

Cahier des charges

DataQuest

Outil fondé sur des technologies d'IA permettant la
génération automatisée de l'inventaire des collections de
la Chambre des Députés

Table des matières

1. Informations sur le marché	9
1.1 Nom du marché	9
1.2 Pouvoir adjudicateur	9
1.3 Personne de contact	9
1.4 Type de secteur d'activité	9
1.5 Type de procédure	9
1.6 Date d'envoi du cahier des charges	9
1.7 Date limite de réception des offres	9
1.8 Demandes de renseignements ou de clarifications	10
1.9 Constatations d'ambiguïtés, d'erreurs ou d'omissions	10
1.10 Propriété du cahier des charges	10
1.11 Acceptation du cahier des charges	10
1.12 Indemnité pour l'élaboration de l'offre	10
1.13 Sous-traitance	10
1.14 Offre collective	11
1.15 Lots	11
1.16 Variantes	11
1.17 Durée du contrat	12
1.18 Prix	12
2. Objet du marché	12
2.1 Équipe en charge du projet	12
2.2 Type de solution recherchée	13
2.3 Besoin et finalités	13
2.3.1 Finalités du projet	13
2.3.2 Objectifs de réussite	13
2.4 Périmètre général du marché	14
2.4.1 Nature des prestations attendues	14
2.4.2 Vision de la solution cible	14
2.4.3 Évolutivité de la solution	14

3.	Contexte, existant, données et périmètres documentaires	15
3.1	Contexte institutionnel et projet.....	15
3.1.1	Présentation de la CHD	15
3.1.2	Contexte archivistique et documentaire	15
3.1.3	Présentation du projet DataQuest.....	15
3.1.4	Cas d'usage <i>Thematic DataQuest</i>	16
3.1.5	Cas d'usage <i>Sensitive DataQuest</i>	17
3.1.6	Projets connexes et dépendances	26
3.2	Processus et environnement existants	27
3.2.1	Processus actuel de traitement des fonds.....	27
3.2.2	Écosystème institutionnel et technique	28
3.2.3	Orientations relatives aux modèles d'IA	29
3.2.4	Type de solution recherché.....	30
3.3	Données et fonds documentaires	30
3.3.1	Sources de données et organisation documentaire	30
3.3.2	Typologies et formats de documents	30
3.3.3	Langues des documents et métadonnées.....	31
3.3.4	Volumétrie et caractéristiques des fonds	31
3.3.5	Qualité et préparation des données	31
3.3.6	Données sensibles	31
3.4	Périmètres du projet.....	32
3.4.1	Périmètres documentaires successifs.....	32
3.4.2	Périmètre fonctionnel de la V1	32
3.4.3	Éléments hors périmètre.....	32
3.4.4	Hypothèses, inconnues et prérequis	33
3.4.5	Extension et mutualisation	33
4.	Processus cible et exigences fonctionnelles.....	34
4.1	Principes du processus cible.....	34
4.1.1	Vision de bout en bout	34
4.1.2	Cadre de description des étapes	34
4.1.3	Constitution d'un projet de traitement.....	35
4.1.4	Réception et préparation des données	35
4.2	Analyse et enrichissement par l'IA	37

4.2.1	Analyse des documents textuels.....	37
4.2.2	Analyse des images et documents numérisés	37
4.2.3	Génération et enrichissement des métadonnées	37
4.2.4	Classification et description thématique	38
4.2.5	Détection des données sensibles	38
4.2.6	Consolidation au niveau du dossier	39
4.3	Validation humaine et restitution	39
4.3.1	Human in the loop.....	39
4.3.2	Boucle d'amélioration contrôlée et jeux de tests	39
4.3.3	Seuils de confiance.....	40
4.3.4	Gestion des anomalies et exceptions.....	41
4.3.5	Restitution dans l'inventaire Excel	41
4.3.6	Rapports de traitement	41
4.3.7	Traçabilité et explicabilité.....	42
4.4	Configuration et administration de la solution	42
4.4.1	Orchestrer les cas d'usages métiers de la CHD.....	42
4.4.2	Administration fonctionnelle	42
4.4.3	Gestion multi-LLM.....	43
5.	Architecture, sécurité et exigences non fonctionnelles	43
5.1	Architecture et infrastructure	43
5.1.1	Principes d'architecture.....	43
5.1.2	Architecture fonctionnelle cible	44
5.1.3	Scénarios d'architecture	44
5.1.4	Hébergement et souveraineté.....	45
5.1.5	Infrastructure et environnements	45
5.1.6	Intégration au SI	46
5.1.7	Fonctionnement filesystem et interopérabilité	46
5.2	Architecture et gouvernance des modèles d'IA.....	46
5.2.1	Architecture multi-LLM.....	46
5.2.2	Utilisation du service Mistral.....	46
5.2.3	Gouvernance technique des modèles.....	47
5.2.4	Maîtrise des appels aux modèles.....	47
5.3	Protection des données et contrôle des accès.....	47

5.3.1	Cycle de vie des données	47
5.3.2	Confidentialité et non-réutilisation	47
5.3.3	Identités et habilitations	48
5.3.4	Chiffrement et gestion des clés	48
5.3.5	Journalisation et audit	48
5.4	Cybersécurité et conformité	48
5.4.1	Sécurisation de la plateforme	48
5.4.2	Sécurité de l'IA générative	49
5.4.3	Conformité réglementaire	49
5.4.4	Développement et DevSecOps	50
5.5	Exploitabilité et qualité de service	50
5.5.1	Performance et capacité	50
5.5.2	Disponibilité, résilience et continuité	50
5.5.3	Sauvegarde et restauration	51
5.5.4	Supervision et observabilité	51
5.5.5	Ergonomie, langues et accessibilité	51
5.5.6	Sobriété environnementale	51
5.6	Pérennité de la solution	52
5.6.1	Licences et maîtrise des coûts	52
5.6.2	Portabilité et réversibilité	52
5.6.3	Documentation technique	52
6.	BUILD : démarche, gouvernance, livrables et planning	53
6.1	Stratégie de réalisation	53
6.1.1	Démarche incrémentale	53
6.1.2	Organisation Générale du BUILD	53
6.1.3	Principes de progression	53
6.2	Initialisation et mise en place du socle	54
6.2.1	Cadrage technique et fonctionnel	54
6.2.2	Finalisation de l'architecture	54
6.2.3	Acquisition des licences et installation	54
6.2.4	Mise en place du socle industriel	54
6.3	Gouvernance du projet	55
6.3.1	Instances de gouvernance	55

6.3.2	Répartition des responsabilités	55
6.3.3	Gestion des risques et dépendances	55
6.3.4	Gestion des évolutions et arbitrages	55
6.3.5	Gestion documentaire et qualité projet.....	55
6.4	Cycle de réalisation d'un incrément	55
6.4.1	Analyse du besoin et qualification du corpus	56
6.4.2	Spécifications et cahier de paramétrage	56
6.4.3	Paramétrage et développements	57
6.4.4	Tests internes du prestataire.....	57
6.4.5	Évaluation de l'IA.....	57
6.4.6	Niveaux de qualité et critères de recevabilité des traitements IA.....	57
6.4.7	Recette fonctionnelle, technique et SSI	60
6.4.8	Admission et décision de mise en production	60
6.4.9	Mise en production et déclenchement de l'hypercare	60
6.5	Trajectoire et planning des incréments	60
6.5.1	Axes de complexification	60
6.5.2	Premier périmètre et quick-win	61
6.5.3	Incréments intermédiaires.....	61
6.5.4	Stabilisation et généralisation	61
6.5.5	Planning prévisionnel.....	61
6.6	Livrables, compétences et préparation de l'exploitation.....	62
6.6.1	Livrable par phase et incrément.....	62
6.6.2	Formation et transfert de compétences.....	62
6.6.3	Accompagnement au changement	63
6.6.4	Clôture du BUILD et transfert à l'exploitation	63
7.	Transition vers l'exploitation, hypercare, autonomie et réversibilité	63
7.1	Principes et périmètre	63
7.1.1	Périmètre de l'après-BUILD	63
7.1.2	Hypercare après chaque mise en production.....	63
7.1.3	Obligations pendant l'hypercare.....	64
7.2	Assistance et niveaux de service pendant l'hypercare.....	64
7.2.1	Canaux, disponibilité et escalade.....	64
7.2.2	Gestion des incidents et problèmes pendant l'hypercare	64

7.2.3	Mesure des niveaux de service.....	64
7.3	Limites contractuelles de l'hypercare	65
7.3.1	Prestations incluses pendant l'hypercare	65
7.3.2	Prestations exclues du marché	65
7.3.3	Clôture de chaque hypercare.....	65
7.4	Autonomie de la CHD après le projet.....	65
7.4.1	Administration des modèles, prompts, règles et workflows	65
7.4.2	Suivi autonome de la qualité de l'IA	66
7.4.3	Suivi des consommations et coûts.....	66
7.4.4	Supervision technique et fonctionnelle	66
7.5	Sécurité, continuité et réversibilité	66
7.5.1	Responsabilités après hypercare	66
7.5.2	Incidents de sécurité pendant le BUILD et l'hypercare	66
7.5.3	Sauvegarde et restauration.....	66
7.5.4	Continuité et reprise d'activité	67
7.5.5	Réversibilité et fin du marché.....	67
7.6	Documentation, transfert et clôture	67
7.6.1	Documentation et base de connaissances transférées	67
7.6.2	Formation et validation de l'autonomie.....	67
7.6.3	Bilan de clôture	67
8.	Déroulement de la procédure d'attribution du marché	68
8.1	Dépôt des offres.....	68
8.2	Examen de l'offre	68
8.3	Délai de validité de l'offre	68
8.4	Procédure d'attribution et contrat	68
8.5	Calendrier prévisionnel estimatif de la procédure	69
8.6	Critères de sélection	69
8.7	Critères d'attribution	70
8.7.1	Critère 1 – Expérience	70
8.7.2	Critère 2 – Approche méthodologique et fonctionnelle	70
8.7.3	Critère 3 – Approche technique	71
8.7.4	Critère 4 – Présentation	72

8.7.5	Critère 5 – Prix.....	73
8.8	Notation.....	73
8.9	Format et contenu de l'offre	73
9.	Conditions d'exécution du marché.....	74
9.1	Règlementation applicable.....	74
9.2	Régime tarifaire.....	74
9.3	Continuité d'activité	75
9.4	Droit d'inspection	75
9.5	Résiliation	75
9.6	Modification et adaptation du marché	76
9.7	Sécurité, confidentialité et protection des données à caractère personnel	76
9.8	Propriété de la solution	76
9.9	Communication autour de la solution.....	77
9.10	Responsabilité	77
9.10.1	Responsabilité délictuelle.....	77
9.10.2	Responsabilité contractuelle	77
9.11	Litige	78
10.	Annexes	78
10.1	Bordereau de soumission	78
10.2	Accord de confidentialité – société.....	78
10.3	Accord de confidentialité – individuel.....	78
10.4	Accord relatif à la protection des données à caractère personnel.....	78

1. Informations sur le marché

1.1 Nom du marché

P1165 – DataQuest. Outil fondé sur des technologies d'IA permettant la génération automatisée de l'inventaire des collections de la Chambre des Députés.

1.2 Pouvoir adjudicateur

Chambre des Députés du Grand-Duché de Luxembourg (ci-après nommée « CHD » ou « pouvoir adjudicateur »), représentée par M. Claude Wiseler, Président de la Chambre des Députés.

Adresse postale : 23, rue du Marché-aux-Herbes, L-1728 Luxembourg

1.3 Personne de contact

Dossier suivi par :

Amandine Gorse

Service des séances, des affaires générales et des archives

Chambre des Députés du Grand-Duché de Luxembourg

Tél. : +352 466 966-650

E-mail : agorse@chd.lu

Et

Sven Oostvogels

Service innovation et données

Ministère de la Digitalisation

Tél. : +352 247-72130

E-mail : data@digital.etat.lu.

