informació ingestada.
- Duplictat de dades i generació de Golden Record.
- Detecció d'incongruències a les dades, dades buides o dades il·legibles.
- Disposar de regles de qualitat predefinides (per exemple, quant a completesa, format i integritat, entre d'altres).
- Disposar de la capacitat d'implementació de noves regles de qualitat ad-hoc, així com de capacitats relacionades amb l'exportació i la reutilització a través d'APIs.
- Definir regles de negoci sense necessitat de disposar de coneixements experts en codificació i programació.

- Tractament de dades assegurances durant el procés de validació de la qualitat (encriptació a curs i en emmagatzematge).

[R50-LOT1] Caldrà incloure mecanismes de seguretat que centralitzin i facilitin la gestió de la seguretat perimetral i gestió de l'autenticació, autorització i auditoria. La solució incorporarà funcionalitats relacionades amb la seguretat i la gestió d'usuaris, permisos i polítiques d'accés:

- Eines i/o mòduls que permetin l'anonimització, pseudonimització, l'emascament i el xifratge de les dades.
- Capacitats d'integració amb sistemes de gestió de l'autenticació i autorització com LDAP, SAML, Kerberos, OAuth2 i OpenID Conect, o tecnologies equivalents.
- Capacitats per realitzar una gestió de polítiques d'autorització i auditoria d'accés usuaris basada en rols i amb una alta granularitat en l'assignació de permisos, permetent fins i tot assignar aquests permisos a nivell de camps, o metadades específiques.

[R51-LOT1] Haurà d'incloure un motor de cerca avançada i intel·ligent (a partir de filtres dinàmics i els algorismes de cerca corresponents, com per exemple, cerques facetades. Aquesta cerca es podrà fer amb un llenguatge tant tècnic, com de negoci, amb l'objectiu de posar a disposició de l'usuari aquella informació a què pugui accedir pel seu rol i responsabilitats determinades.

[R52-LOT1] Haurà d'incloure un repositori centralitzat de metadades que serà accessible via API per a un consum autogestionat de tercers, entre els quals s'inclourà:

- Descobrimet, extracció i associació automatitzada de metadades tècniques i de negoci.
- Traçabilitat i llinatge a nivell de camp.
- Anàlisi automàtica i gestió d'impacte associada als canvis.
- Catàleg de dades i glossari de termes personalitzat, organitzat per orígens de informació.
- Capacitat d'importació de termes a partir de plantilles externes amb les extensions més comunament utilitzades (csv, Excel i equivalents).
- Suport de multitud de connectors a sistemes origen, de manera que permetin la connectivitat d'ampli espectre.

[R53-LOT1] Haurà de permetre la configuració de fluxos de treball de validació en funció del model de govern que es defineixi en base a les següents activitats:

- Implementació de polítiques, processos i procediments del govern de la dada en funció dels rols i les accions que s'executin.
- Implementació de workflows d'autorització/aprovació de continguts.
- Sistema d'alertes i notificacions que faciliti la personalització, col·laboració i la interacció entre els usuaris de l'eina.
- Traçabilitat i control dels passos executats per a un rigorós seguiment, control i auditoria.
- Integració amb altres solucions externes a l'aplicació, com ara eines d'ingesta, virtualització, explotació de dades i APIs.
- Incorporació de regles lògiques de negoci per a l'automatització dels fluxos de treball que siguin necessaris.
- Personalització i suport dels rols del Model de Govern.

[R54-LOT1] Haurà d'oferir la funcionalitat d'Informes i/o quadres de comandament per monitoritzar l'operativa de la eina de Govern de la Dada.

[R55-LOT1] L'eina de govern de la dada estarà integrada, com a mínim, amb la solució de governança de la Generalitat de Catalunya per tal de garantir la federació de dades. La Generalitat fa servir els productes Axon, EDC i IDQ del fabricant Informatica.

7.2.11 Descripció de mitjans personals

La relació de perfils de l'equip de treball i les funcions a desenvolupar per cadascun són les següents:

Cap de projectes:

- Dirigir, gestionar i coordinar el contracte, els serveis/projectes inclosos i els equips.
- Planificar i controlar la prestació i el lliurament dels treballs contractats.
- Gestionar els riscos i els plans de mitigació o contingència.
- Gestionar la facturació i el finançament del servei.
- Reunions de coordinació amb els responsables del Consorci AOC.

Arquitecte / Consultor sènior:

- Establir la visió tècnica global dels equips de desenvolupament.
- Definició global d'arquitectures, processos i models de suport a la gestió del cicle de la dada (ingesta, transformació, emmagatzematge, processament i explotació).
- Garantir el funcionament correcte de la plataforma d'extrem a extrem amb els seus corresponents plans de proves.
- Determinar requisits d'usabilitat i protecció de dades
- Anàlisi funcional.
- Determinar requisits normatius, tècnics i jurídics, a aplicar en tractament de la informació i privadesa.

Desenvolupador / Consultor júnior:

- Donar suport a l'arquitecte de dades / consultor sènior i al cap de projectes mitjançant els desenvolupaments corresponents.
- Dissenyar i desenvolupar APIs i processos ETL.
- Dissenyar i desenvolupar informes d'intel·ligència de negoci.

7.3 Requisits obligatoris LOT2

7.3.1 Govern de la dada

[R1-LOT2] L'adjudicatari haurà de definir la governança del servei, ajudant a definir les polítiques i els protocols necessaris per a garantir la qualitat i la seguretat de les dades. En concret, la consultoria ajudarà a definir les polítiques de gestió de dades, les polítiques de seguretat, les polítiques de privacitat i les polítiques d'accés a les dades. D'aquesta manera

es pretén evitar problemes com la pèrdua de dades, les vulnerabilitats de seguretat i les infraccions de privacitat.

[R2-LOT2] L'adjudicatari haurà d'implementar el model de governança definit emprant l'eina de governança de la dada proveïda al lot 1 utilitzant totes les funcionalitats descrites a l'apartat 7.2.10, per respondre a les necessitats del govern de les dades, incloent-hi aspectes com les polítiques d'accés, la qualitat, el perfilat i el llinatge de les dades.

7.3.2 Marc de treball

[R3-LOT2] Al començament dels treballs, l'adjudicatari definirà i elaborarà un Marc de Treball comú per abordar el desenvolupament i desplegament de casos d'ús, que inclourà la definició de les mesures de seguretat i privadesa, així com els models de govern del dada i govern analític a implantar i configurar a l'Espai de Dades, donant compliment a la normativa aplicable. Es recullen, en aquest apartat, els principals aspectes que ha de contemplar aquest Marc de Treball, podent organitzar o completar l'adjudicatari aquesta relació com consideri adequat.

[R4-LOT2] Lliurables del servei:

- Model de govern de la dada.
- Model de govern analític.
- Mesures de seguretat i privadesa a implementar.
- Qualsevol altre aspecte que l'adjudicatari inclogui en la seva proposta.

7.3.3 Fases

El conjunt d'activitats definides en aquest apartat per als serveis de desenvolupament i implementació de casos d'ús s'estructuraran en les següents fases:

- Fase d'avaluació de la Viabilitat, associada a l'execució, de manera iterativa, de les activitats necessàries per avaluar la viabilitat d'un cas d'ús.
- Fase de desplegament, associada a l'execució, de manera iterativa, de les activitats necessàries per al desenvolupament, la implementació i el desplegament d'un cas d'ús viable.

7.3.4 Detecció i estudi de casos d'ús

[R5-LOT2] Al començament dels treballs, l'adjudicatari realitzarà el conjunt d'accions que calguin per poder tenir una primera valoració de les dades gestionades per l'Espai de Dades així com una definició de casos d'ús d'analítica avançada. En qualsevol cas:

- Es basen en les fonts de dades gestionades per l'Espai de Dades, o bé s'especifica i es desenvolupa la seva integració.
- Es basen en les capacitats de la solució que han estat proporcionades.
- Es disposa una certesa raonable, sobre la base del coneixement ja disponible de les dades, respecte a la seva viabilitat.
- Per realitzar la identificació de nous casos d'ús, El Consorci AOC podrà requerir que es facin dinàmiques de treball amb diferents interlocutors del Consorci AOC i de les Administracions Públiques Catalanes.

El Consorci AOC haurà d'aprovar els casos d'ús presentats i establirà la seva prioritització, amb l'objectiu d'optimitzar l'obtenció de resultats, en el menor temps possible, contribuint als objectius de negoci.

[R6-LOT2] L'adjudicatari haurà, almenys:

- Definir el conjunt de criteris que emprà per a l'anàlisi inicial de casos d'ús.

- Realitzar una anàlisi inicial dels casos d'ús, en base als criteris definits, i obtenir una primera valoració respecte a la seva viabilitat.
- Prioritzar els casos d'ús, d'acord amb el que el Consorci AOC especifiqui.

En cas que, com a resultes de l'anàlisi inicial, s'evidenciessin dificultats en l'assoliment d'algun cas d'ús, l'adjudicatari haurà de proposar les reorientacions que es considerin necessàries, per optimitzar l'obtenció de resultats, en el menor temps possible, sempre contribuint als objectius de negoci.

