Ajuntament de Girona

LEC/ag/so/mas
Exp. 2026012968 – C. 01.001.003

PLEC DE PRESCRIPCIONS TÈCNIQUES REGULADORES DE LA CONTRACTACIÓ DEL SERVEI DE TRANSCRIPCIÓ AUTOMATITZADA I INDEXACIÓ PROBABILÍSITCA DELS MANUALS D'ACORDS PEL PROCEDIMENT OBERT SIMPLIFICAT ABREUJAT

Descripció del contracte	Transcripció automatitzada i indexació probabilística dels manuals d'acords custodiats a l'Arxiu Municipal de Girona
Pressupost base de licitació	71.987,61 €, IVA inclòs, calculat al tipus del 21%
Tipificació contracte de	SERVEIS
Codi CPV	72300000-8, serveis relacionats amb dades
Àrea/servei promotor:	Servei de Gestió Documental, Arxius i Publicacions

1. Introducció / Antecedents

La preservació de la documentació, juntament amb l'accés i la difusió, formen els pilars fonamentals de la missió de l'Arxiu Municipal de Girona. En aquest sentit, la digitalització dels fons documentals esdevé un procediment clau per evitar la manipulació innecessària de la documentació i, especialment, per facilitar la seva divulgació universal mitjançant el web.

Els processos de digitalització de documentació de l'Arxiu Municipal de Girona es van iniciar amb la microfilmació dels conjunts documentals més consultats per la ciutadania: la premsa editada a Girona i, seguidament, els llibres d'Actes del Ple, els Manuals d'acords i els padrons d'habitants. L'evolució tecnològica esdevinguda des de l'any 2000 va permetre millorar els processos i, especialment, la qualitat de les imatges reproduïdes i la seva consulta i reutilització a través del web. L'any 2006 es posava en funcionament un recurs específic per a consultar la premsa digitalitzada a la web del Servei de Gestió Documental, Arxius i Publicacions (prop de més d'un milió de pàgines de la primera vintena de capçaleres digitalitzades): a través de la tecnologia d'OCR (Optical Character Recognition), desenvolupada en aquells moments només per a caràcters mecanoscrits i impresos, que permetia i permet fer cerques per text lliure sobre les pàgines digitalitzades. Aquesta tecnologia va permetre també consultar de forma paral·lela la digitalització i els catàlegs i transcripcions dels documents més antics custodiats a l'AMGi, els *Documents singulars* (Llibres de privilegis, Lletres Reials, Llibre del Sindicat Remença i Pergamins).

Actualment però les innovacions produïdes a través de la Intel·ligència Artificial ja permeten la transcripció automatitzada de textos manuscrits a través de tècniques HTR (Handwritten Text Recognition). Aquestes tecnologies han de facilitar la lectura i, per tant, l'accés a la documentació manuscrita, sovint la més antiga i de difícil comprensió, tant per la diversitat de les escriptures com de les llengües utilitzades. És per aquest motiu doncs, que el Servei de Gestió Documental Arxius i Publicacions impulsa un projecte de transcripció i indexació centrat en el tractament dels llibres manuscrits dels manuals d'acords. Aquest projecte s'inicià el 2024 amb el tractament dels llibres d'Actes del Ple (de 1719 a 1981) i de la Junta de Govern Local (de 1924 a 1981) de l'Ajuntament de Girona. Enguany es vol continuar amb el tractament dels volums manuscrits de Manuals d'acords, precedents de les Actes del Ple, documents anteriors a l'aplicació del Decret de Nova Planta i que, per aquest motiu, presenten característiques pròpies a nivell de llengua i tipus d'escriptura utilitzada, així com de les tipologies documentals contingudes en els llibres.

2. Objecte del contracte

Transcripció automatitzada i indexació probabilística, mitjançant intel·ligència artificial, de 154.891 objectes digitals corresponents a 321 volums manuscrits de la sèrie Manuals d'acords, datats entre 1346 i 1718.