1.4 Type de secteur d'activité

Type de marché : Services

Code CPV : Services de technologies de l'information, conseil, développement de logiciels, Internet et appui (72000000), Logiciels et systèmes d'information (48000000)

1.5 Type de procédure

Il s'agit d'une procédure restreinte sans publication d'avis pour un marché public de faible envergure (Livre I de la Loi du 8 avril 2018 sur les marchés publics).

1.6 Date d'envoi du cahier des charges

07/09/2026

1.7 Date limite de réception des offres

Date limite de remise des offres par mail à archives@chd.lu avec copie à data@digital.etat.lu au plus tard le 07/10/2026, 12h00.

En matière de dépôt, se référer à la section 8.1 *8.1Dépôt des offres* du présent cahier des charges.

1.8 Demandes de renseignements ou de clarifications

Date limite pour la réception des questions ou demandes de clarification : 16/09/2026 à 12h00.

L'envoi des questions se fait exclusivement par mail à archives@chd.lu avec copie à data@digital.etat.lu.

Il y sera répondu simultanément à tous les opérateurs économiques par publication sur le site web du GovTechLab au plus tard le 25/09/2026.

1.9 Constatations d'ambiguïtés, d'erreurs ou d'omissions

L'opérateur économique qui constate dans le présent cahier des charges des ambiguïtés, erreurs ou omissions est tenu, sous peine d'irrecevabilité, de les signaler au pouvoir adjudicateur au plus tard pour le 16/09/2026 à 12h00 par mail à archives@chd.lu avec copie à data@digital.etat.lu.

Les réponses seront transmises à tous les opérateurs économiques au même moment que les réponses aux questions.

1.10 Propriété du cahier des charges

Le présent cahier des charges ainsi que, le cas échéant, tous les documents annexes subséquents de ce marché appartiennent à la CHD.

Tous les documents produits par le pouvoir adjudicateur et l'opérateur économique dans le cadre du présent marché sont strictement confidentiels et ne peuvent donc pas être divulgués à des tiers à l'exception des documents publiés sur le site internet du GovTechLab.

1.11 Acceptation du cahier des charges

Par la remise d'une offre, l'opérateur économique reconnaît qu'il a réuni toutes les informations nécessaires pour établir une offre valable, c'est-à-dire qu'il a pu se rendre compte des difficultés et particularités des travaux, fournitures ou services à exécuter ou à livrer et les prend en considération dans l'élaboration de son offre.

Toute modification apportée aux éléments inclus dans le présent cahier des charges de la part de l'opérateur économique est considérée comme nulle et non avenue.

1.12 Indemnité pour l'élaboration de l'offre

Le pouvoir adjudicateur ne versera pas d'indemnité aux soumissionnaires.

1.13 Sous-traitance

Un opérateur économique peut envisager de soumettre une offre de candidature en ayant recours à la sous-traitance. La part de marché et les activités que l'opérateur économique entend sous-traiter devra être indiquée dans son offre.

Dans le cas du recours à la sous-traitance, la remise d'une offre collective est interdite.

L'opérateur économique principal ayant recours à des sous-traitants assume l'entière responsabilité envers le pouvoir adjudicateur. Les tâches confiées à des sous-traitants telles que prévues par l'opérateur économique principal devront être clairement identifiées dans l'offre qui sera remise.

L'opérateur économique principal est tenu d'identifier tous les sous-traitants pour autant que ceux-ci soient connus au moment de l'établissement de l'offre.

L'opérateur économique joindra à son offre une fiche de sous-traitance dans laquelle il indique les noms et adresses des sous-traitants, la part de marché ou l'activité qu'il entend sous-traiter à chacun, ainsi que tous les documents dont il est question ci-dessus. Cette sous-traitance ne pourra toutefois pas dépasser 45%.

Ne peut être prise en considération une offre en nom personnel émanant d'un opérateur économique si celui-ci figure également en tant que sous-traitant dans une entreprise générale ou s'il remet parallèlement une offre en association momentanée avec un ou plusieurs autres opérateurs économiques. Au cours de l'exécution du projet, tout changement de sous-traitant identifié dans l'offre ou sous-traitance supplémentaire sera soumise à l'approbation écrite préalable du pouvoir adjudicateur.

Une copie de la lettre d'engagement entre l'opérateur économique principal et le sous-traitant sera demandée par le pouvoir adjudicateur.

1.14 Offre collective

Les groupements d'opérateurs économiques, y compris les associations temporaires ou momentanées, peuvent participer au marché. Ils ne sont pas contraints d'avoir une forme juridique déterminée pour présenter une offre collective.

En cas d'une offre collective, celle-ci est obligatoirement accompagnée d'un engagement solidaire, daté et signé, dans lequel les opérateurs économiques désignent parmi eux un mandataire. Ce mandataire sera considéré comme l'interlocuteur unique entre les membres du groupement d'opérateurs économiques et le pouvoir adjudicateur. L'offre indique soit la proportion assumée dans l'exécution du marché, et, le cas échéant, dans chacun de ses éléments, par chacun des opérateurs, soit l'apport proportionnel effectué par chacun d'eux dans l'exécution du marché dans son ensemble ou dans celle de ses différents éléments.

L'ensemble des opérateurs économiques de l'offre collective devront répondre aux critères de sélection sous peine d'irrecevabilité de celle-ci.

Ne peut être prise en considération une offre en nom personnel émanant d'un opérateur économique si celui-ci figure également en tant que sous-traitant dans une entreprise générale globale ou partielle, ou s'il remet parallèlement une offre en association avec un ou plusieurs autres opérateurs économiques.

Le contrat, après attribution du marché, sera signé uniquement entre le pouvoir adjudicateur et le mandataire de l'offre collective.

Une copie du protocole d'entente entre les membres du groupe sera demandée par le pouvoir adjudicateur.

1.15 Lots

La division en lots n'est pas justifiée pour ce marché considérant son envergure, sa portée et son ambition fonctionnelle.

1.16 Variantes

Les variantes ne sont pas acceptées.

1.17 Durée du contrat

Le contrat durera le temps de l'implémentation du projet selon le planning proposé par le prestataire sélectionné jusqu'à la clôture de la dernière phase d'hypercare pour un maximum de 11 mois à compter du début des travaux. La durée maximale de l'implémentation du projet ne doit pas dépasser 8 mois et la dernière phase d'hypercare 3 mois.

Le démarrage des prestations prévu le 11 janvier 2027.

1.18 Prix

Le présent marché sera placé sous le régime de l'offre à prix global non révisable ([art. 8 du Règlement Grand-Ducal du 8 avril 2018 portant exécution de la loi sur les marchés publics](#)).

Le détail du prix sera proposé par l'opérateur économique dans son offre et ne devra pas dépasser 140.000 EUR hors TVA contingence incluse.

Le détail du prix doit inclure un budget de contingence de 10% qui sera utilisé par le pouvoir adjudicateur en cas de changement de son fait et pour des évolutions éventuelles. Ces 10% ne pourront être facturés par l'adjudicataire qu'après une validation formelle écrite du pouvoir adjudicateur et une estimation par ce dernier, des travaux supplémentaires à réaliser.

2. Objet du marché

2.1 Équipe en charge du projet

Name	Organisation	Service	Rôle
Amandine GORSE	CHD	Service des séances, des affaires générales et des archives	Cheffe de projet
Maxime DETANT	CHD	Service des séances, des affaires générales et des archives	Expert e-archiviste
Maria MATHIEU	CHD	Service des séances, des affaires générales et des archives	Responsable de service
Thomas VENEZIANO	CHD	Service des Technologies de l'information	AMOA

Loïc SAINT-GHISLAIN	CHD	Service des Technologies de l'information	PMO
Sven OOSTVOGELS	Ministère de la Digitalisation	Service Innovation et données	Sponsor
Paula ALMEIDA	Ministère de la Digitalisation	Service Innovation et données	Déléguée à l'archivage
Christine ZOLLER	Ministère de la Digitalisation	Service Innovation et données	GovTech Lab

2.2 Type de solution recherchée

Le présent marché porte sur la mise en œuvre d'un outil d'intelligence artificielle (IA) permettant l'identification/catégorisation et la description archivistique d'informations au sein des collections de la CHD et la génération automatisée de réponses.

La solution devra répondre à deux cas d'usage principaux :

- La rédaction automatisée d'inventaires archivistiques (*Thematic DataQuest*) ;
- L'identification des données sensibles au sein des documents de la CHD (*Sensitive DataQuest*).

La CHD recherche un outil basé sur des technologies d'IA qui permet d'automatiser les cas d'usages métiers spécifiques décrits ci-dessus.

2.3 Besoin et finalités

2.3.1 Finalités du projet

Les résultats attendus par les métiers portent notamment sur l'enrichissement et la fiabilisation des métadonnées, l'identification des données sensibles afin de renforcer la gestion de la confidentialité et la conformité réglementaire, ainsi que sur l'amélioration de l'exploitation et de la valorisation des fonds documentaires. L'outil mis en œuvre devra constituer une aide à la décision pour les archivistes, sans jamais se substituer à leur expertise, ni à leur validation finale.

À travers ce projet, la CHD vise à améliorer l'efficacité opérationnelle des agents, à disposer d'un inventaire fiable des données sensibles présentes dans les documents afin de répondre aux obligations légales en vigueur, et à réduire les coûts liés au traitement et à la gestion de ses archives numériques.

2.3.2 Objectifs de réussite

La réussite du projet sera appréciée au regard de la capacité de la solution à produire des résultats exploitables dans les délais impartis, afin de démontrer rapidement sa faisabilité et sa valeur ajoutée pour les équipes de la CHD. Elle devra garantir la souveraineté, la confidentialité et la sécurité des données traitées, maintenir un niveau élevé de maîtrise humaine et atteindre des niveaux de qualité objectivement mesurables. La réussite inclut également la facilité de prise en main et l'autonomie effective de la CHD : à l'issue du projet, des administrateurs fonctionnels formés devront pouvoir faire évoluer les règles, prompts, seuils, référentiels, profils de

configuration et workflows livrés par le titulaire, en créer de nouveaux, les tester, les publier et revenir à une version antérieure sans modification du code ni intervention du titulaire.

Enfin, son architecture et ses fonctionnalités devront offrir la flexibilité nécessaire pour intégrer, à terme, de nouveaux fonds documentaires, de nouveaux cas d'usage et de nouveaux périmètres fonctionnels, en cohérence avec l'évolution des besoins de la CHD.

2.4 Périmètre général du marché

2.4.1 Nature des prestations attendues

Le marché comprend l'ensemble des prestations nécessaires à la conception, au déploiement et à l'exploitation de la solution d'IA. Il couvre également une phase de cadrage et d'analyse destinée à analyser les besoins techniques de la CHD. Le titulaire assurera la conception de l'architecture de la solution. Il fournira les licences logicielles nécessaires à son fonctionnement. Les prestations incluent également l'installation, le paramétrage et l'intégration de la solution dans l'environnement technique de la CHD, ainsi que la mise en œuvre des cas d'usage métiers identifiés (*Thematic DataQuest* et *Sensitive DataQuest*).

Le marché comprend en outre les opérations de recette, de mise en production, de formation et de transfert de compétences, ainsi qu'une phase d'hypercare de trois mois déclenchée après chaque mise en production de tout ou partie d'un incrément. Les prestations récurrentes de RUN, de support hors hypercare et de maintenance applicative corrective, préventive, adaptative ou évolutive ne font pas partie du présent marché. Les corrections d'anomalies et ajustements de paramétrage nécessaires à la conformité des éléments livrés restent toutefois dus pendant la recette et l'hypercare de l'incrément concerné.

2.4.2 Vision de la solution cible

La solution attendue par la CHD devra être appréhendée comme un ensemble cohérent dont l'architecture sera suffisamment modulaire pour permettre d'inclure facilement d'autres besoins ou d'inclure la solution dans une solution plus globale. L'objectif du marché n'est pas de déployer un orchestrateur technique comme finalité première mais de démontrer la valeur ajoutée de composants répondant de manière pragmatique et évolutive aux besoins opérationnels de la CHD.