[R7-LOT2] Com a fruit de la primera fase del projecte, l'adjudicatari lliurarà al Consorci AOC un document especificant:

- Criteris d'anàlisi i prioritització de casos d'ús.
- Anàlisi inicial i prioritització dels casos d'ús.
- Planificació per al desenvolupament i el desplegament dels casos d'ús.
- Qualsevol altre aspecte que l'adjudicatari inclogui a la seva proposta.

7.3.5 Serveis de desenvolupament i implementació de casos d'ús

[R8-LOT2] L'adjudicatari haurà de portar a terme les tasques associades al desenvolupament i la implementació de casos d'ús d'anàlisi avançada, a la monitorització i manteniment dels models analítics (algorismes programari), que donen resposta als casos d'ús, i a la capacitat als usuaris finals, consumidors dels serveis analítics, derivats dels casos d'ús.

[R9-LOT2] Com a resultat de cada cas desenvolupat, l'adjudicatari lliurarà al Consorci AOC:

- Informes amb el coneixement adquirit de les dades, amb informació com el llistat i volums de dades adquirides (nombre de registres i camps per registre).
- Scripts desenvolupats per a la planificació i sistematització de processos d'anàlisi.

[R10-LOT2] L'adjudicatari haurà de garantir l'explicabilitat i la reducció de biaix en tots els models analítics que es desenvolupin en el marc del projecte, per donar resposta als casos d'ús, recolzant-se en les capacitats que es requereixen, a aquest efecte, a la capa analítica de la Solució.

[R11-LOT2] El licitador haurà d'identificar aquells algorismes sobre els quals el Consorci AOC pot exigir-ne la propietat intel·lectual, per no haver estat desenvolupats en el marc del contracte. Els models analítics, un cop implementats sobre la Solució i desplegats per utilitzar-los, podran ser consumits pels ens usuaris dels serveis del Consorci AOC, sense més limitacions que les capacitats de recursos maquinari que pugui tenir la Solució.

[R12-LOT2] La consultoria objecte d'aquest lot podrà abordar un màxim de dos casos d'ús en paral·lel.

[R13-LOT2] En cas que l'adjudicatari proposi la incorporació de fonts de dades externes a les disponibles des de l'Espai de dades, que siguin necessàries o d'utilitat per al desenvolupament i desplegament d'algun cas d'ús, s'incorporaran al data lake de la Solució. Qualsevol font externa que l'adjudicatari proposi incorporar al data lake de la solució, haurà de complir les condicions de seguretat i protecció de dades amb acord a la normativa aplicable.

Activitats associades al desenvolupament i desplegament de cada cas d'ús

Les activitats que es recullen a continuació es basen en el model estàndard CRISP-DM (Cross Industrial Standard Process for Data Mining), que estableix les fases necessàries per cobrir el cicle de vida propi de les tècniques de mineria de dades, i que són comunament aplicades al desenvolupament de models analítics.

Així mateix, la prestació d'aquest servei s'haurà de basar en:

- [R14-LOT2] El conjunt de millors pràctiques, definides per MLOps (Machine Learning Ops), a l'hora de desenvolupar models de machine Learning, amb l'objectiu de la posada en producció.
- [R15-LOT2] S'empraran metodologies de desenvolupament àgil, en concret es farà ús de SCRUM. Es duran a terme lliuraments continus, al llarg del transcurs del servei, els quals hauran de ser operatius, iteratius i incrementals, és a dir, entre una iteració (SPRINT) i la següent s'hauran d'introduir noves funcionalitats al desenvolupament, fins a completar l'abast del cas d'ús.

Coneixement del negoci

La definició de cada cas d'ús objecte del present lot contindrà com a mínim:

- [R16-LOT2] Els objectius de negoci, els criteris de negoci per a l'avaluació del cas d'ús, els objectius analítics, la complexitat, els passos a seguir i les tècniques analítiques a emprar per al desenvolupament i la implementació de cada cas.
- [R17-LOT2] Els indicadors i criteris que permetin valorar si es respon amb èxit al problema que es vol resoldre. Per això, es comptarà amb la participació de perfils tècnics i de negoci del Consorci AOC i, si s'escau, dels ens del món local.
- [R18-LOT2] Elaboració del pla de treball i el seu calendari. Així mateix, es revisarà i actualitzarà la planificació general del projecte, realitzada en el servei d'anàlisi inicial i prioritització.
- [R19-LOT2] La definició de cada cas d'ús així com del pla de treball que l'implementarà haurà de ser aprovat pel Consorci AOC.

Coneixement, recopilació i descripció de les dades

- [R20-LOT2] Caldrà identificar clarament les dades que seran objecte del cas d'ús i la seva adequació al processament futur. Aquesta tasca implica l'elaboració d'informes, amb la llista de les dades adquirides i la localització. Si les dades no es troben a l'Espai de Dades objecte d'aquesta licitació caldrà especificar també les tècniques utilitzades en la recollida, els volums de dades (nombre de registres i camps per registre), el significat i el format de cada camp, etc. En el cas que sigui necessari, i l'AOC ho determini, el licitador haurà de proveir les dades a càrrec del pressupost definit per aquesta finalitat.
- [R21-LOT2] Es podrà portar a terme un anàlisi exploratori de les dades objecte de cada cas d'ús revelar les seves propietats.
- [R22-LOT2] S'efectuaran verificacions sobre la qualitat de les dades per determinar la consistència dels valors individuals dels camps, la quantitat i la distribució dels valors nuls, la localització de valors fora de rang, etc., amb l'objectiu d'assegurar la completesa i correcció de les dades.
- [R23-LOT2] Se seleccionarà el subconjunt de les dades adquirides a la fase anterior, que estigui relacionat amb les tècniques de modelatge analític que s'aplicaran. Si s'escau, les dades seleccionades es podran pre-processar fent servir tècniques com:
 - Optimitzara la qualitat de les dades, a fi de preparar-les per a la fase de modelatge analític, mitjançant tècniques com la normalització de les dades, la discretització de camps numèrics, el tractament de valors absents o la reducció del volum de dades, entre d'altres.
 - Es podran realitzar operacions de preparació de les dades, com ara la generació de nous atributs a partir d'atributs ja existents o la transformació de valors per a atributs existents.

- Es podran realitzar transformacions sintàctiques de les dades, sense modificar-ne el significat, amb la idea de permetre o facilitar l'ús d'alguna tècnica de modelatge analític en particular.

Modelat analític

[R24-LOT2] Es seleccionarà/n la/les tècnica/es de modelatge analític més apropiada/es al tipus de problema a resoldre en cada cas d'ús. Prèviament a la construcció del model analític, es definirà el mètode d'avaluació a emprar per determinar la qualitat i la validesa (en base a paràmetres com el rendiment, fiabilitat, robustesa o explicabilitat, entre d'altres) del model analític.

[R25-LOT2] Una vegada seleccionada/des la/les tècnica/es de modelatge analític, es construirà el model i s'executarà sobre les dades prèviament preparades. Es poden generar, en aquesta fase, un o més models analítics.

[R26-LOT2] S'interpretaran els resultats de l'anàlisi, tenint en compte el coneixement preexistent i els criteris preestablerts. Es valoraran factors com l'exactitud i la generalitat del model. e) Avaluació dels resultats

Avaluació dels resultats

[R27-LOT2] De forma conjunta amb els responsables del servei del Consorci AOC, es realitzarà l'avaluació del model analític, en relació amb els indicadors establerts, i es determinarà si hi ha alguna raó per a la qual el model sigui deficient, o si es considera òptim per passar a la fase de desplegament. En concret:

- S'avaluarà el procés complet de modelatge, a fi d'identificar elements que es puguin millorar.
- S'establiran els passos següents a la llum de l'avaluació dels resultats davant dels objectius. S'avaluarà la viabilitat del cas d'ús i, prèvia aprovació del Consorci AOC es decidirà si s'han generat resultats satisfactoris i es pot de passar a la fase de desplegament.
- Si els resultats no són satisfactoris o es detecta alguna oportunitat de millora, s'aborda una altra iteració des de la fase de preparació de dades o des de la fase de modelatge, tenint en compte els resultats obtinguts.
- En cas que es detecti la inviabilitat del cas d'ús, es proposarà la seva substitució per un de complexitat equivalent i es començarà amb un nou projecte de modelatge analític. Per proposar la substitució d'un cas d'ús, l'adjudicatari elaborarà un informe on es detallin els treballs realitzats, es determinin les causes per les quals el cas d'ús no és viable, justificant el nombre d'iteracions realitzades i identifiquin les condicions que haurien de donar-se perquè el cas d'ús pogués ser viable. L'informe haurà de ser validat per part del Consorci AOC.
- Si els resultats són satisfactoris es dissenyarà i elaborarà un pla d'implantació, que inclourà les tasques a realitzar per a la publicació o consum del model analític (desplegament del cas d'ús), que haurà de ser aprovat pel Consorci AOC prèviament al desplegament. Els resultats dels models analítics s'emmagatzemaran o desplegarà a l'Espai de dades, des d'on podran ser consultats pels usuaris finals, a través de les capacitats d'explotació de la solució, o ser consumits per altres sistemes/serveis.

Desplegament

[R28-LOT2] L'adjudicatari abordarà les tasques associades al consum dels models analítics mitjançant:

- El desenvolupament d'APIs, que permetin la publicació del model analític, així com el consum dels serveis analítics exposats per altres sistemes d'informació de l'organització.