Signatures:
1. Cap del Servei de Gestió Documental, Arxius i Publicacions de l'AJUNTAMENT DE GIRONA - 01/04/2026 18:03

Descripció de l'objecte

L'objecte del present contracte se centra en el tractament dels Manuals d'acords, corresponents al període comprès entre 1346 i 1718. Aquest interval cronològic constitueix la totalitat dels Manuals d'acords del Consell de la ciutat.

Des de creació del govern municipal, arran del privilegi (anomenat *Privilegi d'en Provençal*) concedit pel rei Pere II a la ciutat de Girona l'any 1284, i fins l'any 1719, l'exercici d'aquest requeia sobre un òrgan col·legiat format per 6 (o 4, segons l'època) jurats escollits entre els diferents estaments de la ciutat. Aquests, a més, recolzaven la seva acció amb els acords presos pel Consell de la ciutat, format per 80 (o 60, segons l'època) consellers.

Els Manuals d'acords del Consell de la ciutat són el precedent dels llibres d'Actes del Ple i contenen els acords presos pels jurats i Consell de la ciutat. Consten de 321 volums i presenten les següents característiques:

- Documents manuscrits en escriptura gòtica catalana, fins a mitjan segle XV aproximadament, o en humanística cursiva, a partir de mitjan segle XV. A partir del final del segle XVII alguns volums contenen documents impresos cosits.
- Les llengües utilitzades fins al segle XVI són el català i el llatí. A partir del segle XVI comença a utilitzar-se també el castellà, especialment en els documents generats per l'administració reial. En els períodes en què la ciutat estigué sota domini francès (1643-1646 i 1694-1695), també hi ha documentació en francès inserida en els manuals.
- La majoria de volums corresponen cadascun d'ells a un sol any, especialment a partir de 1380. La sèrie és bastant completa i només presenta salts en el període inicial, entre 1346 i 1380, període en què s'està organitzant l'administració municipal.
- Presenten diversitat de tipologies documentals i d'escriptura perquè, a més de les actes del Consell, s'hi copiaven o cosien els documents tractats en les sessions, les actes de les reunions de les comissions o els documents de presa de possessió de càrrecs, entre altres. Entre la documentació cosida destaquen les lletres reials que rebia la ciutat de part del rei o dels lloctinents reials. A més, s'hi registraven altres documents produïts pels jurats de la ciutat i signats pel notari de la ciutat: crides públiques, establiments, arrendaments o correspondència emesa, entre altres.
- Els volums són en paper amb cobertes de pergamí, en alguns casos documents reutilitzats, presenten un bon estat de conservació i els que són objecte del present contracte han estat restaurats.

3. Especificacions tècniques

Caldrà aplicar les especificacions tècniques de caràcter diplomàtic per entrenar l'eina de reconeixement de textos manuscrits (HTR) i poder establir un model a aplicar als conjunt de document imatge.

L'inici de cada fase requerirà la validació per part del Servei de Gestió Documental i Publicacions de l'informe de la fase anterior.

Les fases del procés seran les següents:

A. Entrega de documentació en format digital

A.1. Canal de traspàs: els 154.891 fitxers TIFF i JPEG a tractar, amb un pes de 1'8 TB, caldrà fer-los arribar a l'empresa adjudicatària a través d'un mitjà segur. Per agilitzar el procés, si es creu convenient es pot realitzar el traspàs per correu postal mitjançant un dispositiu d'emmagatzematge.

A.2. **Informe de verificació de la quantitat d'objectes digitals transferits, amb els corresponents formats**: tot seguit a la recepció dels materials l'empresa adjudicatària realitzarà un informe detallat dels fitxers rebuts sobre la quantitat, el volum i format. Fins que aquesta verificació no s'hagi confirmat per l'Arxiu Municipal i validat per ambdues parts no es podrà continuar el projecte. L'empresa adjudicatària facilitarà aquest informe el més ràpid possible per a identificar possibles errades.

Signatures:
1. Cap del Servei de Gestió Documental, Arxius i Publicacions de l'AJUNTAMENT DE GIRONA - 01/04/2026 18:03

B. Procés d'anàlisi de la documentació

B.1. Identificació de les zones del document amb informació concreta: l'empresa adjudicatària realitzarà una anàlisi diplomàtica de la documentació que l'Ajuntament de Girona haurà de verificar. En aquest punt es podrà demanar mútuament més informació per a consensuar conceptes i identificar regles a aplicar, que es posaran per escrit.