L'architecture proposée devra reposer sur des modules indépendants, afin de garantir la flexibilité, la maintenabilité et l'évolutivité de la solution dans le temps. Cette approche devra permettre à la CHD de conserver sa liberté de choix quant aux technologies utilisées, de remplacer ou faire évoluer certains composants sans remise en cause de l'ensemble du dispositif, et d'intégrer de nouveaux services, modèles d'IA ou sources de données en fonction de l'évolution des besoins.

La solution devra privilégier le paramétrage fonctionnel plutôt que le développement spécifique. Les opérations courantes de création ou d'évolution d'une règle, d'un prompt, d'un seuil, d'un référentiel, d'un profil de traitement ou d'un contrôle devront être réalisables par un administrateur métier de la CHD au moyen d'interfaces documentées, compréhensibles et sécurisées, sans intervention du titulaire.

2.4.3 Évolutivité de la solution

La solution mise en œuvre sera limitée aux deux cas d'usage initiaux du programme DataQuest (*Thematic DataQuest* et *Sensitive DataQuest*). Sa modularité et son évolutivité permettront de réaliser d'autres cas d'usage qui sont pour l'instant hors périmètre du projet. Elle devra être conçue dès l'origine comme une plateforme évolutive, capable d'accompagner le développement

progressif de nouveaux usages métiers au sein de la CHD. À ce titre, la solution devra permettre l'intégration de nouveaux fonds documentaires, de nouveaux processus de traitement, de nouvelles catégories d'utilisateurs ainsi que de services complémentaires répondant à des besoins métiers futurs. Elle devra également faciliter l'ajout, le remplacement ou la combinaison de différents modèles d'IA, sans remise en cause de l'architecture globale ni des composants déjà déployés.

La CHD recherche un outil basé sur des technologies d'IA permettant d'adresser les deux cas d'usage métiers dans une logique de pérennité, de modularité et d'amélioration continue.

3. Contexte, existant, données et périmètres documentaires

3.1 Contexte institutionnel et projet

3.1.1 Présentation de la CHD

La Chambre des Députés du Grand-Duché de Luxembourg est l'unique assemblée législative nationale. Elle exerce le pouvoir législatif, contrôle l'action du Gouvernement et participe au débat démocratique ainsi qu'aux processus institutionnels nationaux et européens. Son organisation repose sur 60 députés élus, des commissions parlementaires, des groupes et sensibilités politiques et une administration parlementaire assurant le soutien aux activités législatives et administratives.

En raison de la nature sensible des informations traitées et de son rôle stratégique au sein des institutions de l'État, la CHD est soumise à des exigences fortes en matière de confidentialité, de souveraineté numérique, de continuité d'activité, de traçabilité, d'intégrité et d'auditabilité des données et des systèmes d'information.

3.1.2 Contexte archivistique et documentaire

La CHD possède des archives physiques et nativement numériques et numérisées (depuis les années 1990, uniquement pour les documents publics). Le service en charge des archives assure la gouvernance, la collecte, la conservation, la sécurisation et la valorisation des archives de l'institution tout au long de leur cycle de vie.

Dans un contexte de transformation numérique, la CHD fait face à une augmentation significative des volumes documentaires ainsi qu'à une diversification des formats et des sources de données. Le développement des archives électroniques constitue un enjeu majeur afin de garantir la pérennité, l'intégrité, la traçabilité, l'accessibilité et la valeur probante des informations conservées.

Les approches reposant exclusivement sur des traitements manuels montrent aujourd'hui leurs limites face à la masse documentaire à traiter, à l'hétérogénéité des fonds et aux exigences croissantes de qualité, de conformité et d'efficacité opérationnelle. Le recours à des outils d'assistance à la description, à l'indexation et à la qualification documentaire constitue dès lors un levier essentiel d'optimisation des processus d'archivage électronique.

3.1.3 Présentation du projet DataQuest

Cette démarche s'inscrit dans un contexte de renforcement des exigences réglementaires, notamment à la suite de la loi du 17 août 2018 relative à l'archivage, qui impose une maîtrise accrue des conditions de conservation, des délais de restriction et des modalités de communicabilité des archives. Or, la CHD stocke actuellement plus de 4,3 millions de documents

numériques qu'il faut traiter suivant les exigences de la loi du 17 août 2018 relative à l'archivage. Cela représente environ 5365 jours-personne de travail, soit 25 ans de travail à temps-plein pour 1 archiviste. Au regard de ces éléments de volumétrie, le traitement manuel de l'ensemble de ces documents est inenvisageable.

Un tel projet a été envisagé pour se rapprocher de la conformité à la loi, fournir une meilleure efficacité de recherche, ainsi qu'une meilleure gestion de la confidentialité (et diffusion). Ce projet permettra également une réduction des coûts et un gain de temps par rapport au traitement manuel.

Dans une première phase, le projet a pour objectif à la fois de mettre en œuvre l'utilisation de l'IA dans le traitement des fonds documentaires, mais aussi d'industrialiser à plus grande échelle. Le périmètre de DataQuest est limité aux archives électroniques, qu'elles soient nativement numériques ou issues de la numérisation de documents papier.

Les objectifs poursuivis sont de réduire le temps consacré aux tâches de description à faible valeur ajoutée, d'harmoniser la qualité des inventaires produits et d'accélérer le traitement de volumes documentaires importants.

3.1.4 Cas d'usage *Thematic DataQuest*

3.1.4.1 *Description*

La CHD souhaite mettre en œuvre un outil permettant la rédaction automatique d'inventaires archivistiques. L'objectif est d'automatiser, de manière contrôlée et supervisée, la production des informations descriptives nécessaires à l'élaboration des inventaires, tout en conservant la responsabilité finale de validation et de correction par les archivistes.

La solution attendue devra permettre la complétion des résultats dans une matrice d'inventaire organisée en colonnes thématiques. L'analyse pourra être menée au niveau des pièces, mais devra notamment produire une restitution consolidée au niveau du dossier.

Une matrice d'inventaire préremplie constituera la base de travail et comportera notamment :

- L'identifiant unique du dossier
- L'identifiant du plan de classement correspondant
- La référence au tableau de tri applicable (code de série)

Par ailleurs, les fonds documentaires auront préalablement fait l'objet d'une organisation et d'un classement archivistique. Ce contexte documentaire constitue un élément essentiel du traitement et pourra être exploité par l'IA afin d'améliorer la compréhension du contenu, de la fonction administrative du dossier et de son positionnement dans l'arborescence documentaire.

3.1.4.2 *Résultat attendu*

À partir de l'analyse du contenu des documents et des informations de contexte, l'IA devra identifier les informations pertinentes et compléter automatiquement les champs descriptifs manquants de l'inventaire.

La solution devra proposer :

- Un titre de dossier synthétique, normalisé selon la définition donnée par les Archives Nationales du Luxembourg et représentatif du contenu¹,
- Une description archivistique permettant d'identifier l'objet, le contexte et la nature du dossier²,
- La période de création des documents, comprenant :
 - o La date de début
 - o La date de fin

Les résultats générés devront :

- Respecter les règles de saisie et de rédaction validées par la CHD. Ces règles devront être gérées comme des paramètres métier versionnés, modifiables et activables par un administrateur fonctionnel de la CHD, sans modification du code ; elles pourront varier selon le fonds, le projet de traitement ou le niveau de détail attendu.
- Être restitués directement dans la matrice d'inventaire.
- Conserver une cohérence avec les règles de description archivistique en vigueur. La CHD devra pouvoir faire évoluer les règles existantes, en créer de nouvelles, les tester sur un corpus de référence, les publier à une date choisie et revenir à une version antérieure en toute autonomie.
- Être exportés dans un format bureautique modifiable par les archivistes.

3.1.5 Cas d'usage *Sensitive DataQuest*

3.1.5.1 Description

Dans le cadre de la gestion des archives électroniques et de la maîtrise des risques liés à l'information, la CHD souhaite mettre en œuvre une solution permettant d'identifier et qualifier les données sensibles présentes au sein des fonds documentaires.

L'objectif est de renforcer la connaissance des contenus archivés, de faciliter l'application des règles d'accès et de diffusion, et de limiter les risques de divulgation non autorisée d'informations sensibles, tout en maintenant une validation finale par les experts métier.

Comme pour le *Thematic DataQuest*, le traitement s'appuiera sur les informations déjà disponibles au sein des fonds documentaires :

- Contenu textuel des documents
- Métadonnées existantes dans la matrice d'inventaire
- Plan de classement du dossier

3.1.5.2 Définition des cas sensibles

Les définitions des cas sensibles reposent sur :

- L'analyse sémantique selon les définitions données ci-dessous.
- L'exploitation d'exemples de référence
- L'analyse de caractéristiques documentaires

¹ Le titre doit énoncer de manière claire, précise, explicite et non-équivoque l'objet de l'unité de description de façon qu'un tiers comprenne directement le sujet du dossier. La formulation du titre doit se baser sur la réponse à au moins une des questions suivantes :

- De quoi s'agit-il (l'objet) ?
- Comment l'objet est-il traité (l'action) ?
- Quel est le type de documents ?

² Suivant la définition des Archives nationales de Luxembourg : Si toutes les informations ont déjà été données dans le titre, il n'est pas nécessaire de remplir ce champ. Il s'agit d'indiquer toute information pouvant compléter le titre en communiquant des informations supplémentaires, plus précises, sur le contenu d'un dossier. Si le dossier comprend des procès-verbaux de 1939 à 1950, et que vous constatez que les années 1947 et 1948 sont absentes, il convient de noter cet état de fait dans le champ description.

Cette approche multicritère vise à améliorer la qualité de la qualification et à limiter les faux positifs ou faux négatifs.

Le tableau ci-dessous donne une première approche des définitions des cas sensibles. Les définitions sont données dans leur version synthétiques. Le titulaire devra toujours se référer au document de référence des définitions produit et fourni par la CHD. Ces définitions pourront évoluer au cours du projet, au regard du contexte, des évolutions législatives, de l'avis des Archives Nationales du Luxembourg, etc. Le système devra donc être apte à accepter et faire évoluer la nature des définitions au gré de son utilisation par les administrateurs métier. L'implémentation des nouvelles définitions devra pouvoir se faire par les administrateurs métier (archivistes).

Cas sensible	Définition	Exemples	Caractéristiques
Droit d'auteur	Un document est protégé par le droit d'auteur lorsqu'il constitue ou contient une œuvre originale, c'est-à-dire une création intellectuelle propre à son auteur, exprimée sous une forme perceptible. Le droit d'auteur protège l'œuvre incorporée dans le document, et non le support matériel lui-même. Les droits d'auteur protègent les œuvres littéraires et artistiques originales, quels qu'en soient le genre et la forme ou l'expression (...). Ils ne protègent pas les idées, les méthodes de fonctionnement, les concepts ou les informations, en tant que tels	– Articles de presse – Articles scientifiques – Brochures produites par d'autres institutions – Supports de formation – Extraits d'ouvrages littéraires	Tout document où un auteur est mentionné.
Données à caractère personnel	Une donnée à caractère personnel est une information qui permet d'identifier un être humain (personne physique), directement (par exemple son nom/prénom), ou indirectement (par exemple son numéro de téléphone, son numéro de contrat, son pseudonyme). Définition du RGPD : Toute information concernant une personne physique identifiée ou identifiable (personne concernée) ; une personne qui peut être identifiée, directement ou indirectement, notamment par référence à un numéro d'identification ou à un ou plusieurs éléments spécifiques, propres à son identité physique, physiologique, psychique, économique, culturelle ou sociale.³	– Passeports – Livrets de famille – Lettres contenant des données personnelles – Annuaire, listes d'adresses – Visas – Informations sur les députés ou le personnel (groupe sanguin, plaque d'immatriculation)	Coordonnées (Adresse postale, adresse électronique, numéro de téléphone) Identifiants techniques (Adresse IP, identifiant de connexion, cookies de suivi) Critères spécifiques (Âge, taille, voix, orientation sexuelle, données de santé) <u>Ce que n'est pas la donnée :</u> Les noms/prénoms en lien avec les députés et agents de la CHD, ainsi que les opinions des députés exprimées publiquement.