- La construcció d'informes i visualitzacions ad-hoc, que facilitin la presentació de resultats i consum per part dels usuaris.

Pla de monitorització i manteniment del model

[R29-LOT2] Es dissenyaran i elaboraran les estratègies de monitorització i manteniment del model, una vegada desplegat, les quals permetran valorar si el model està funcionant i sent utilitzat apropiadament, emprendre les tasques de manteniment que siguin necessàries i controlar les diferents versions del model.

Revisió del projecte analític

[R30-LOT2] S'avaluarà tot el projecte de desenvolupament i desplegament del cas d'ús i s'identificaran lliçons apreses i accions de millora. Es realitzarà un informe final de conclusions que cobrirà, per a cada cas d'ús, almenys, els aspectes següents:

- Informes associats al coneixement de les dades.
- La construcció i l'ús del model analític: procés d'elaboració, fonaments, resultats, ús, consum i conclusions del model, entre d'altres aspectes.
- El pla d'implantació.
- El pla de monitorització i manteniment del model.
- Lliçons apreses i accions de millora, associades a la fase de revisió del projecte.

Capacitació associada als casos d'ús

[R31-LOT2] L'Adjudicatari oferirà un servei de formació sobre aquells casos d'ús en què sigui necessari impartir capacitació sobre la publicació o consum del model analític associat. En aquests casos, es durà a terme la capacitació adreçada als usuaris finals o consumidors dels serveis analítics derivats de cada cas d'ús. Aquesta capacitació es basarà principalment en l'elaboració de manuals, guies d'usuari o vídeos autoexplicatius, tot i que podrà ser necessari recórrer puntualment a sessions presencials.

Lliurables

En resum, el conjunt de lliurables associats a un cas d'ús viable seran:

- Pla de treball de cada cas d'ús.
- Mètode d'avaluació del model analític per determinar-ne la qualitat i la validesa.
- Datasets emprats a l'entrenament i la validació dels models analítics.
- Codi font, desenvolupat en el marc del contracte, dels processos ETL, els algorismes (models analítics) o les API per a consum dels models, entre d'altres.
- Informe final o de conclusions, que ha d'incloure tots els lliurables associats a les seves versions finals.
- Documentació i elements audiovisuals associats a la capacitació.

Si es determina la inviabilitat d'un cas d'ús, els lliurables seran:

- Pla de treball de cada cas d'ús. Informes associats al coneixement de les dades.
- Mètode d'avaluació del model analític per determinar-ne la qualitat i la validesa.
- Datasets emprats a l'entrenament i la validació dels models analítics.
- Codi font, desenvolupat en el marc del contracte, dels processos ETL, els algorismes (models analítics).

- Informe de necessitat de substitució del cas d'ús.

7.3.6 Servei de Capacitació

Aquest servei inclou les accions de capacitació adreçades als professionals del Consorci AOC i usuaris finals dels models analítics desenvolupats. Les accions de capacitació, incloses en aquest servei, han de ser graduals i continuades durant tot el període d'execució del contracte, i no s'ha de focalitzar només en els moments de desplegament dels casos d'ús.

[R32-LOT2] Les accions de capacitació seguiran les següents directrius:

- S'han d'utilitzar models de capacitació presencial i en línia.
- Per a cada grup d'usuaris es definiran activitats formatives específiques, indicant les propostes de temaris i els perfils del personal docent que les impartirà, sense perjudici que part del temari que contingui cada activitat formativa pugui ser coincident.
- La durada i el nombre de les activitats formatives, la forma d'impartició i el nombre d'assistents mínims i màxims a aquestes, s'han d'adequar a les necessitats del temari de cadascuna de les propostes dissenyades per l'adjudicatari.
- L'adjudicatari realitzarà una proposta de cadascuna de les activitats formatives específiques (Pla de capacitació), que haurà de ser aprovada pel Consorci AOC.

Lliurables

Els lliurables relacionats amb el servei de capacitació inclouran:

- Materials de formació elaborats per a les sessions presencials i de capacitació en línia: presentacions, casos pràctics o vídeos, entre d'altres.
- Manuals d'usuari final, si s'escau.

7.3.7 Seguiment

[R33-LOT2] L'adjudicatari haurà de portar a terme la monitorització i avaluació del rendiment dels models i el seu manteniment durant el període de durada del contracte. S'entén per manteniment, la reestimació, el recalibrat o el reentrenament dels models, aplicant una o altra opció, en funció del resultat de l'avaluació del rendiment.

7.3.8 Casos d'ús: prioritats, abast i àmbits

Un dels objectius que busca l'aplicació de tècniques d'aprenentatge automàtic i analítica de dades en les organitzacions és el d'obtenir informació per millorar la gestió i la seva eficiència, que en darrera instància suposa un estalvi en temps i diners. En l'àmbit de les administracions públiques, l'aplicació del coneixement obtingut com ajuda per la presa de decisions i l'automatització de processos pot ser un accelerador en el procés de transformació digital, en benefici de tota la ciutadania.

7.3.8.1 Priorització dels casos d'ús

Es prioritzaran els casos d'ús que aportin resultats a curt termini, amb valor públic i impacte per les administracions locals, a nivell municipal i supramunicipal. En aquest sentit, es prioritzaran els àmbits on ja es disposi de dades d'interès públic per tots els municipis de Catalunya, com per exemple el datawarehouse de l'AOC i les plataformes de dades obertes.

7.3.8.2 Abast dels casos d'ús

Els casos d'ús que són objecte de la present licitació hauran d'aplicar, amb criteri general, les següents tècniques:

- Estandardització del format de les dades de les administracions locals i definició d'un catàleg de dades amb la finalitat de facilitar la seva consulta, accés i estudi.
- Identificar bones pràctiques i iniciatives innovadores en l'àmbit de les administracions locals
- Descobriments de coneixement (Insights) per millorar la presa de decisions i la planificació
 - Descobriments de relacions no conegudes entre variables
 - Identificació de patrons de comportament
 - Obtenció de models predictius tant univariants com multivariants
 - Detecció d'anomalies i estudi del seu efecte en el comportament esperat de les dades
 - Càlcul de indicadors que permetin agrupar el coneixement contingut a les dades
 - Realitzar estudis de classificació, tant jeràrquics com no jeràrquics, per tal de descobrir agrupacions d'individus i variables segons la semblança en el seu comportament
 - Disseny i entrenament de xarxes neuronals pel desenvolupament de sistemes de generació de llenguatge natural especialitzat en l'àmbit de les administracions locals.
- Desenvolupament de petites solucions que implementin, en forma d'aplicació web o de API, el coneixement obtingut. Per exemple en forma de sistemes de classificació, generació de previsions, de llenguatge natural o automatització de tasques.
- Automatització de processos (si escau)

7.3.8.3 Àmbits dels casos d'ús

El projecte tractarà casos d'ús dels àmbits que siguin d'interès per la millora de la gestió pública local. Sense fer una anàlisi exhaustiva, alguns dels àmbits de potencial interès són:

- **Serveis adreçats a la ciutadania**
 - Serveis digitals
 - Inclusió digital
 - Serveis socials
 - Mobilitat
 - Gestió de l'espai públic
 - Turisme
 - Comerç
 - Medi ambient i clima
 - Gestió residus i selectiva
 - Aigua i energia
 - Seguretat
 - Xarxes socials
 - Atenció ciutadana
- **Gestió interna de l'Administració**
 - Transformació digital interna
 - Pressupost

- Recursos humans
- Planificació de serveis, infraestructures i equipaments
- Tramitació: terminis i eficiència

L'anàlisi de les dades d'aquests àmbits es realitzarà amb visió municipal i també supramunicipal: comarcal, província i Catalunya.

7.3.8.4 Exemple de cas d'ús: Índex de Maduresa Digital

Per posar un exemple, i en la línia de trobar indicadors que facilitin la presa de decisions, el Consorci AOC ja ofereix al conjunt de les Administracions Públiques Catalanes l'Índex de Maduresa Digital (IMD). És un índex compost que avalua el grau de transformació digital de les administracions locals catalanes. Es calcula mitjançant la suma ponderada de més de 30 indicadors organitzats en tres dimensions: drets digitals, activitat dels serveis digitals i govern obert.

Es tracta d'una iniciativa única i innovadora, tant a nivell estatal i internacional. Hi ha una gran quantitat d'índexs que mesuren el desenvolupament del govern digital, com l'Índex d'Economia i Societat Digital (DESI) de la CE, l'Índex de Desenvolupament de Govern Electrònic (EGDI) de la ONU o l'e-Government Benchmark de la CE, però tots estan definits a nivell de país i no es poden aplicar a nivell regional o local. A més a més, es basen en enquestes o dades estadístiques, mentre que l'IMD utilitza dades públiques i objectivables.

El conjunt de dades de l'IMD conté una sèrie històrica de més de 30 indicadors relacionats amb el govern digital de 988 entitats locals catalanes. Les dades de l'IMD estan permanentment actualitzades i disposen d'un ampli nombre de metadades, incloent la descripció de cada un dels indicadors utilitzats.

Un cas d'ús a cobrir durant el contracte seria el de millorar, amb les dades disponibles i emprant tècniques d'anàlisi avançada de dades aquest índex per tal de garantir el seu ajust al que precisament es vol mesurar.