B.2. Identificació de l'idioma: s'avaluarà si l'idioma pot ser un factor que distorsioni el resultat i en cas afirmatiu es realitzarà la transcripció pura sense cap correcció de l'idioma no previst.

B.3. Anàlisi de la lletra al llarg de tots els documents.

B.4. Anàlisi de l'estat de la documentació.

B.5. Anàlisi de les còpies en digital, avaluació de la qualitat de les còpies en digital.

B.6. **Informe final de fase**, que inclourà l'avaluació individual de cadascun dels punts d'aquest apartat, indicant-ne els punts febles, i el cronograma exhaustiu dels processos a executar en les fases següents.

C. Procés d'entrenament

C.1. Per a poder fer la indexació automàtica i la cerca textual sobre els objectes digitals, l'empresa adjudicatària implementarà un sistema de models de Machine Learning / Deep Learning per a reconèixer text manuscrit i poder cercar dins d'aquest. El procés implicarà les següents fases:

C.1.1. Selecció d'imatges representatives: l'empresa adjudicatària **conjuntament amb la persona designada del Servei de Gestió Documental, Arxiu i Publicacions** determinaran els objectes digital més representatius per a l'entrenament del sistema. Les mostres hauran de ser suficient per a cobrir tot l'espectre de possibilitats i peculiaritats de la documentació a tractar.

C.1.2. **Comunicació dels criteris de selecció i justificació de les imatges representatives**. Aquesta comunicació serà indispensable per a la continuació de la fase d'entrenament i contindrà un gràfic per visualitzar la cronologia de les imatges seleccionades. També s'ajuntarà un fitxer de full de càlcul indicant cada imatge amb les següents dades:

- Nom de la imatge
- Registre General del volum
- Any del volum
- Pàgina
- Criteris de selecció, amb columnes separades en el cas de més d'un criteri.

C.1.3. Preparació de la col·lecció per l'entrenament. L'empresa adjudicatària tindrà una persona especialista en el tractament de documentació de l'època per a fer l'entrenament, que inclourà la descripció de l'estructura de les pàgines, la transcripció de cada paraula, les abreviatures incloses, el marcatge dels blocs i les línies de text de cada imatge.

C.1.4. Entrenament de models i avaluació de qualitat dels resultats. L'empresa adjudicatària realitzarà entre tres i quatre entregues de resultats en un entorn de demostració on es puguin avaluar els resultats, segons l'entrenament dels models. **Per a cada entrega es comunicarà** quantes imatges més es necessiten per al següent entrenament, optimitzant els resultats, i el percentatge de millora esperat, i també el guany de l'anterior respecte el resultat l'esperat.

C.2. **Informe final de fase**: inclourà un resum de la definició dels criteris de selecció i la relació d'imatges seleccionades en el conjunt d'iteracions amb la justificació de cadascuna d'elles. També el resum dels balanços de les diferents iteracions que ja s'han anat comunicant, amb les mètriques d'avaluació (p. ex. *Character Error Rate* o *Word Error Rate*), així com la descripció dels mètodes de validació aplicats.

D. Indexació probabilística

D.1. Com a resultat del procés s'obtindrà la informació relativa als possibles continguts per a cada pàgina de la transcripció probabilística i la informació d'ubicació dins cada pàgina.

D.2. **Informe final de projecte**: contindrà com a mínim l'avaluació del model, la descripció d'aquest model obtingut per aplicar a altra documentació semblant, format dels fitxers entregats, així com la definició i especificacions concretes per a cada element, i descripció dels mètodes de validació aplicats. Aquest informe serà revisat per personal designat pel Servei de Gestió Documental, Arxius i Publicacions, per a donar com a vàlids els resultats.

D.3. **Resultats:** s'entregaran en formats oberts, estandarditzats i documentats. Cada fitxer ha d'incloure, com a mínim:

- La transcripció textual de cada paraula, línia o fragment del document.
- Les metadades associades (pàgina, línia, coordenades dins la imatge o zona del text).
- Els valors de confiança o precisió associats a cada reconeixement.
- La identificació de la imatge d'origen.
- La identificació de la sèrie documental a la qual pertany.