³ CNPD, Avis 4/2007 sur le concept de données à caractère personnel, cf. https://cnpd.public.lu/dam-assets/fr/publications/groupe-art29/wp136_fr.pdf.

Actes d'état civil	Un acte de l'état civil est un acte dressé par un officier d'état civil destiné à prouver l'état d'une personne (naissance, mariage, décès).	– Actes de naissance – Actes de mariage – Actes de décès	L'acte est rédigé par un officier de l'état civil et concerne un fait d'état civil. L'acte est enregistré dans les registres de l'état civil. Il comporte des mentions obligatoires : - L'identification de la commune ; - La date et le lieu de l'établissement de l'acte ; - Les informations d'identité des personnes concernées (nom, prénom, date et lieu de naissance, etc.) ; - Les déclarations recueillies par l'officier de l'état civil ; - La signature de l'officier de l'état civil (ou une validation électronique conforme). Il est délivré sous forme de copie intégrale ou d'extrait. Il comporte des éléments de sécurité (sceau ou le cachet de la commune, signature manuscrite ou électronique qualifiée, code de vérification). Il est émis par une commune luxembourgeoise.
Actes notariés	Document prouvant un accord, établi par un notaire et signé en sa présence par les parties concernées qui sont nommées dans l'acte. Au Luxembourg, la notion d'acte notarié (ou acte authentique reçu par un notaire) est principalement régie par la <u>Loi modifiée du 9 décembre 1976 relative à l'organisation du notariat</u> , ainsi que par les dispositions du Code civil concernant les actes authentiques.	– Actes de vente – Testaments – Donation – Reconnaissance de dette – Actes de constitution de sociétés – Actes de constitution d'association – Contrat de mariage – Constitution d'hypothèque – Actes de successions	Chaque notaire doit disposer d'un cachet ou sceau personnel comportant, en français ou en allemand, son identité, sa qualité, sa résidence et les armes couronnées du Grand-Duché de Luxembourg selon un modèle uniforme. Les actes notariés doivent mentionner l'identité du notaire, celle des parties (et, le cas échéant, des témoins), ainsi que le lieu et la date de l'acte. Les dates, les mois et les sommes doivent être écrits en

			toutes lettres, sauf certaines évaluations et, pour les opérations de compte, seuls les totaux et soldes. Les pièces annexées doivent être signées ou paraphées par les comparants et le notaire, qui en fait mention dans l'acte. Les actes ne peuvent comporter ni surcharge, ni interligne, ni ajout. Les ratures doivent rester lisibles et être constatées et approuvées par les signataires. Enfin, les actes doivent être rédigés en français ou en allemand, au choix des parties <u>Ce que n'est pas le document :</u> Un document signé uniquement par des particuliers ou par un avocat n'est pas un acte notarié
Documents qui porteraient atteinte aux relations extérieures, à la sécurité du Grand-Duché de Luxembourg ou à l'ordre public	Documents dont la divulgation, la communication ou la diffusion non autorisée serait susceptible de compromettre les intérêts diplomatiques, les relations internationales, la sûreté, la défense, la sécurité des personnes et des infrastructures, les capacités opérationnelles des autorités publiques ou le maintien de l'ordre public au Grand-Duché de Luxembourg. Il s'agit d'une clause juridique de protection des intérêts essentiels de l'État, utilisée également dans la législation sur l'accès aux documents administratifs, dont l'interprétation repose sur une appréciation concrète du risque lié à la communication du document plutôt que sur une qualification prédéfinie des documents eux-mêmes. Cette même formule apparaît	– Documents sur les dispositifs de sécurité lors des visites – Plans des bâtiments de la CHD – Documents confidentiels – Documents sur des négociations internationales – Documents sur la défense militaire – Rapport diplomatique – Documents en lien avec les relations extérieures, la sécurité du Grand-Duché et l'ordre public	Son contenu concerne un intérêt essentiel de l'État, porte l'effet prévisible de divulgation, contient des informations stratégiques ou confidentielles, est produit ou détenu par une autorité exerçant des missions régaliennes, ou fait l'objet de mesures particulières de protection. <u>Ce que n'est pas le document :</u> Ne relèvent pas automatiquement de cette catégorie les documents ayant simplement trait à des relations extérieures, à la sécurité ou à l'ordre public, lorsque leur contenu est déjà public, purement administratif ou

	aussi dans la <u>Loi du 14 septembre 2018 relative à une administration transparente et ouverte</u>, qui exclut de la communication les documents dont la divulgation porterait atteinte.		dépourvu d'impact réel sur les intérêts protégés. Un document ne relève pas automatiquement de cette catégorie parce qu'il : - Est ancien - Provient d'un ministère - Comporte des données personnelles (celles-ci relèvent d'un autre régime de protection) - Est marqué « confidentiel » en interne - Traite de sujets politiques ou administratifs La qualification repose toujours sur l'existence d'un risque concret d'atteinte aux intérêts protégés par la loi.
Affaires portées devant les instances juridictionnelles, extrajudiciaires ou disciplinaires	Les affaires portées devant les instances juridictionnelles, extrajudiciaires ou disciplinaires sont les dossiers relatifs à une procédure engagée devant une autorité légalement compétente afin de trancher un litige, de statuer sur un recours, de régler un différend ou d'apprécier la responsabilité ou le comportement d'une personne, que cette autorité appartienne à l'ordre juridictionnel, exerce une mission de règlement extrajudiciaire des différends ou dispose d'un pouvoir disciplinaire. Cette expression constitue une notion descriptive employée par le législateur pour désigner une catégorie de dossiers bénéficiant d'un délai particulier de communicabilité, sans qu'elle fasse l'objet d'une définition autonome dans un règlement grand-ducal ou dans la loi elle-même. L'expression vise non pas une catégorie	- Documents des juridictions administratives ou autre type de juridiction - Jugements des tribunaux - Documents sur des procédures extrajudiciaires (médiateur, conseil arbitral de la sécurité sociale) - Documents rendant compte de procédures disciplinaires	En pratique, l'interprétation repose sur plusieurs critères. Une « affaire portée devant une instance juridictionnelle, extrajudiciaire ou disciplinaire » présente généralement les caractéristiques suivantes : - Existence d'un litige, d'une contestation ou d'une mise en cause ; - Saisine d'une autorité légalement compétente ; - Existence d'une procédure organisée par un texte ; - Instruction ou examen contradictoire (selon les cas) ; - Adoption d'une décision, d'un avis ou d'une sanction.

	documentaire particulière, mais un type de procédure. Autrement dit, ce n'est pas la nature du document qui est déterminante, mais le fait qu'il soit rattaché à une procédure de règlement d'un litige ou de mise en cause d'une personne devant une autorité compétente.		L'expression vise donc davantage le contexte procédural que la nature intrinsèque des documents.
Documents ayant trait à la prévention, à la recherche ou à la poursuite de faits punissables	Documents relatifs à la détection, l'enquête, l'instruction, la poursuite ou la répression d'infractions pénales, dont la divulgation pourrait compromettre le travail des autorités compétentes ou le bon déroulement des procédures. Cette expression n'est pas isolée dans le droit luxembourgeois. Elle est reprise, avec des formulations très proches, dans d'autres textes, notamment en matière d'accès aux documents administratifs et de protection des données. Son interprétation peut donc être éclairée par ces autres régimes juridiques : - La loi du 14 septembre 2018 relative à une administration transparente et ouverte, qui permet de refuser la communication de documents lorsque celle-ci porterait atteinte à la prévention, à la recherche ou à la poursuite de faits punissables ; - Le Règlement (UE) 2016/679 (RGPD) et surtout la directive (UE) 2016/680 (« Police-Justice »), qui distinguent les traitements de données effectués aux fins de la prévention, de l'investigation, de la détection ou de la poursuite d'infractions pénales et de l'exécution des sanctions pénales.	- Procès-verbaux de police judiciaire - Rapports d'enquête - Procès-verbaux d'auditions - Réquisitions judiciaires - Commissions rogatoires - Dossiers d'instruction - Rapports d'expertise judiciaire - Rapports de constatation - Analyses balistiques ou ADN - Dossiers d'instruction - Pièces à conviction documentaires - Notes opérationnelles - Notes de renseignements criminels - Analyses de renseignement criminel - Demandes d'entraide pénale internationale - Inventaires de pièces à conviction	Le critère déterminant n'est pas l'autorité qui détient le document, mais le lien direct ou suffisamment étroit entre le document et une activité relevant de la fonction répressive de l'État. <u>Ce que le document n'est pas :</u> Un document administratif ordinaire ne relève pas de cette catégorie au seul motif qu'il est détenu par la police ou le parquet : il doit présenter un lien avec une activité de prévention, de recherche ou de poursuite.
Documents dont la communication	Documents contenant des informations économiques, commerciales, techniques ou	- Procédés de fabrication - Méthodes de production - Formules ou recettes	Le document contient des informations non publiques et renferme des

porterait atteinte au caractère confidentiel des informations commerciales et industrielles	stratégiques non publiques dont la divulgation pourrait porter préjudice aux intérêts, à la compétitivité ou aux activités d'une entreprise ou d'un organisme. Cette expression est très proche de celle utilisée dans la loi du 14 septembre 2018 relative à une administration transparente et ouverte, qui prévoit une exception au droit d'accès lorsque la communication porterait atteinte : - Aux intérêts commerciaux ; - Au secret des affaires ; - Au caractère confidentiel d'informations commerciales ou industrielles.	- Savoir-faire technique - Stratégies commerciales - Plans de développement - Études de marché non publiées - Listes de clients ou de fournisseurs - Conditions contractuelles confidentielles - Prix négociés - Marges commerciales	informations qui ne sont pas librement accessibles, qui ne figurent pas dans des registres publics et qui ne sont pas connues des concurrents ou du public. Le caractère non public est un indice essentiel. Ensuite, les informations présentent une valeur économique : elles procurent ou sont susceptibles de procurer un avantage économique ou concurrentiel. La communication du document serait susceptible de causer un préjudice économique. Le document fait généralement l'objet de mesures destinées à préserver sa confidentialité La directive (UE) 2016/943 permet d'identifier trois critères particulièrement utiles liés au secret d'affaires
Documents ayant trait au secret fiscal	Documents contenant des informations couvertes par le secret fiscal relatives à la situation fiscale d'une personne physique ou morale, dont la divulgation non autorisée porterait atteinte à la confidentialité protégée par la législation en vigueur Le secret fiscal est une notion juridique prévue par la législation fiscale luxembourgeoise, notamment par l'article 22 de la Loi générale des impôts du 22 mai 1931 (Abgabenordnung). Le secret fiscal impose notamment aux personnes qui participent à l'établissement ou à l'exécution des règles fiscales de préserver la confidentialité des informations dont elles prennent connaissance dans l'exercice de leurs fonctions. Il s'agit donc d'une obligation légale	- Certificats de rémunération et de retenue d'impôt - Fiches de retenue d'impôt - Déclarations fiscales - Avis d'impôt - Dossiers d'imposition - Bons à liquider/feuilles de traitement - Rapport de contrôle fiscal - Correspondance avec un contribuable sur son imposition - Renseignement fiscal transmis par une autorité étrangère	Le premier critère est que le document a été produit, reçu ou utilisé dans le cadre de l'application de la législation fiscale. Le document contient des informations permettant de connaître la situation fiscale : - D'une personne physique ; - D'une société ; - D'une association ; - D'une succession ; - D'une autre personne soumise à l'impôt. Il contient des déclarations ou données fiscales. Les documents constitués dans le cadre d'un contrôle fiscal sont généralement particulièrement susceptibles d'être couverts. Il faut

	de confidentialité applicable aux informations fiscales. Les documents couverts par le secret fiscal sont des documents contenant ou révélant des informations entrant dans le champ de cette obligation. Les informations suivantes sont couvertes par le secret fiscal : - Informations communiquées par un contribuable à l'administration fiscale - Résultats de contrôles fiscaux - Correspondances entre le contribuable et l'administration fiscale - Informations relatives à une enquête ou à une procédure fiscale - Renseignements obtenus auprès d'une autre administration fiscale	- Analyse fiscale concernant un contribuable déterminé - Dossier de réclamation fiscale individuel	également tenir compte des informations qui, prises isolément, semblent anodines mais permettent, combinées entre elles, de déduire une information fiscale protégée. <u>Ce que n'est pas le document :</u> Les documents publics, réglementaires ou administratifs dont la communication ne permet pas de connaître la situation fiscale d'une personne physique ou morale. Un document émanant de l'Administration des contributions directes n'est pas automatiquement couvert par le secret fiscal. Un document émanant de l'Administration des contributions directes n'est pas automatiquement couvert par le secret fiscal. Une mention interne de confidentialité constitue un indice, mais ne crée pas à elle seule le secret fiscal. Un rapport général sur la fiscalité luxembourgeoise peut traiter d'impôts sans contenir d'information individuelle protégée.
--	---	---	--

3.1.5.3 Résultats attendus

La solution attendue devra permettre :

- L'identification des documents (sur la base de leur contenu) selon des catégories prédéfinies,
- L'alimentation d'une matrice d'analyse structurée recensant les critères de sensibilité identifiés,
- La mise à disposition de résultats révisables et validables par les archivistes.