7.3.9 Descripció de mitjans materials.

Les empreses licitadores podran proposar la consulta a fonts de dades que impliquin un cost, sempre que aquestes consultes aportin valor als objectius que estableix el projecte i el seu import estigui dins del pressupost definit a aquesta finalitat.

7.3.10 Descripció de mitjans personals

La relació de perfils de l'equip de treball i les funcions a desenvolupar per cadascun són les següents:

Cap de projecte:

- Dirigir, gestionar i coordinar el contracte, els serveis/projectes inclosos i els equips.
- Planificar i controlar la prestació i el lliurament dels treballs contractats.
- Gestionar els riscos i els plans de mitigació o contingència.
- Gestionar la facturació i el finançament del servei.
- Reunions de coordinació amb els responsables del Consorci AOC servei.
- Establir la visió tècnica global dels equips de desenvolupament.
- Definició del model de governança i definició del catàleg de dades, en col·laboració amb l'empresa adjudicatària del Lot 1 i el Consorci AOC
- Anàlisi funcional.

- Identificació de KPIs a través de les diferents oportunitats que puguin sorgir per als casos d'ús.
- Dimensionat d'usuaris de la plataforma, configurant diferents rols pel mòdul del govern de la dada.
- Garantir la fiabilitat de les dades, realitzant creuaments i integració de dades amb les diferents eines corporatives.

Data Scientist / Consultor de dades:

- Definició global d'arquitectures, processos i models de suport a la gestió del cicle del dada (ingesta, transformació, emmagatzematge, processament i explotació).
- Determinar requisits d'usabilitat i protecció de dades
- Determinar requisits normatius, tècnics i jurídics, a aplicar en tractament de la informació i privadesa.
- Extraure coneixement de les dades per als diferents casos d'ús, tant per a personal tècnic qualificat, com per a usuaris sense coneixements avançats.
- Desenvolupament d'anàlisi avançada, models amb Machine Learning i Deep Learning.
- Crear models de dades, consulta de dades, gestió de dades, neteja de dades i vinculació de dades.
- Disseny de quadres de comandament.
- Configurar mòdul BI i preparar entorns per a anàlisi i reporting.
- Utilitzar les diferents eines per extreure i analitzar les dades generades per la activitat dels usuaris.

Consultor junior / Programador:

- Desenvolupament de APIs i aplicacions web per emprar els resultats i models obtinguts dels anàlisi de dades portats a terme.
- Suport als perfils de Cap de projectes i Data Scientist.

8 Formes de seguiment i control de l'execució de les condicions

L'adjudicatari haurà d'explicar en la seva proposta quin és el model de relació que proposa per garantir l'èxit del projecte.

Com a mínim, però, caldrà que s'estableixin els següents nivells d'interlocució:

Reunions d'estratègia i direcció amb les següents característiques:

- Interlocutors: responsable del servei per part de l'adjudicatari. Cap de Servei del servei i/o Cap de Projecte per part del Consorci AOC i personal directiu de l'AOC si s'escau.
- Periodicitat: trimestral.
- Objectiu: fer el seguiment del contracte, analitzant diversos aspectes: productivitat, control d'hores, temes de facturació, seguiment de fites (a alt nivell), etc.
- Lliurables:
 - Informe resum de les actuacions ja resoltes, en curs i pendents, amb el seu grau de prioritització.
 - Escandall d'hores totals realitzades en el període amb caràcter mensual.

Reunions de seguiment amb les següents característiques:

- Interlocutors: les persones assignades per l'adjudicatari per dur a terme el servei. Per part del Consorci AOC serà el Cap de Servei i/o el Cap de Projecte, o algun dels tècnics assignats al projecte.
- Periodicitat: setmanal.
- Objectiu: seguiment del compliment de l'ANS, rendiment de la plataforma, incidències més destacables, i seguiment dels evolutius i correctius a prioritzar.

La comunicació, gestió i dedicació de les tasques, incidències i propostes de millora es realitzarà mitjançant alguna de les següents eines: JIRA (com a eina de negoci i seguiment), Teams i Zendesk del Consoci AOC.

9 Aspectes de seguretat i ENS

Els sistemes que integrin la Plataforma de dades de les Administracions Locals de Catalunya hauran d'estar integrats amb el landing zone de seguretat de Amazon Web Services del Consorci AOC i estar certificats amb un nivell mig d'acord amb el que estableix l'Esquema Nacional de Seguretat, regulat pel Real Decret 311/2022, de 3 de maig. D'acord amb l'anàlisi de riscos efectuat, la categorització dels serveis objecte d'aquesta licitació és la següent:

			Metodologia ENS					
Actiu essencial	Descripció	Subsistema	D	C	I	A	T	DP
Data Lake	Dades en brut carregades al sistema abans de ser tractades.	AWS	Baix	Mig	Mig	Mig	Mig	NA
Datawarehouse	Dades ja processades amb estructura i format especialment pensat per facilitar el seu anàlisi	AWS	Baix	Mig	Mig	Mig	Mig	NA

C – Confidencialitat I – Integritat T – Traçabilitat A – Autenticitat D - Disponibilitat

DP – Protecció de dades

Tot i que l'adquisició de la infraestructura AWS sobre la que es muntarà la plataforma no està contemplada en aquesta licitació, les empreses adjudicatàries d'ambdós lots hauran de garantir que qualsevol implantació o servei que s'hi munti complirà amb els requisits de seguretat definits a l'Esquema per aquest nivell.

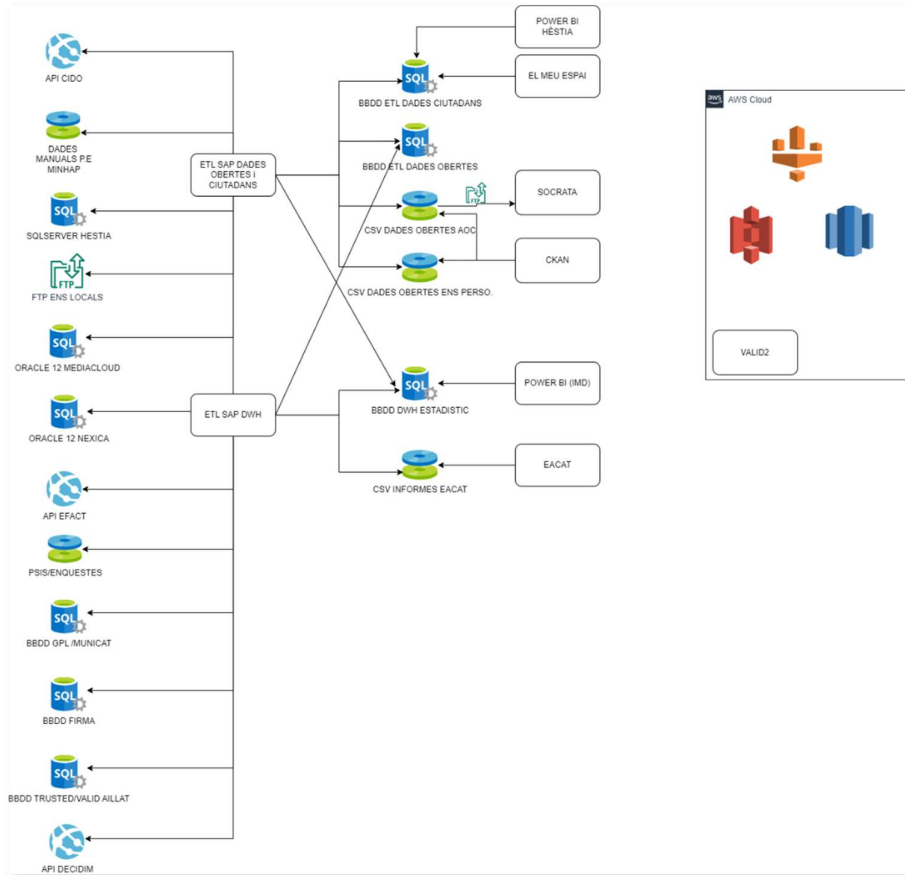
Annex I: Descripció Infraestructura de dades actual de l'AOC

A continuació es descriu la infraestructura de dades actual del Consorci AOC acotada al context de:

- La generació de les dades estadístiques dels serveis gestionats pel Consorci AOC, el seu emmagatzematge al *data warehouse* del Consorci AOC, i la seva publicació via informes públics, per exemple en: <https://www.aoc.cat/indicadors>
- La recopilació de múltiples dades de les administracions catalanes per al seu emmagatzematge i publicació en format de dades obertes a: <https://dadesobertes.seu-e.cat>
- La recopilació de múltiples dades dels ciutadans per posar-se'ls a la seva disposició a "El Meu Espai" : <https://www.aoc.cat/serveis-aoc/el-meu-espai>

Es començarà descrivint les diferents fonts origen de dades, es continuarà amb el tractament que es realitza a aquestes dades mitjançant processos ETL, i per últim es detallaran les bases de dades de destí i les aplicacions de negoci que accedeixen a aquestes.