Els fitxers finals s'organitzaran en carpetes per sèrie documental i Registre General.

El lliurament final es considerarà complet quan l'Ajuntament de Girona disposi de tota la informació assenyalada en aquest apartat.

Requisits per a l'ús de tècniques d'intel·ligència artificial

Respecte al contingut a transcriure, no s'autoritza a utilitzar aquest contingut per entrenar altres sistemes d'altres clients. Tampoc es podrà partir d'un model pre-entrenat amb dades d'altres clients.

4. Perfils professionals

Per a la prestació del servei es considera imprescindible que l'empresa adjudicatària disposi de recursos humans adequats, tant en nombre com en qualificació tècnica. Mitjançant la firma del contracte, l'adjudicatari es compromet a garantir els mitjans personals necessaris per a la correcta execució dels serveis.

Atès el caràcter tècnic i especialitzat del contracte, i la complexitat de la transcripció automatitzada de documentació manuscrita històrica, els perfils professionals han de comptar amb experiència acreditada en projectes similars. Aquesta experiència és essencial per assegurar l'execució adequada de les tasques, atenent la varietat tipogràfica, lingüística i cronològica dels documents i l'ús d'eines d'intel·ligència artificial.

Es consideren adequats els perfils que s'especifiquen a continuació, deixant no obstant a l'empresa la determinació del personal que posi a disposició del contracte. S'estima que seran necessaris, com a mínim, els següents perfils per executar aquest servei:

- 1 cap de projecte.

  Tasques: planificació, logística i coordinació per part de l'empresa, tramitació, generació d'informes i documentació, direcció tècnica i supervisió de les feines.

  L'experiència prèvia en projectes similars és imprescindible per garantir una gestió eficient del projecte, el compliment dels terminis i la correcta gestió dels recursos humans i tècnics.

- 1 programador informàtic amb experiència en tecnologia d'indexació automàtica de text manuscrit.

  Tasques: programació informàtica.

  Els documents històrics presenten variabilitat grafològica i lingüística que requereix models adaptats. L'experiència assegura l'aplicació de metodologies adequades i la minimització d'errors en la transcripció automàtica.

- 1 especialista en arxivística i paleografia amb experiència en lectura d'escriptura gòtica catalana, dels segles XIV i XV, i escriptura humanística dels segles XV al XVIII, en català i llatí.

  Tasques: entrenament del sistema de verificació de resultats d'indexació i de verificació i correcció de les transcripcions utilitzades en la indexació automàtica dels documents mitjançant un mostreig sistemàtic.

Aquest perfil és clau per garantir la coherència i precisió del resultat final. La seva experiència permet interpretar correctament abreviatures, formes antigues i errors manuscrits, validar les transcripcions generades pel sistema d'intel·ligència artificial i establir els criteris de normalització necessaris per a una indexació uniforme de tota la sèrie documental.

No obstant, l'empresa adjudicatària establirà el nombre de persones que consideri adequat per a la realització d'aquest servei amb els perfils exposats o similars.

El cap de projecte serà l'interlocutor amb el personal encarregat del Servei de Gestió Documental, Arxius i Publicacions, per a totes les tasques de planificació, direcció i seguiment de les actuacions contemplades en aquest Plec.

Correspon a l'adjudicatari l'execució, la direcció i la coordinació tècnica dels mitjans personals que realitzin les actuacions objecte del contracte.

5. Mitjans materials

Els treballs objecte d'aquest contracte es realitzaran a les instal·lacions de l'empresa adjudicatària. Aquesta posarà a disposició de l'equip de treball tots els mitjans materials per a realitzar les tasques objecte del contracte: l'equipament informàtic i infraestructura tecnològica i de processament d'imatges, el material d'oficina, així com els elements ofimàtics i de comunicacions necessaris.

Signatures:
1. Cap del Servei de Gestió Documental, Arxius i Publicacions de l'AJUNTAMENT DE GIRONA - 01/04/2026 18:03