3.1.6 Projets connexes et dépendances

Le présent projet s'inscrit dans un écosystème applicatif et documentaire plus large, structuré autour du projet de conservation de la mémoire digitale de la CHD (COMED), des infrastructures mises à disposition par le Centre des Technologies de l'Information de l'État (CTIE) ainsi que des différentes initiatives de transformation numérique et d'exploitation de l'intelligence artificielle portées par la CHD. Il dépend du projet de traitement des archives numériques de la CHD qui débute en septembre 2026.

L'objectif est de prévoir, au regard du statut des projets en question, une intégration cohérente des futurs traitements d'IA dans la chaîne documentaire existante, tout en évitant la duplication de fonctionnalités déjà couvertes par ces autres projets.

3.1.6.1 Projet COMED

Le projet COMED a pour finalité la mise en place d'une plateforme dédiée au traitement des documents intermédiaires et définitifs en préparation de leur versement dans le Système d'Archivage Électronique (SAE).

Le principe retenu est de permettre aux différents services de déposer leurs documents dans une arborescence proche de leur organisation future. La plateforme contrôle ensuite la présence des métadonnées obligatoires. Lorsque certaines informations sont manquantes, COMED sollicite l'archiviste afin qu'il les complète ou les définisse.

La solution prévoit également l'automatisation de plusieurs tâches archivistiques, notamment :

- La rédaction des bordereaux de versement,
- La rédaction des bordereaux de destruction,
- Le contrôle de la complétude des métadonnées associées aux dossiers.

À plus long terme, le projet pourra être enrichi grâce à la mise en œuvre d'une charte de nommage et d'une charte de métadonnées communes à l'ensemble des services. Ces prérequis ne sont toutefois pas encore en place à ce stade, mais constituent un levier important pour améliorer l'automatisation et la qualité des traitements documentaires.

Le cas d'usage *Thematic DataQuest* constitue un traitement préalable des versements dans COMED. L'identification des données sensibles peut également s'intégrer dans le processus de préparation des versements. Le projet ne doit pas reproduire des fonctionnalités déjà couvertes par COMED ou d'autres briques du système d'information. Sa valeur ajoutée réside dans l'analyse des contenus, l'extraction de connaissances, la qualification automatique et l'assistance aux archivistes, tandis que les fonctions de gestion documentaire, de versement et de conservation demeurent assurées par les plateformes métier existantes ou en cours de déploiement.

3.1.6.2 Projet de traitement des archives numériques

Le projet vise à assurer le traitement de l'arriéré des archives numériques de la CHD et à garantir leur conservation, leur intégrité, leur accessibilité et leur lisibilité, conformément à la loi du 17 août 2018 relative à l'archivage. Le projet doit également contribuer à réduire la dette numérique, à

limiter les risques liés à l'obsolescence technique et à faciliter l'accès aux archives. Il porte sur les documents bureautiques stockés sur les serveurs représentant environ 4,3 millions de fichiers.

Les travaux incluent notamment le récolement, le tri, le classement, la saisie des métadonnées, la préparation à la migration vers un système d'archivage électronique et la réalisation d'instruments de recherche.

Le projet permettra de préparer les fonds documentaires à décrire par l'outil et d'élaborer la matrice d'inventaire préremplie nécessaire à l'outil permettant la rédaction automatique d'inventaires d'archives.

3.2 Processus et environnement existants

3.2.1 Processus actuel de traitement des fonds

À ce jour, le traitement des fonds archivistiques numériques demeure principalement manuel et n'est réalisé que de manière partielle.

3.2.1.1 *Plan de classement*

Le plan de classement constitue le référentiel structurant sur lequel repose l'organisation, la gestion et la valorisation des archives de la CHD. Historiquement élaboré pour les archives papier, il est actuellement formalisé dans un document d'une cinquantaine de pages qui organise les fonds principalement par commissions parlementaires, services versants et catégories documentaires spécifiques. Toutefois, ce référentiel demeure partiel en raison de versements encore incomplets et d'une couverture limitée de certains fonds.

Dans le cadre de la transformation numérique de la gestion documentaire et archivistique, un travail de refonte du plan de classement numérique est actuellement engagé. L'objectif est d'aboutir à un modèle unifié couvrant à terme l'ensemble des archives, qu'elles soient nativement numériques ou issues du patrimoine papier.

Le plan de classement numérique revêt une importance particulière pour les futurs traitements automatisés et les cas d'usage d'IA. La qualité, la stabilité et la granularité de ce référentiel conditionneront directement la pertinence des analyses automatisées et la cohérence des inventaires produits. À ce titre, il constitue l'un des principaux prérequis à la mise en œuvre des futurs dispositifs d'assistance archivistique fondés sur l'IA.

3.2.1.2 *Matrice d'inventaire*

La loi du 17 août 2018 relative à l'archivage impose désormais des exigences précises en matière de délais de restriction et de communicabilité des archives. Dans ce cadre, une matrice d'inventaire (fichier Excel) a été créée en reprenant les principales métadonnées de description et de gestion documentaire :

- Cote,
- Identifiant de la série du plan de classement,
- Code-série du tableau de tri (conservation ou destruction),
- Titre,
- Description du contenu,
- Période de création des documents (dates de début et de fin),
- Langues
- Mots-clés matière,
- Noms cités (personnes, organismes, lieux).

La matrice d'inventaire sera susceptible d'évoluer.

Un volet spécifique consacré à l'identification des données sensibles a également été intégré, permettant de signaler notamment la présence de :

- Droits d'auteur,
- Données à caractère personnel,
- Actes d'état civil,
- Actes notariés,
- Documents qui porteraient atteinte aux relations extérieures, à la sécurité du Grand-Duché de Luxembourg ou à l'ordre public,
- Affaires portées devant les instances juridictionnelles, extrajudiciaires ou disciplinaires,
- Documents ayant trait à la prévention, à la recherche ou à la poursuite de faits punissables,
- Documents dont la communication porterait atteinte au caractère confidentiel des informations commerciales et industrielles,
- Documents ayant trait au secret fiscal.

La matrice d'inventaire est également constituée d'un onglet plan de classement (composé d'un code rubrique et intitulé rubrique). Aujourd'hui, elle est principalement et très majoritairement utilisée pour les archives physiques.

3.2.2 Écosystème institutionnel et technique

La réussite du projet repose sur une gouvernance associant les acteurs institutionnels, les responsables de la conformité et de la sécurité ainsi que les éventuels prestataires techniques. Les responsabilités détaillées, les processus de validation et les modalités opérationnelles seront précisés dans les sections techniques et dans la phase BUILD.

3.2.2.1 *Chambre des Députés*

La Chambre des Députés (CHD) est le maître d'ouvrage du projet et le principal responsable des traitements réalisés dans le cadre de ses missions institutionnelles. Elle définit les objectifs métiers, les cas d'usage, les règles de gestion documentaire et les exigences fonctionnelles attendues. Elle assure également la validation des résultats produits, notamment pour les activités de description archivistique, de qualification documentaire et d'identification des données sensibles.

La CHD pilote les arbitrages fonctionnels, coordonne les différentes parties prenantes internes et veille à la conformité des solutions mises en œuvre avec les besoins de l'institution, les exigences réglementaires et les politiques de gouvernance de l'information.

3.2.2.2 *Hébergeur*

L'hébergeur intervient en tant qu'acteur de référence pour les infrastructures, les services numériques mutualisés et les cadres techniques de l'administration luxembourgeoise. Selon l'architecture retenue, son rôle peut consister à fournir ou encadrer les environnements d'hébergement, les services d'authentification, les mécanismes de sécurité, les plateformes d'échange ou les composants techniques réutilisables. Le Centre des Technologies de l'Information de l'État (CTIE) est actuellement pressenti comme hébergeur.

3.2.2.3 *Ministère de la Digitalisation*

Le ministère de la Digitalisation intervient comme acteur de soutien et de financement du projet. Le lancement de cette initiative s'inscrit dans le cadre des dispositifs d'accompagnement et des mécanismes de financement mis en place par le Ministère pour favoriser la transformation numérique des administrations et institutions publiques luxembourgeoises.

Le Ministère n'a pas vocation à intervenir dans la conduite opérationnelle, les choix archivistiques ou les validations métiers du projet.

3.2.2.4 DPO

Le DPO veille au respect des obligations relatives à la protection des données à caractère personnel. Il accompagne le projet dans l'identification des impacts éventuels sur les données personnelles et contribue à la définition des mesures de conformité appropriées. Son implication peut être particulièrement importante pour les cas d'usage portant sur l'identification automatique de données sensibles.

3.2.2.5 Fournisseur d'infrastructure

Les éventuels fournisseurs d'infrastructure assurent la mise à disposition et l'exploitation des ressources techniques nécessaires au fonctionnement de la solution : hébergement, stockage, sauvegarde, réseaux, supervision, continuité d'activité et dispositifs de sécurité associés.

Leur intervention devra respecter les exigences de souveraineté, de sécurité, de disponibilité et de réversibilité définies par la CHD. Ils pourront être sollicités dans le cadre d'environnements cloud, SaaS ou d'infrastructures mutualisées selon les choix retenus lors de la conception détaillée.

3.2.2.6 Fournisseur de modèles d'IA

Les fournisseurs de modèles d'IA mettent à disposition les composants nécessaires à l'analyse documentaire, à l'extraction d'informations, à la génération de contenus descriptifs ou à la qualification des données sensibles. Leur rôle se limite à la fourniture des capacités technologiques nécessaires à la réalisation des traitements. Les règles métier, les référentiels archivistiques, les critères de qualification et les validations finales demeurent sous la responsabilité de la CHD. Une attention particulière devra être portée aux garanties offertes en matière de confidentialité, d'utilisation des données, de localisation des traitements, de traçabilité et d'explicabilité des résultats.

3.2.3 Orientations relatives aux modèles d'IA

Le projet devra s'inscrire dans les orientations de transformation numérique et d'IA portées par l'État luxembourgeois. À ce titre, il convient de prendre en considération le partenariat conclu en juin 2025 entre le Gouvernement du Luxembourg et Mistral AI, afin de favoriser le développement et l'adoption de solutions d'IA au sein des administrations publiques luxembourgeoises.