Un esquema resumit és el següent, on a l' esquerra hi ha les fonts de dades, al mig hi ha els dos ETL que processen les dades, seguits a continuació per les bases de dades destí; i per últim les diferents aplicacions/serveis que hi accedeixen:



9.1 Descripció de les fonts de dades d'origen

9.1.1 Bases de dades gestionades pel consorci AOC

A continuació s'enumeren les bases de dades del consorci AOC, aquestes es troben dividides en quatre centres de dades diferenciades (Mediacloud, Nexica, Azure i AWS). La descripció de cada servei mencionat es pot trobar a www.aoc.cat/serveis.

- Oracle 12i multiservei a Mediacloud: aquesta base de dades inclou les dades de negoci dels següents serveis del consorci: VIA OBERTA, ENOTUM, ETAULER, ETRAM2, OVER, REPRESENTA, VALISA, SIR, MUX2, PADRO, IDCAT MÒBIL, SIRI, SEU ELECTRONICA i TRANSPARÈNCIA.
- SQLServer a Mediacloud: per a la plataforma EACAT
- SQLServer a Mediacloud: per a la plataforma ETRAM1
- Oracle 12i a Mediacloud: pel servei de signador centralitzat (trustedX)
- Oracle 12i a Mediacloud: pel servei VALID
- Oracle 12i a Nexica: pel servei VALID DR
- Oracle 12i de Nexica: que dona servei a ERES i iARXIU
- Mongo DB a Mediacloud: que dona servei a el VALIDADOR
- Mongo DB a Mediacloud: que dona servei al HUB de carpetes ciutadanes
- PostgreSQL a mediacloud: per al servei de dades obertes CKAN
- SQLServer a Azure: per al servei HÈSTIA
- Mysql a Azure: per al servei Decidim
- Mysql a AWS: per al servei VALID2
- Elasticsearch a AWS: per al servei MUX3
- Elasticsearch a AWS: per al servei DESAL2

9.1.2 Bases de dades gestionades per altres administracions

El consorci en la prestació dels seus serveis també accedeix a bases de dades d'altres administracions mitjançant VPN, API keys o SFTP, entre elles:

1. Oracle GAH: en aquesta base de dades es troba informació sobre la gestió i pressupostos GPL dels ens locals, els càrrecs electes, informació institucional dels ens del món local, relacions entre organismes, etc.
 - API JSON de l'organisme CIDO: d'on s'obté informació de diferents AAPP com per exemple: ajuts beques cursos premis subvencions, calendaris i padrons fiscals, cartipàs, convenis col·lectius municipis, estatuts organismes, ofertes públiques de convocatòries personal i els seus resultats, ordenances fiscals, ordenances reguladores i reglaments, plecs de clàusules generals, pressupostos i plantilles, tot això amb l'objectiu de mostrar-les a la plataforma de dades obertes del Consorci AOC.
 - API JSON de la DIBA: per accedir a dades de SAIP i del registre d'activitats.
 - API JSON de la plataforma de dades obertes de la Generalitat (Socrata): s'accedeix entre d'altres a les dades del registre de contractes de tot Catalunya i als de la plataforma de licitació PSCP amb la intenció de mostrar-les com a dades obertes dins la plataforma de dades obertes del Consorci AOC.

- API JSON global de descarrega del servei Decidim: s'obtenen les dades d'activitat de tots els Decidim de Catalunya (encara que no siguin hostatjats pel Consorci AOC).
- Descarrega SFTP/FTP/FTPS d'altres administracions: en la prestació del servei de dades obertes el consorci accedeix a uns 60 d'organismes locals per sincronitzar els seus fitxers de dades obertes amb la plataforma de dades obertes del Consorci AOC, actualment són uns 600 fitxers diferents.

Es pot obtenir la llista completa de tots els organismes i dels seus datasets en el següent enllaç: <https://dadesobertes.seu-e.cat/organization>

- Pujada SFTP altres administracions: en el marc de col·laboració amb la plataforma de dades de la Generalitat el Consorci AOC també puja informació al SFTP de la plataforma Socrata per a la seva publicació en dades obertes. Concretament es pugen diàriament uns 25 fitxers entre ells: càrrecs del món local, estatuts, pressupostos, cartipàs, convocatòries de personal, ordenances fiscals, ordenances reguladores i reglaments, convocatòries de subvencions i ajudes, organismes dependents, pressupostos i liquidacions, convenis funcionaris laborals i sindicals, calendaris i padrons fiscals, liquidacions per principals ingressos tributaris, padró municipal d'habitants per municipi any, endeutament, dades d' e-administració, cost efectiu dels serveis, i objectius d' estabilitat pressupostària, etc.
- Descarrega manual d'altres fitxers: en els casos en què no és possible automatitzar la descàrrega de dades d'altres administracions aquesta es realitza de forma manual periòdicament, com per exemple amb les dades del Ministeri d'administració pública i Hisenda en el cas de: endeutament, termini mig de pagament, convocatòria i concessions de subvencions, etc.

9.1.3 Bases de dades gestionades per proveïdors externs del Consorci AOC

En alguns serveis del consorci AOC les bases de dades es troben a centres de dades contractats directament per proveïdors externs, seria el cas de:

- Base de dades mysql al CPD de Firma professional: utilitzada pel servei de certificació digital tant a nivell de generació de certificats per a Administracions Públiques (T-CAT) com per al certificat idCAT destinat a la ciutadania.
- API d' accés a les factures de les Administracions Públiques catalanes de CERES: és un API que permet descarregar les factures i els estats de les mateixes per a un període de temps determinat.
- Descàrrega manual de dades de Google analytics: per a tots els serveis del consorci AOC, com per exemple les visites als més de 1200 portals de transparència.
- Descàrrega manual de les dades de les estadístiques en survey monkey: amb les enquetes de tots els serveis prestats pel consorci AOC.

9.2 Instàncies ETL del Consorci AOC

El consorci AOC disposa de dues instàncies ETL per a l' extracció, transformació i càrrega de les dades a les bases de dades finals. Les dues instàncies estan basades en SAP data services versió 9.2 que ja no té suport del fabricant, i que per tant cal substituir de forma inclosa en aquesta licitació..

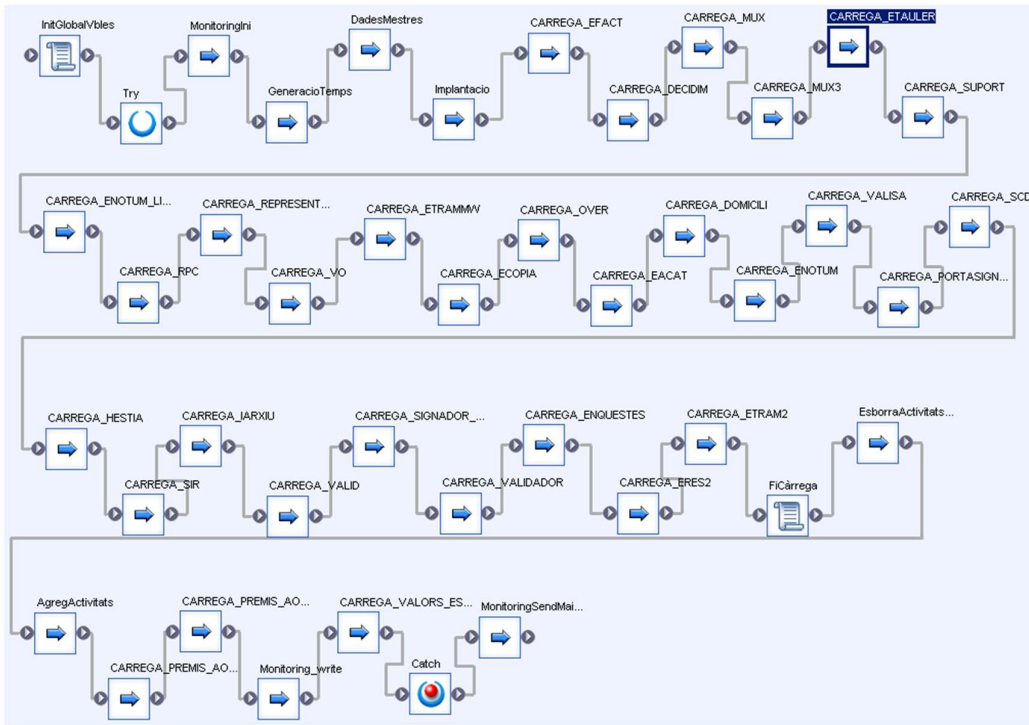
Ambdós ETLs tenen capacitat d'executar fitxers Python per a aquells casos en què l'ETL no tinguin les funcionalitats requerides per al tractament de dades.