Dans ce contexte, les technologies proposées par Mistral AI constituent une orientation privilégiée au sein de l'écosystème numérique de l'État luxembourgeois et pourront être considérées comme une cible de référence pour l'implémentation des cas d'usage prévus dans le cadre du présent projet.

Toutefois, cette orientation ne devra en aucun cas créer une dépendance exclusive à un fournisseur, à un modèle de langage ou à une plateforme donnée. Elle devra notamment permettre l'utilisation de modèles Mistral lorsque ceux-ci sont disponibles et autorisés, tout en restant compatible avec d'autres LLM ou services IA validés dans le cadre des politiques techniques et de sécurité applicables.

Le titulaire devra donc privilégier une approche fondée sur l'interopérabilité, la portabilité des composants et des services d'IA, afin de garantir la capacité d'évolution de la solution dans le temps, d'accompagner les orientations futures du CTIE et de permettre l'intégration de nouveaux modèles ou fournisseurs autorisés sans remise en cause significative de l'architecture globale du projet.

3.2.4 Type de solution recherché

La CHD privilégie la sélection de briques basées sur des technologies d'IA et répondant aux cas d'usage métiers. Ces briques devront être modulaires pour permettre leur intégration dans les outils de la CHD le moment venu. L'objectif recherché est de disposer d'un outil flexible, évolutif et modulaire capable d'enchaîner différents traitements documentaires au sein d'un même processus métier.

L'outil devra pouvoir configurer et exécuter des chaînes de traitement composées de plusieurs étapes successives :

1. Réception ou sélection d'un dossier documentaire,
2. Analyse du contexte documentaire et des données contextuelles (plan de classement, inventaire vierge),
3. Détection des données sensibles,
4. Génération et complétion des métadonnées descriptives et sensibles,
5. Contrôle qualité automatique,
6. Itération humaine par les archivistes en modifiant les paramètres,
7. Restitution des résultats en langage naturel et selon les règles de saisie de la CHD dans les matrices d'inventaire.

La composition de ces chaînes devra rester configurable afin de pouvoir adapter les traitements selon les spécificités de chaque cas d'usage. La plateforme peut être envisagée comme une « boîte à outils documentaire et IA » permettant d'activer, de combiner ou de réutiliser différents traitements selon les besoins.

Selon les situations, les processus pourront :

- Être entièrement automatisés sur certaines tâches simples,
- Combiner traitements classiques et traitements IA,
- Intégrer plusieurs modèles ou moteurs spécialisés,
- Conserver systématiquement une étape de validation humaine lorsque cela est nécessaire.

L'utilisation de l'IA devra ainsi être considérée comme un moyen d'assister les experts métier et de renforcer leur efficacité, sans remettre en cause leur rôle décisionnel.

3.3 Données et fonds documentaires

3.3.1 Sources de données et organisation documentaire

La solution s'appuie sur une organisation documentaire structurée autour de dossiers et plans de classement, qui constituent le contexte principal d'analyse et de traitement.

Les fichiers documentaires contenus dans les dossiers ainsi que la matrice d'inventaire Excel partiellement complétée (identifiant du dossier, identifiant du plan de classement correspondant, référence au tableau de tri applicable soit le code série) constituent les principales entrées du système. Le plan de classement associé apporte un contexte complémentaire permettant d'orienter les traitements, la qualification documentaire et l'interprétation des contenus.

Les principaux résultats produits seront restitués au sein de la matrice d'inventaire Excel, qui demeure l'outil de travail de référence pour les archivistes.

3.3.2 Typologies et formats de documents

La solution devra permettre l'import, l'export, le traitement et la restitution des documents et données dans les formats suivants, considérés comme formats de référence du projet : CSV encodé en UTF-8, EML, ICS encodé en UTF-8, JPEG2000, JSON, MP3, MP4, PDF, SIARD,

TIFF, TXT encodé en UTF-8, WAV, XLS, XLSX, XML, PPT, PPTX, DOC et DOCX. Le titulaire devra garantir la prise en charge native de ces formats, sans altération de leur contenu, de leur structure, de leurs métadonnées associées ou de leurs caractéristiques techniques. Pour les formats dont le traitement peut s'avérer complexe, le titulaire devra proposer une solution basée sur l'analyse des métadonnées. Le système devrait être flexible pour traiter les formats similaires à ceux mentionnés ci-dessus. Les échanges de données devront respecter les spécifications et normes propres à chacun de ces formats afin de garantir leur interopérabilité, leur pérennité et leur réutilisation. Toute conversion éventuelle entre formats devra préserver l'intégrité, la lisibilité et l'exploitabilité des informations contenues dans les fichiers concernés.

3.3.3 Langues des documents et métadonnées

Les fonds documentaires de la CHD sont multilingues et contiennent principalement des documents en français, allemand, luxembourgeois et anglais.

La solution devra détecter automatiquement la langue des documents analysés. Quelle que soit leur langue d'origine, les titres, descriptions archivistes et métadonnées générés devront être rédigés exclusivement en français, afin de garantir l'homogénéité des inventaires et des traitements.

3.3.4 Volumétrie et caractéristiques des fonds

À titre indicatif, la CHD recense actuellement environ 4,3 millions de documents. Sur la base de l'hypothèse de travail retenue, selon laquelle seuls les documents nécessitant une analyse seraient soumis aux traitements d'IA (archives uniquement), le volume potentiellement concerné peut être estimé à environ 1 million de documents (soit 20 % du volume total).

Cette estimation, qui devra être confirmée et affinée lors de la phase de cadrage avec les équipes métiers et informatiques, constitue un ordre de grandeur permettant d'apprécier les besoins de dimensionnement, de performance et de montée en charge de la solution.

3.3.5 Qualité et préparation des données

La qualité des traitements IA dépend directement de la qualité des fonds documentaires analysés. Les principales difficultés identifiées concernent notamment :

- Les fichiers corrompus ou inexploitable,
- Les documents numérisés de qualité insuffisante,
- Les résultats OCR incomplets ou erronés,
- La présence de doublons,
- L'absence ou l'incomplétude des métadonnées,
- Les arborescences documentaires irrégulières,
- Les incohérences de nommage,
- Les formats non supportés ou exclus du périmètre.

Ces situations peuvent dégrader la qualité des analyses, limiter la précision des résultats et nécessiter des traitements préparatoires (nettoyage, normalisation, restructuration ou exclusion).

3.3.6 Données sensibles

La liste précise des données sensibles ainsi que leurs définitions est disponible dans la partie 3.1.5.2 Définition des cas sensibles. Cette liste doit être validée par les juristes/DPO de la CHD, avant leur utilisation dans les processus et sera susceptible d'évoluer sur paramétrage des administrateurs métier (archivistes).

3.4 Périmètres du projet

3.4.1 Périmètres documentaires successifs

Afin de maîtriser les risques et d'évaluer progressivement les performances de la solution, une démarche d'intégration progressive des fonds documentaires sera déployée. Cette approche permettra de déployer les cas d'usage sur des périmètres de complexité croissante, tout en ajustant les paramètres, les modèles et les règles métier au fur et à mesure des retours d'expérience.

Les fonds retenus pour les différentes phases seront caractérisés selon plusieurs critères, notamment :

- Les formats,
- Les langues utilisées,
- La qualité des documents et des métadonnées disponibles.

La démarche pourra ainsi débuter sur des fonds homogènes, bien classés et disposant de métadonnées de qualité, avant d'être progressivement étendue à des ensembles documentaires plus volumineux, plus hétérogènes ou présentant des enjeux accrus en matière de sensibilité des données. Cette montée en charge progressive permettra de sécuriser le déploiement de la solution et d'améliorer continuellement la qualité des traitements réalisés.

3.4.2 Périmètre fonctionnel de la V1

La V1 couvre la mise en œuvre des cas d'usage *Thematic DataQuest* et *Sensitive DataQuest* appliqués aux fonds documentaires sélectionnés par la CHD dans le cadre des phases pilotes et de montée en charge progressive.

Le périmètre initial concerne les documents et dossiers préalablement préparés, organisés selon un plan de classement et mis à disposition sous forme de répertoires, fichiers documentaires ainsi que la matrice d'inventaire Excel. Les traitements porteront principalement sur les formats bureautiques précisément définis lors de la phase de cadrage.

Les fonctionnalités minimales attendues dans le cadre du marché comprennent :

- L'analyse des documents et dossiers mis à disposition,
- La prise en compte du contexte de classement et des métadonnées existantes,
- La détection automatique de la langue des documents (en tant qu'unité de base de l'information enregistrée, quel que soit son support),
- La génération des informations attendues dans le cadre du *Thematic DataQuest* (titre, description, dates, etc.),
- L'identification et la qualification des critères du *Sensitive DataQuest*,
- La restitution des résultats dans la matrice d'inventaire,
- Les mécanismes de validation et de correction par les archivistes,
- La traçabilité des traitements réalisés et des résultats proposés (et supervision/indicateurs),
- La capacité à traiter des formats spécifiques, avancés ou multiformats.

3.4.3 Éléments hors périmètre

Dans sa version initiale (V1), la solution reposera principalement sur l'exploitation de fonds documentaires organisés sous forme de répertoires, de dossiers et de fichiers, ainsi que sur les matrices d'inventaire utilisées par les archivistes. Afin de limiter la complexité technique et de concentrer les travaux sur les cas d'usage prioritaires (*Thematic DataQuest* et *Sensitive DataQuest*), aucune connexion directe aux applications métiers, aux systèmes documentaires existants ou aux bases de données de la CHD n'est prévue dans le périmètre de la V1.

Les traitements seront donc réalisés à partir d'extractions documentaires, de répertoires de travail et des métadonnées mises à disposition dans le cadre du projet.

Toutefois, la solution devra être conçue de manière à faciliter les évolutions futures. Des capacités d'interopérabilité pourront être demandées dans les phases ultérieures du projet afin de permettre l'intégration avec des applications métiers, des plateformes documentaires, des systèmes d'archivage électronique, des référentiels de métadonnées ou des bases de données institutionnelles.

3.4.4 Hypothèses, inconnues et prérequis

À ce stade du projet, plusieurs paramètres structurants demeurent à préciser afin de confirmer les choix de conception, de dimensionnement et de déploiement de la solution.

Les principaux sujets restant à arbitrer concernent notamment :

- L'architecture cible, notamment l'organisation des chaînes de traitement et les modalités d'intégration avec l'environnement existant,
- L'infrastructure d'hébergement, incluant les choix de déploiement, les exigences de souveraineté, les environnements techniques et les responsabilités d'exploitation,
- Le ou les services LLM à utiliser, notamment au regard des orientations portées par l'État luxembourgeois, des solutions mises à disposition par le CTIE et des exigences de sécurité, de confidentialité et d'explicabilité,
- La volumétrie utile à traiter, tant en nombre de documents qu'en volume de données, ainsi que les critères de sélection des documents réellement concernés par les traitements IA. Pour indication, pour un fonds d'environ 100 000 documents, le temps de traitement ne doit pas excéder une durée de 8h00,
- La définition définitive des catégories de données sensibles, qui devra être validée conjointement avec les représentants métier, archivistiques et juridiques,
- Les modalités de mise à disposition et de transfert des données, incluant les formats d'échange, les mécanismes de chargement des fonds, les règles de préparation documentaire et les éventuels flux entre systèmes.

Ces éléments présentent aujourd'hui un niveau d'incertitude qui ne permet pas de figer définitivement la cible technique ou opérationnelle. Ils justifient la réalisation d'une phase de cadrage incluse dans le marché et destinée à consolider les hypothèses de travail, affiner le périmètre réel des traitements et sécuriser les choix de conception avant le déploiement des premiers cas d'usage.

3.4.5 Extension et mutualisation

La solution devra être conçue comme un socle mutualisable susceptible d'accueillir, à terme, de nouveaux cas d'usage au-delà des seuls périmètres *Thematic DataQuest* et *Sensitive DataQuest*. Ce projet doit répondre à un objectif de scalabilité.