Els dos ETL són els següents:

9.2.1 ETL SAP Dataservices per al DW estadístic dels serveis AOC

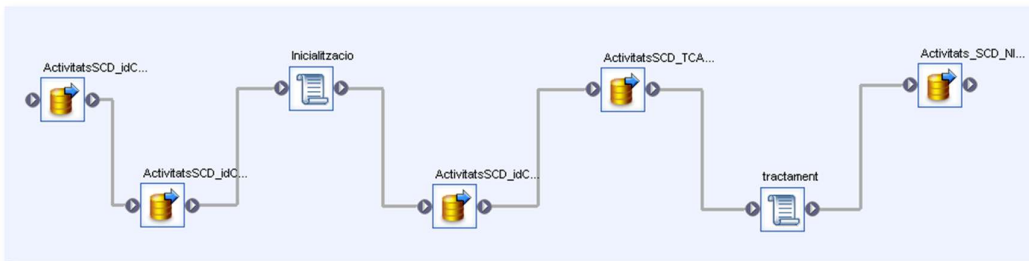
Aquest primer ETL obté dades d'activitat de tots els serveis del consorci a nivell estadístic, ho fa a partir de les fonts de dades detallades anteriorment, i ho desa principalment en la base de dades estadística del data warehouse. De la mateixa forma aquest ETL també ajuda a la realització de l' estudi dels reconeixements AOC, que cada any es lliuren als organismes del món local, per fer això darrer inclou dades addicionals més enllà dels propis serveis del Consorci AOC.

- Exemple flux principal de l' ETL amb tots els subfluxos:

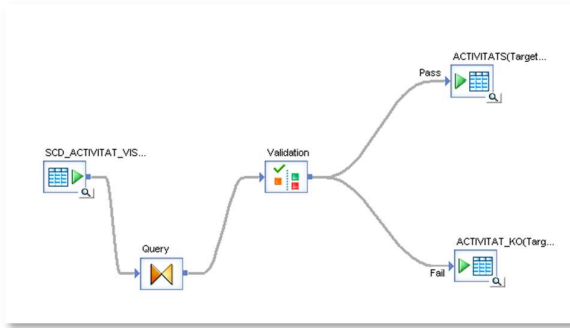


- Exemple subflux de càrrega del servei SCD:

Cada flux de càrrega es sol estructurar en dos nivells, un primer nivell que obté i desa les dades en brut i un segon que agrega i transforma les dades cap a la seva destinació final



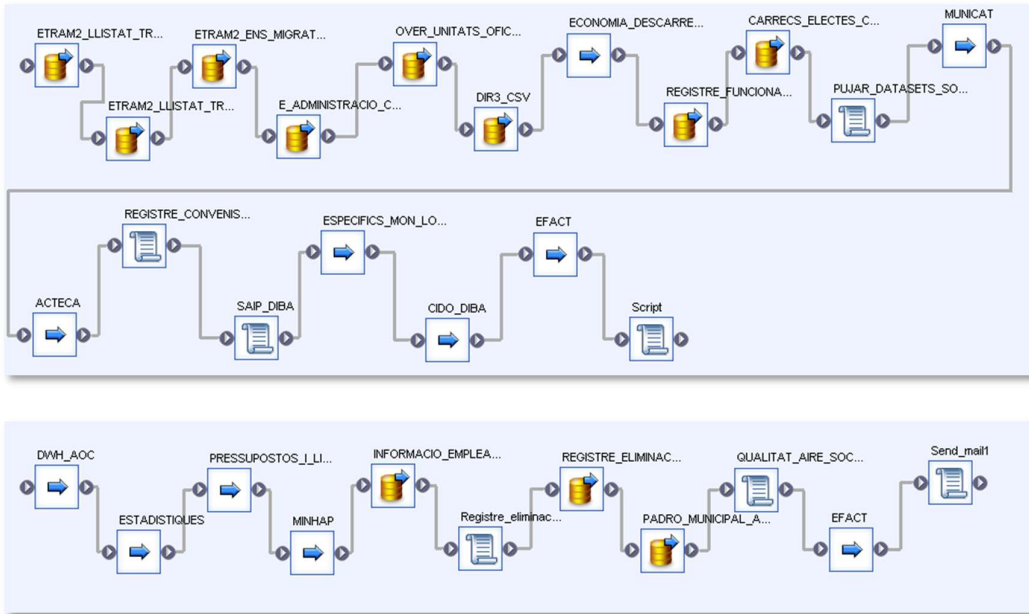
A nivell de data warehouse les dades estadístiques s'acaben guardant principalment a la taula ACTIVITATS, que conté les activitats que realitza cada ens del món local en els serveis del Consorci AOC, juntament amb altres dimensions com la temporalitat d' aquesta activitat.



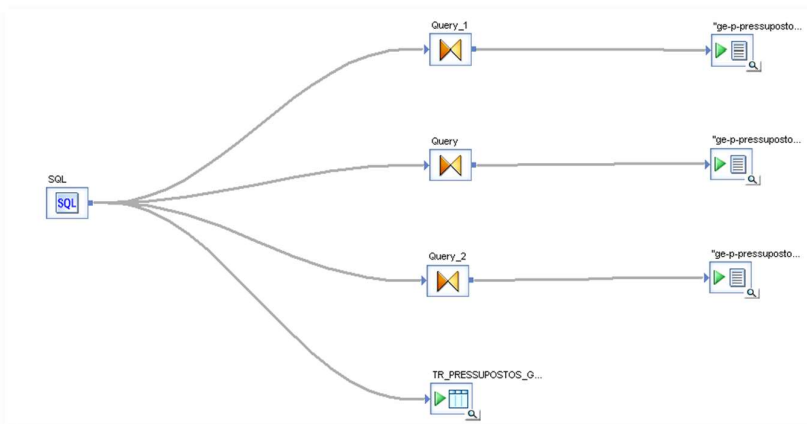
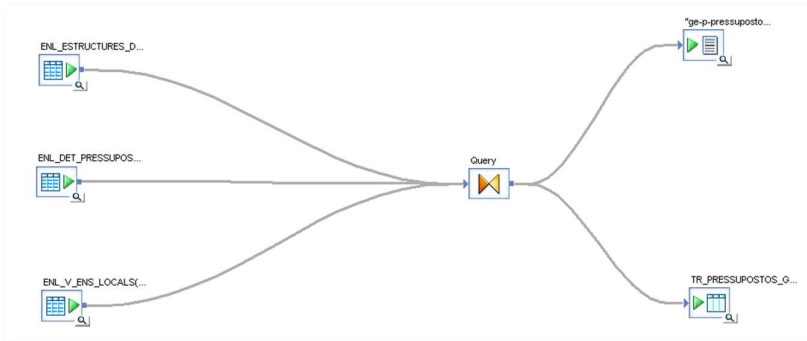
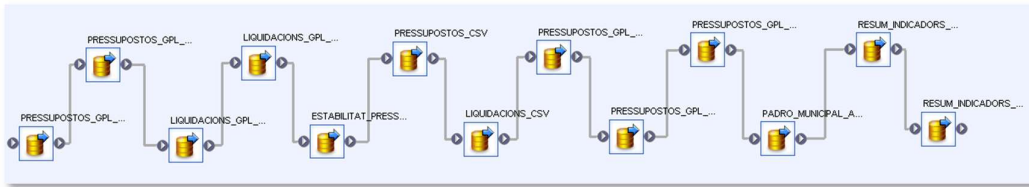
9.2.2 ETL SAP Dataservices per a dades obertes, i ciutadans

Aquest segon ETL obté les dades obertes de les diferents Administracions Públiques que seran publicades als portals de transparència; a més té la responsabilitat d'alimentar part de les dades de "El Meu Espai".

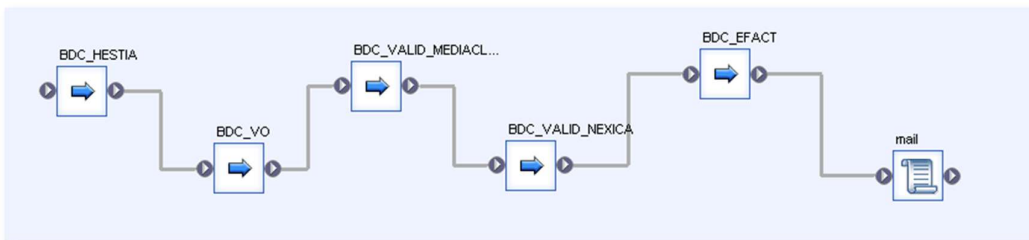
- *Exemple flux principal de càrrega de dades obertes:*



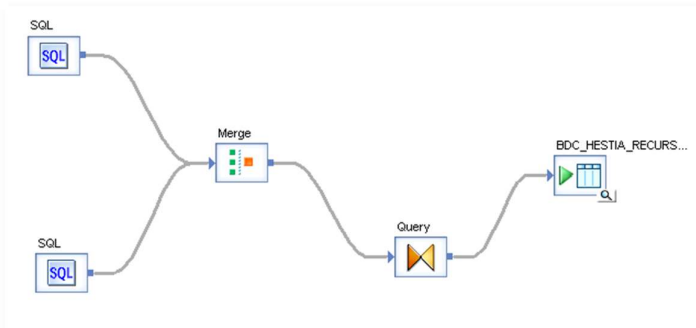
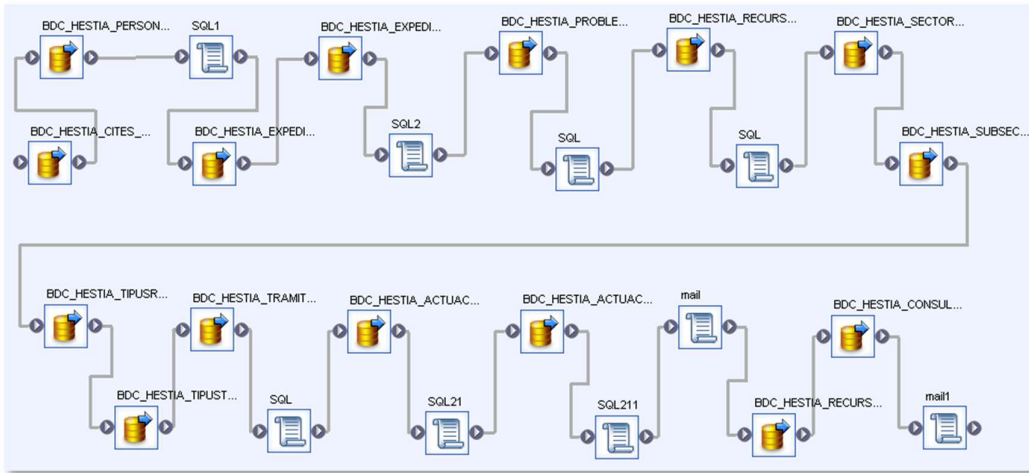
- Exemple subflux de càrrega de dades obertes de pressupostos i liquidacions:



- Exemple flux principal de càrrega de dades per al Meu Espai:



- Exemple de flux de càrrega d' HÈSTIA:



9.3 Bases de dades i repositoris de destí

La destinació de les dades de tots els fluxos de càrrega ETL anteriors són les següents:

- Base de dades de l'ETL dels ciutadans: es tracta d'un servidor SQL Server que dona servei específic a "El Meu Espai" a l'hora de mostrar les seves dades als ciutadans.
- Base de dades SQL Server de l'ETL de dades obertes: es tracta d'un servidor SQL Server que dona servei a la generació de fitxers CSV de dades obertes.

Aquest fitxers CSV de dades oberts es troben en un filesystem de l'ETL de dades obertes, que conté:

- Els fitxers CSV generats per Consorci AOC en la prestació dels seus serveis a les AAPP.
- Els fitxers CSV pujats directament pels organismes del món local. Aquests són sincronitzats per l'ETL de dades obertes cada nit accedint als múltiples FTPs dels ens del món local.
- Base de dades SQL Server de l'ETL del DWH del Consorci AOC: es tracta d'un servidor SQL Server on s' emmagatzema:
 - un bolcat inicial de primer nivell de les dades d'activitat dels diferents serveis del Consorci AOC com s'ha comentat anteriorment, així com les dades necessàries per a l'índex de maduresa digital.
 - el tractament posterior de segon nivell de les dades per a que siguin analitzables fàcilment a nivell estadístic i es puguin mostrar.

Donada la complexitat interna de les aplicacions, aquestes bases de dades s'hauran de mantenir, si més no temporalment, a la nova plataforma de dades del món local, movent aquests SQLServers al cloud corresponent de forma complerta i incloent aquestes tasques en el marc d'aquesta licitació. I després, si s'escau, en una segona fase passant aquestes bases de dades al nou sistema de base de dades que es decideixi.

9.4 Aplicacions que accedeixen a les bases de dades de destí

A continuació es detallen les aplicacions de negoci que accedeixen als repositoris de destí de dades que s'han comentat anteriorment.

Aquestes aplicacions s'hauran de tenir-se en compte en els moviments de les dades, inclosos en aquest plec, cap a la nova plataforma de dades locals del món local, amb l'objectiu que la migració tingui el mínim impacte possible cap als usuaris finals, ja siguin ciutadans o empleats públics.

- Plataforma de dades obertes de l'AOC (CKAN): una de les principals aplicacions que accedeix al repositori de fitxers CSV de l'ETL de dades obertes és el propi servei de dades obertes del Consorci AOC: <http://dadesobertes.seu-e.cat> amb més de 600 datasets provinents de diverses fonts de dades (a nivell de món local, autonòmic i estatal), i que son processats per l'ELT de dades obertes com s'ha comentat anteriorment.

Consorci Administració Oberta de Catalunya | Dades obertes | Consorci Administració Oberta de Catalunya | Contacte

Dades obertes

Organitzacions | Grups | Conjunts de dades | admin

Cercar conjunts de dades:

O filtra per categoria

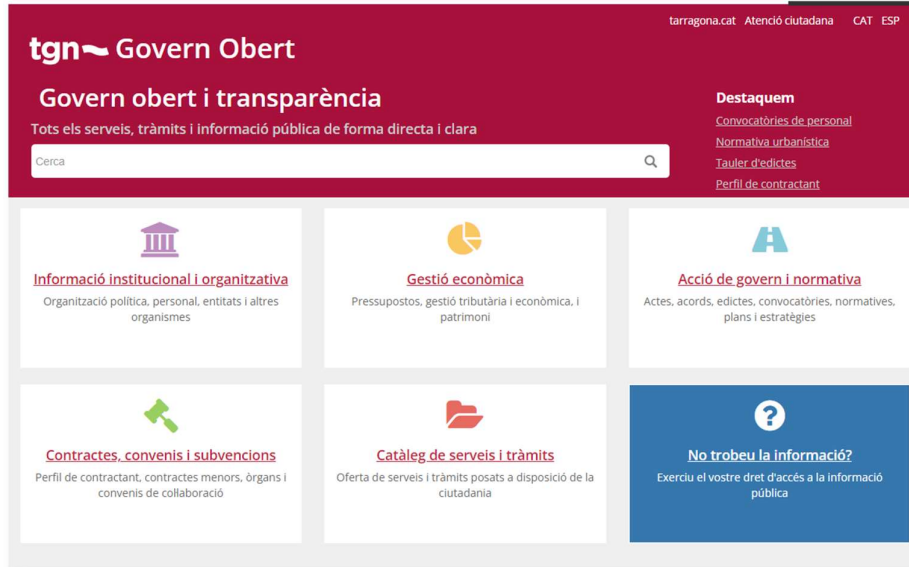
Ciència i Tecnologia (6)	Comerç (31)	Cultura i oci (52)	Demografia (47)	Economia (60)	Educació (10)
Energia (8)	Esport (7)	Habitatge (2)	Hisenda (5)	Indústria (13)	Legislació i justícia (12)
Medi Ambient (24)	Medi Rural i Pesca (1)	Salut (20)	Sector Públic (136)	Seguretat (6)	Societat i benestar (37)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

Cal comentar també que la plataforma de dades obertes del consorci dona també servei a la creació de portals de dades obertes pròpies generades pels propis ens locals, tot accedint al repositori de l'ETL de dades obertes per a mostrar els datasets corresponents, sent ara mateix una seixantena d'organitzacions les que disposen d'aquest tipus de portal:

Ordena per: |

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

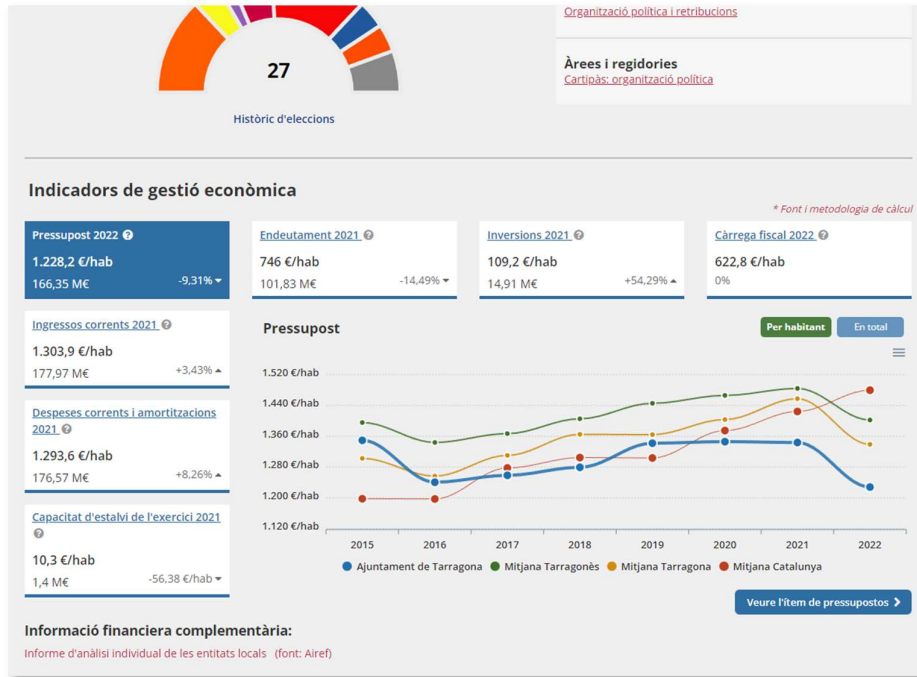
A la seva vegada diferents serveis accedeixen a la plataforma de dades obertes del Consorci AOC per mostrar-ne la seva informació, com per exemple els més de 1200 portals de seu electrònica i/o transparència del món local proveïts també pel Consorci AOC (per exemple www.seu-e.cat/web/tarragona), destacant per aquest motiu la importància de tenir-ho en compte en la migració cap a la nova plataforma de dades del món local.



En aquests portals de transparència també és mostren incrustades les dades obertes de la personalitzades per a cada ens:



També es mostren diferents gràfics alimentats per aquestes dades obertes de la plataforma CKAN, com per exemple els indicadors de gestió econòmica, l'estructuració per programes del pressupost i la liquidació del mateix, etc.