Au-delà des besoins actuellement identifiés pour les archives, la plateforme devra être en mesure de supporter :

- L'augmentation progressive des volumes documentaires,
- L'intégration de nouveaux fonds,
- L'ajout de nouveaux traitements,
- L'apparition de nouveaux cas d'usage métiers,
- L'extension à d'autres environnements ou domaines fonctionnels de la CHD et autres acteurs de l'État luxembourgeois.

L'investissement réalisé dans cette plateforme devra ainsi permettre de constituer un socle réutilisable pour les futures initiatives d'IA documentaire de la CHD, tout en conservant la

souplesse nécessaire pour accompagner les évolutions technologiques et les orientations de l'État luxembourgeois.

4. Processus cible et exigences fonctionnelles

4.1 Principes du processus cible

4.1.1 Vision de bout en bout

Le processus cible repose sur une chaîne de traitement combinant des traitements automatisés par l'IA, des contrôles obligatoires et des validations humaines, depuis la préparation des fonds jusqu'à la restitution des résultats dans les outils de travail des archivistes.

La chaîne de traitement débute par la réception et la préparation des fonds documentaires, incluant le dépôt des données, les contrôles de recevabilité, les prétraitements nécessaires (décompression, OCR, extraction de texte, normalisation, etc.) ainsi que la prise en charge des différents formats autorisés.

Une fois ces contrôles réalisés, les documents sont soumis aux traitements automatisés d'IA qui assurent l'analyse des contenus textuels et des documents numérisés, l'extraction et l'enrichissement des métadonnées, la génération des descriptions archivistiques ainsi que l'identification et la qualification des données sensibles. Les résultats sont ensuite consolidés au niveau du dossier afin de produire une vision cohérente de l'ensemble documentaire.

Dans un troisième temps, les propositions générées par l'IA sont soumises à une validation humaine systématique selon le principe du *Human in the Loop*. Les archivistes peuvent alors contrôler, corriger, compléter ou rejeter les résultats, en particulier lorsque les niveaux de confiance sont insuffisants ou que des anomalies sont détectées. Après validation, les informations sont restituées dans la matrice d'inventaire Excel et accompagnées des éléments de traçabilité, d'explicabilité et des rapports de traitement nécessaires au suivi et à l'audit des opérations réalisées.

4.1.2 Cadre de description des étapes

Afin d'assurer une compréhension homogène des processus et de faciliter l'analyse des propositions des soumissionnaires, chaque étape de la chaîne de traitement sera décrite selon une grille commune. Cette approche permettra d'identifier de manière claire les responsabilités, les données manipulées, les contrôles réalisés et les résultats attendus.

Pour chaque étape, les éléments suivants seront précisés :

- Objectif : finalité métier de l'étape et valeur apportée au processus,
- Déclencheur : événement ou condition initiant l'exécution de l'étape,
- Entrées : données, documents, métadonnées ou paramètres nécessaires au traitement,
- Traitements : opérations réalisées, qu'elles soient automatisées ou assistées par l'IA ou non,
- Résultats : informations, métadonnées, alertes ou livrables produits,
- Intervention humaine : actions de contrôle, de validation, de correction ou de décision réalisées par les utilisateurs,
- Exceptions : cas particuliers, anomalies ou situations nécessitant un traitement spécifique,
- Traçabilité : informations conservées pour assurer le suivi, l'auditabilité des opérations,
- Critères de réussite : conditions permettant de considérer l'étape comme correctement exécutée et validée.

4.1.3 Constitution d'un projet de traitement

La solution devra permettre la création d'un projet de traitement correspondant à un périmètre documentaire défini par l'utilisateur. Chaque projet regroupera au minimum le corpus et la matrice d'inventaire concernés, un profil de configuration versionné, les règles et référentiels applicables, les modèles et prompts autorisés, les seuils de confiance, le workflow, les jeux de tests associés, les droits d'accès, ainsi que les modalités de restitution et de validation.

Lors de la création du projet, l'utilisateur devra pouvoir sélectionner les répertoires et documents concernés, renseigner les informations de contexte utiles, choisir le processus métier applicable et appliquer un profil de configuration prédéfini. Ce profil pourra être dupliqué, adapté et enregistré comme modèle réutilisable pour d'autres projets, y compris pour tout processus ultérieurement intégré à la plateforme.

L'utilisateur devra pouvoir définir ou surcharger, selon ses habilitations, les paramètres d'exécution : périmètre analysé, règles et guides rédactionnels, catégories de données sensibles, seuils de confiance, modalités de validation humaine, traitement des exceptions et formats de restitution. Une prévisualisation des impacts et un contrôle de cohérence devront précéder la publication d'une configuration.

Chaque projet devra disposer d'un identifiant unique, d'un statut, d'une version de configuration et d'un historique permettant de tracer les traitements, tests, validations, corrections, résultats et publications tout au long de son cycle de vie. Il devra être possible de cloner, archiver, exporter, réimporter et rejouer un projet dans des conditions reproductibles.

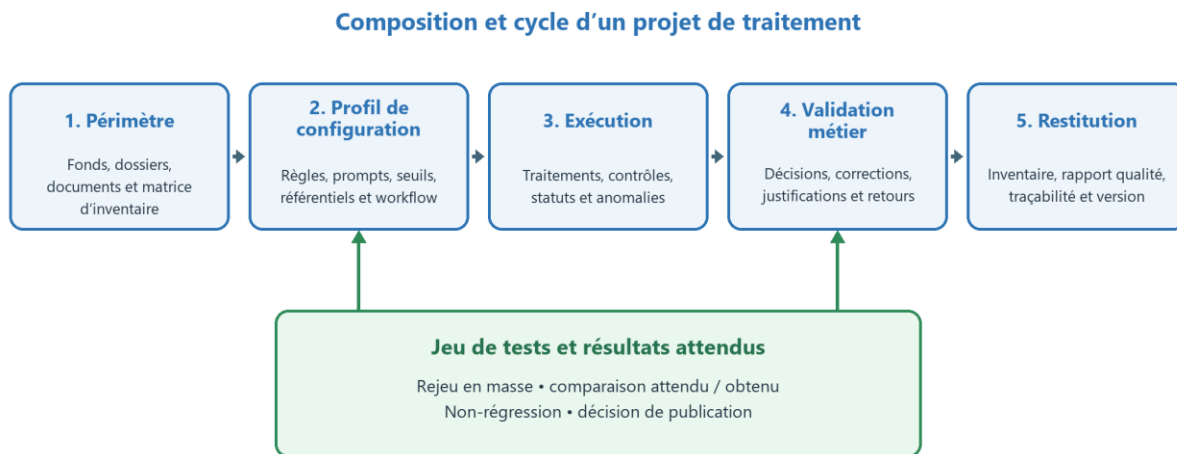


Figure - Composition et cycle d'un projet de traitement

4.1.4 Réception et préparation des données

4.1.4.1 Dépôt des données

La solution devra permettre le dépôt de la matrice d'inventaire Excel et des répertoires documentaires associés au sein d'un plan de classement constituant le périmètre d'analyse. Ces données pourront correspondre à un fonds, un versement, un ensemble de dossiers ou une campagne de traitement définie par les archivistes.

4.1.4.2 Contrôle de recevabilité

Avant tout lancement des traitements d'analyse, la solution devra réaliser des contrôles automatiques de recevabilité visant à vérifier la qualité, l'intégrité et l'exploitabilité des données fournies.

Ces contrôles devront notamment porter sur :

- La conformité des formats de fichiers par rapport à la liste des formats pris en charge,
- La taille des fichiers et la détection d'éventuelles anomalies de volumétrie,
- La détection des doublons,
- L'identification des fichiers ou dossiers manquants par rapport à l'inventaire,
- La présence d'archives compressées (ZIP ou équivalent) non préalablement traitées,
- L'identification des documents corrompus, incomplets ou illisibles,
- La qualité des documents numérisés et des résultats OCR lorsqu'ils sont existants.

À l'issue de ces vérifications, la solution devra produire un rapport de recevabilité présentant les contrôles effectués, les anomalies détectées, leur niveau de criticité ainsi que les éventuelles actions correctives à réaliser. Ce rapport constituera un point de contrôle obligatoire avant le démarrage des traitements d'IA et permettra aux archivistes de décider du lancement, du report ou de l'exclusion de certains dossiers du périmètre analysé.

4.1.4.3 Prétraitement des documents

La solution devra assurer les opérations de préparation documentaire nécessaires à l'analyse des fonds. Cette phase comprendra :

- L'extraction de texte,
- La mise en œuvre des traitements OCR pour les documents numérisés ou les images,
- La conversion des formats lorsque cela est nécessaire,
- La segmentation des contenus en unités exploitables (pages, sections, pièces ou dossiers),
- La lecture et l'exploitation des métadonnées techniques disponibles.

Les métadonnées techniques disponibles et complétées au préalable devront pouvoir être exploitées en complément de l'analyse du contenu.

L'ensemble de ces opérations devra garantir la préservation du lien entre les informations extraites et le document source. La solution devra ainsi être en mesure de conserver une référence vers le fichier d'origine, son emplacement, son identifiant et, lorsque cela est pertinent, la localisation des informations détectées au sein du document.

4.1.4.4 Gestion des différents formats

Le soumissionnaire devra préciser, pour chaque format pris en charge les traitements réalisables par la solution.

Le soumissionnaire devra également indiquer de manière détaillée les limites de gestion associées à chaque format, notamment les formats de fichiers pris en charge et non pris en charge, les limites de taille des fichiers, les limites de volume (nombre de pages, lignes, onglets, durée, résolution des images, qualité des scans, qualité audio ou vidéo, etc.), les contraintes liées à la qualité des contenus (lisibilité des documents, etc.), la prise en charge des objets complexes tels que les macros, formulaires, objets embarqués, animations, liens externes, signatures électroniques ou contenus multimédias intégrés, les éventuelles limitations linguistiques ou fonctionnelles.

Le soumissionnaire précisera également les comportements de la solution lorsque ces limites sont atteintes.

4.2 Analyse et enrichissement par l'IA

4.2.1 Analyse des documents textuels

La solution devra intégrer des fonctions d'analyse sémantique, de classification documentaire, de résumé et d'extraction d'informations afin d'exploiter le contenu des documents et leur contexte de classement.

L'analyse sémantique devra permettre d'interpréter le sens des contenus, d'identifier les thématiques traitées, les principaux concepts, les entités nommées, les événements, les dates et les informations significatives présentes dans les documents. Ces analyses devront tenir compte à la fois du contenu des fichiers, des métadonnées disponibles et du contexte apporté par le plan de classement et le dossier de rattachement.

Les capacités de résumé et de synthèse devront permettre de produire des descriptions courtes, pertinentes et compréhensibles du contenu des dossiers.

Enfin, les mécanismes d'extraction d'informations devront identifier et structurer les données utiles telles que les titres, dates, objets, acteurs, organismes, mots-clés, informations de contexte ou critères de sensibilité pertinents.

4.2.2 Analyse des images et documents numérisés

La solution devra être capable de traiter les images, documents numérisés et fichiers ne contenant pas de couche texte exploitable au moyen de mécanismes de reconnaissance optique de caractères (OCR). Les traitements devront permettre d'extraire le contenu textuel des documents scannés, qu'ils soient composés de pages isolées ou de dossiers complets. L'analyse devra prendre en charge les principaux formats image retenus dans le périmètre du projet.

La solution devra restituer le texte extrait de manière structurée afin qu'il puisse être utilisé par les traitements d'analyse sémantique, de génération de métadonnées, de classification documentaire et de détection des données sensibles. Les informations extraites devront conserver un lien avec le document source

Pour chaque document traité, la solution devra restituer un indice de confiance relatif à la qualité de la reconnaissance réalisée. Cet indicateur devra pouvoir être exploité dans les mécanismes de contrôle qualité, les seuils de validation et les processus de *Human in the Loop*. Les documents dont le niveau de confiance est inférieur aux seuils définis devront pouvoir être isolés et orientés vers un traitement ou une vérification manuelle avant toute exploitation des résultats.