- **El Meu Espai:** permet a la ciutadania accedir a les seves dades en la seva relació amb les administracions públiques catalanes i d'una forma interadministrativa. Es pot consultar la fitxa de "el Meu Espai" a <https://www.aoc.cat/serveis-aoc/el-meu-espai/>, i es pot accedir a l'entorn productiu 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)

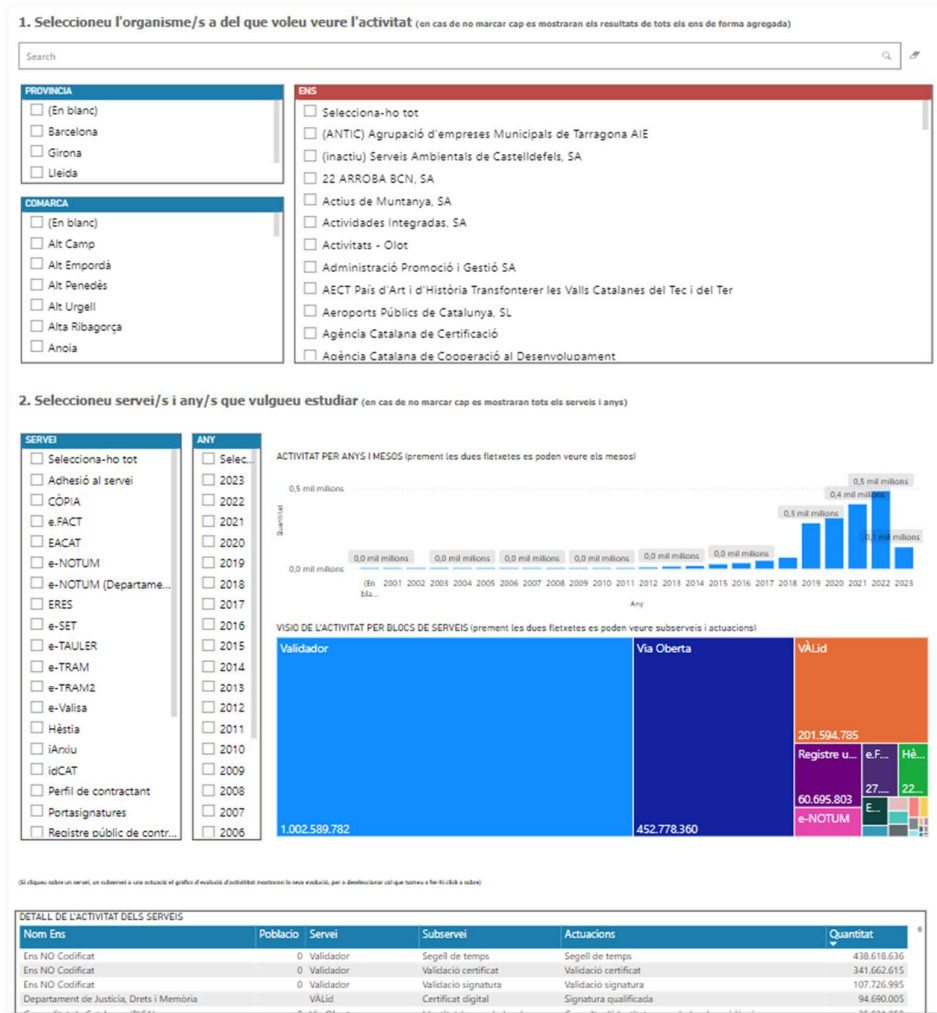
Obté informació sobre els ajuts socials que t'ofereixen les diferents administracions públiques en funció de la teva situació personal i familiar.

[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]

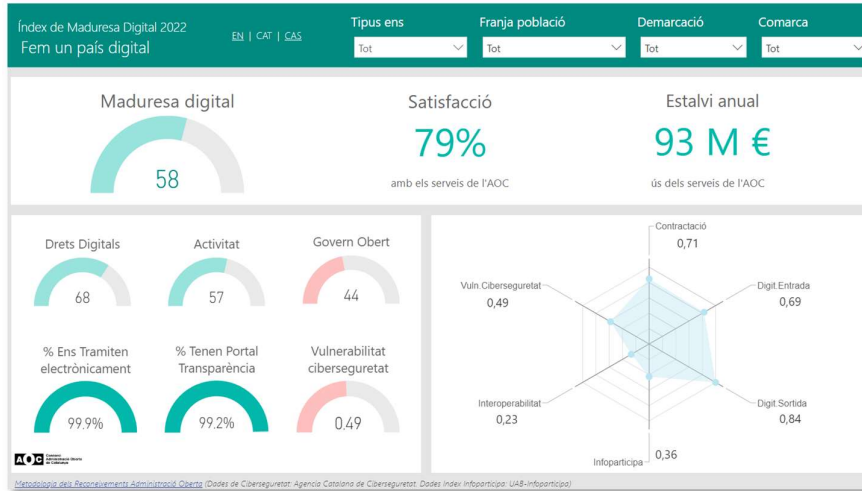
[ACCEDIR AL RECOMANADOR](#)

Ens ajudes a millor

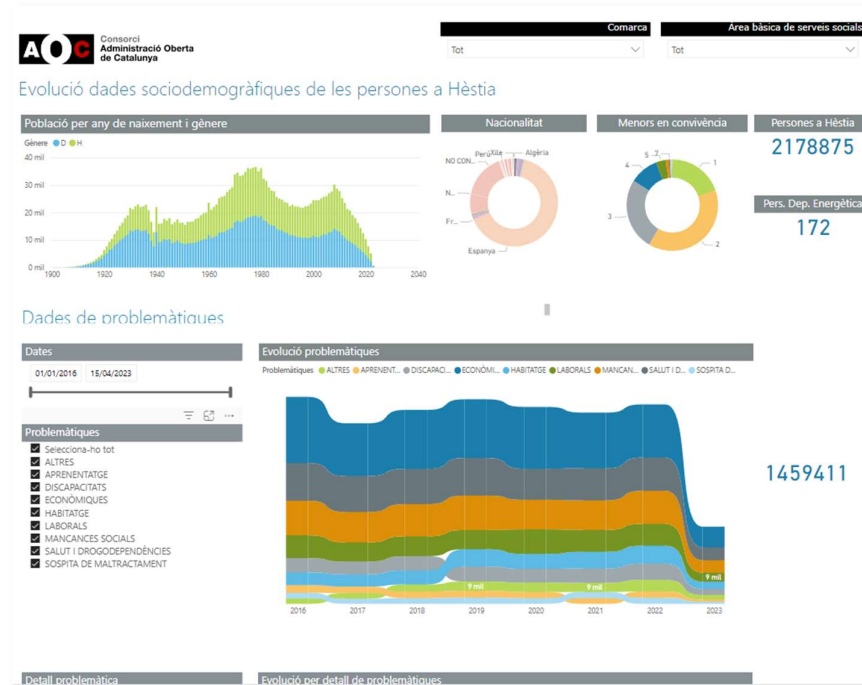
- **Power bi del quadre de comandament i informes dels serveis AOC:** es disposa d'un informe base de quadre de comandament dels serveis del que en deriven més de 200 informes públics, per exemple els que es mostren aquí: <https://www.aoc.cat/indicadors>



- **Power bi de l'index de maduresa digital:** cal destacar també un altre informe crític corresponent a l'index de maduresa digital de les administracions públiques catalanes: <https://www.aoc.cat/blog/2023/revisa-imd-2022/>



- **Power bi de quadres de comandament específics:** que inclouen ratios del servei concret:



- **Plataforma de dades obertes de la Generalitat (Socrata):** és un altra aplicació que accedeix a les dades obertes generades pel consorci AOC es la plataforma Socrata de la Generalitat <https://analisi.transparenciacatalunya.cat/ca/> . En aquest cas concret des del repositori de l'ETL de dades obertes es pugen a Socrata via SFTP uns 20 datasets amb dades del món local de forma diària.

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: és l'extranet de les administracions públiques catalanes (<https://www.aoc.cat/serveis-aoc/eacat-tramits/>), i també accedeix al repositori de l'ETL del Data warehouse per a la generació d'informes CSV d'activitat dels serveis Via Oberta i Padró, per als diferents ens que en sol·licitin la descàrrega.

Annex II: Dades disponibles del món local actualment

Les principals fonts de dades del món local són:

- *Data warehouse* de l'AOC
 - o La informació d'activitat dels serveis del Consorci AOC es pot consultar, de forma agregada, en el següent informe: [Indicadors – Consorci AOC](#)
- Plataforma compartida de dades obertes de l'AOC, amb conjunts de dades globals (publicats i gestionats pel Consorci AOC) i específics d'ens locals:
 - o [Dades Obertes – Consorci AOC](#)
- Plataforma de dades obertes de la Generalitat de Catalunya:
 - o [Dades Obertes de Catalunya](#)
- Plataforma de dades obertes de l'Administració General de l'Estat
 - o [Datos Abiertos del Gobierno de España](#)
- Plataformes de dades obertes d'altres administracions locals amb plataforma pròpia (Ajuntament de Barcelona, Diputació de Barcelona, etc).
 - o [Open Data BCN](#)
 - o [Portal de Dades Obertes de la Diputació de Barcelona](#)