4.2.3 Génération et enrichissement des métadonnées

La solution devra être capable de générer, et compléter les métadonnées descriptives et de gestion à partir du contenu des documents, des métadonnées existantes et du contexte apporté par le plan de classement.

Les traitements devront notamment permettre de proposer :

- Le titre du dossier,
- La description archivistique,
- La période de création des documents (date de début et date de fin).

Les propositions devront être produites sous une forme structurée et directement exploitable dans la matrice d'inventaire. La solution devra également être capable d'enrichir uniquement les champs manquants ou, le cas échéant, de suggérer une amélioration des informations déjà présentes sans les remplacer automatiquement.

Afin de garantir la traçabilité des traitements, chaque valeur devra être associée à son origine. Il devra ainsi être possible de distinguer clairement :

- Les données initialement présentes dans l'inventaire,
- Les valeurs générées automatiquement par l'IA,
- Les valeurs modifiées ou validées manuellement par un utilisateur.

Cette distinction devra être conservée tout au long du cycle de traitement afin de permettre les opérations de contrôle, de validation, d'audit et de justification des résultats.

4.2.4 Classification et description thématique

La solution devra être capable d'analyser le contenu des documents et des dossiers afin de proposer automatiquement des thèmes, catégories, mots-clés, typologies documentaires et éléments de description archivistique pertinents. Ces propositions pourront être utilisées pour alimenter les champs de l'inventaire, faciliter les opérations de recherche, améliorer la cohérence des descriptions et assister les archivistes dans leurs travaux de qualification documentaire.

Afin de garantir l'adaptabilité et la pérennité de la solution, les mécanismes de génération et de classification ne devront pas être figés. Les référentiels, règles métier, prompts, paramètres de traitement et modèles d'IA devront être configurables afin de pouvoir évoluer en fonction des besoins de la CHD.

Ces différents éléments devront être versionnés afin d'assurer la traçabilité des résultats produits, de permettre la reproductibilité des traitements et de conserver l'historique des évolutions apportées aux règles de classification ou aux modèles utilisés. Il devra ainsi être possible d'identifier la version du référentiel, du prompt ou du modèle ayant contribué à la production d'un résultat donné.

4.2.5 Détection des données sensibles

La solution devra être capable d'identifier automatiquement les informations relevant des catégories de données sensibles définies par la CHD. Cette détection devra s'appuyer sur l'analyse du contenu des documents, des métadonnées disponibles, du contexte documentaire et, lorsque cela est pertinent, des caractéristiques de forme des documents (formulaires, cachets, logos, signatures, modèles documentaires, etc.).

Pour chaque occurrence détectée, la solution devra être en mesure de :

- Qualifier la catégorie de sensibilité concernée,
- Localiser précisément l'information détectée dans le document, la page ou la zone concernée,
- Restituer le passage, l'extrait ou l'élément justificatif ayant conduit à la détection,
- Attribuer un niveau de confiance à la qualification proposée,
- Fournir les éléments d'explication nécessaires à la compréhension du résultat.

Les résultats devront être restitués dans la matrice d'inventaire. Les justifications devront pouvoir être restituées dans une forme structurée et exploitable par les archivistes, afin de faciliter les opérations de contrôle et de validation.

La qualification proposée par l'IA ne pourra pas être considérée comme définitive sans validation humaine. Les archivistes devront pouvoir consulter les occurrences détectées, confirmer ou corriger la catégorie proposée, modifier le niveau de qualification retenu et documenter, le cas échéant, la décision finale associée aux règles de communicabilité ou de restriction d'accès.

4.2.6 Consolidation au niveau du dossier

La solution devra être capable de consolider les résultats obtenus au niveau des différentes pièces afin de produire une qualification cohérente à l'échelle du dossier. Cette consolidation devra s'appliquer à l'ensemble des résultats trouvés.

Les règles de consolidation devront également permettre de gérer les situations d'incohérence ou de contradiction, par exemple lorsque plusieurs documents présentent des dates divergentes, des qualifications différentes ou des niveaux de sensibilité distincts. Dans ces cas, la solution devra signaler les écarts détectés, conserver les éléments justificatifs associés et proposer un résultat pertinent ainsi qu'une synthèse argumentée destinée à faciliter la décision des archivistes.

Les résultats consolidés devront être accompagnés d'indicateurs de confiance, des informations ayant contribué à leur élaboration et des éléments nécessaires à leur validation.

4.3 Validation humaine et restitution

4.3.1 Human in the loop

La solution devra intégrer des mécanismes de validation humaine aux étapes présentant un impact sur la qualité des inventaires et la qualification documentaire. Les processus de génération de métadonnées, d'identification des données sensibles et de consolidation des résultats au niveau du dossier devront notamment pouvoir faire l'objet d'un contrôle par les archivistes avant toute validation définitive.

Pour chaque proposition produite par la solution, l'archiviste devra disposer des éléments justificatifs nécessaires à sa compréhension et à son évaluation. Il devra être en mesure :

- D'accepter la proposition,
- De la modifier ou la compléter,
- De la rejeter.
- De saisir une valeur alternative lorsque cela est nécessaire.

Les mécanismes de validation devront être particulièrement mobilisés dans les situations présentant un faible niveau de confiance, des incohérences documentaires, des informations contradictoires ou des enjeux de sensibilité des données.

La solution devra conserver l'historique des actions réalisées ainsi que la trace des décisions prises. Il devra notamment être possible d'identifier :

- La proposition initialement générée,
- La décision prise par l'utilisateur,
- Les éventuelles modifications apportées,
- L'auteur de la validation,
- La date et l'heure de l'intervention,
- Les commentaires ou justifications associés.

Ces informations devront être conservées afin de garantir la traçabilité, l'auditabilité et la reproductibilité des traitements réalisés.

4.3.2 Boucle d'amélioration contrôlée et jeux de tests

La solution devra permettre aux administrateurs métier de constituer et gérer des jeux de données de référence comprenant les documents d'entrée, les résultats attendus, les cas négatifs, les cas limites et les décisions de validation. Ces jeux seront versionnés, rattachés à un cas d'usage et pourront être exécutés en masse à la demande.

L'interface devra comparer automatiquement les résultats obtenus aux résultats attendus, mettre en évidence les écarts, calculer les indicateurs de qualité applicables et permettre à l'administrateur de corriger ou compléter la vérité terrain. Toute modification d'une règle, d'un prompt, d'un seuil, d'un référentiel, d'un workflow ou d'un modèle devra pouvoir déclencher le rejeu des tests pertinents avant publication.

Les corrections réalisées par les archivistes pourront alimenter une amélioration contrôlée des traitements sous forme d'exemples, de règles, d'instructions mais ne doivent en aucun cas servir d'entraînement ou de réentraînement au modèle. Aucune correction utilisateur ne devra modifier automatiquement un modèle ou un comportement en production : la proposition d'amélioration devra être explicite, versionnée, testée, approuvée et réversible.

La boucle d'amélioration devra permettre à la CHD, sans assistance du titulaire :

- De créer, importer, exporter et versionner des jeux de tests et leurs résultats attendus,
- De rejouer tout ou partie des tests en lot et d'analyser les écarts par cas d'usage, catégorie, langue, format, fonds et niveau de confiance,
- D'ajuster les règles, prompts, exemples, seuils et profils de configuration dans un environnement de test,
- De soumettre une version à validation, de la publier en production et de revenir à la version précédente,
- De conserver la traçabilité de l'auteur, de la justification, des tests exécutés, des résultats obtenus et de la décision de publication.

4.3.3 Seuils de confiance

La solution devra intégrer des seuils de confiance configurables permettant de distinguer les résultats exploitables automatiquement, ceux nécessitant une validation humaine et ceux devant être considérés comme insuffisamment fiables pour être utilisés.

Trois niveaux minimums sont attendus :

- Résultats acceptables : niveau de confiance élevé permettant de proposer directement la valeur avec une validation simplifiée,
- Résultats à contrôler : niveau de confiance intermédiaire nécessitant une revue et une validation explicite par un archiviste avant toute utilisation,
- Résultats à rejeter ou à retraiter : niveau de confiance faible ne permettant pas de retenir la proposition sans analyse complémentaire ou intervention manuelle,

Les seuils retenus devront être paramétrables et pouvoir varier selon plusieurs critères, notamment :

- Le cas d'usage concerné (*Thematic DataQuest*, *Sensitive DataQuest* ou autre),
- La catégorie d'information traitée,
- Le niveau de sensibilité des données,
- Le type de métadonnée générée,
- Le format documentaire analysé,
- La qualité de l'OCR,
- Le fonds ou la série documentaire concerné.

La solution devra permettre d'adapter ces seuils au fil du temps en fonction des retours d'expérience et des exigences métier. Les niveaux de confiance utilisés, ainsi que les décisions prises à partir de ceux-ci, devront être conservés afin d'assurer la traçabilité et l'auditabilité des traitements réalisés.

4.3.4 Gestion des anomalies et exceptions

La solution devra intégrer des mécanismes permettant de gérer les anomalies susceptibles d'intervenir au cours des traitements, sans interrompre l'ensemble du processus. Les cas suivants devront notamment être pris en charge :

- Fichiers dans des formats non supportés,
- Documents corrompus ou inexploitable,
- Erreurs d'extraction de contenu,
- Métadonnées incohérentes,
- Réponses incomplètes ou non exploitables produites par les traitements IA,
- Indisponibilité temporaire d'un modèle ou d'un service externe,
- Dépassement des délais de traitement ou des temps de réponse attendus,
- Erreurs techniques survenues lors d'une étape de la chaîne de traitement.

La solution devra être capable d'isoler les éléments en anomalie, de journaliser les erreurs rencontrées et de produire des rapports permettant d'en comprendre l'origine et l'impact. Les dossiers ou documents concernés pourront être orientés vers une validation, une correction ou un traitement manuel lorsque cela est nécessaire.

Afin de faciliter l'exploitation opérationnelle, la solution devra permettre de relancer uniquement les traitements en échec ou les dossiers concernés par une anomalie, sans nécessiter la réexécution complète des traitements sur l'ensemble du fond. Des mécanismes de reprise ciblée et de suivi des statuts devront ainsi être prévus afin d'optimiser les temps de traitement et de limiter les impacts des erreurs sur les campagnes en cours.

4.3.5 Restitution dans l'inventaire Excel

La solution devra permettre l'alimentation automatique de la matrice d'inventaire Excel avec l'ensemble des informations produites et validées au cours des traitements. Les résultats restitués comprendront les métadonnées descriptives générées ou enrichies, les catégories de données sensibles identifiées, les statuts de traitement, les niveaux de confiance, les alertes détectées ainsi que les résultats des validations réalisées par les archivistes.

La restitution devra respecter la structure de l'inventaire définie par la CHD et permettre l'alimentation des colonnes existantes sans modification du format de référence. Les informations générées par l'IA devront pouvoir être distinguées des données d'origine et des valeurs modifiées manuellement afin de garantir la traçabilité des interventions réalisées.

La solution devra également préserver l'intégrité des données initialement présentes dans l'inventaire. Les informations existantes ne devront pas être écrasées sans règle explicite ou validation humaine préalable. Les résultats des traitements, les corrections apportées par les utilisateurs, les statuts de validation et les éventuelles alertes devront être conservés et restitués de manière cohérente afin de faciliter les opérations de contrôle et de suivi.

4.3.6 Rapports de traitement

La solution devra produire des rapports de traitement permettant de suivre les campagnes réalisées et d'en contrôler la qualité. Ces rapports devront présenter a minima :

- Le périmètre traité (fonds, dossiers, volumes et formats analysés),
- Les fichiers rejetés, en erreur ou non pris en charge,
- Les métadonnées générées et les résultats obtenus,
- Les validations, corrections et rejets réalisés par les archivistes,
- Les indicateurs de qualité et de confiance associés aux traitements,
- Les anomalies détectées et les éventuelles reprises effectuées.

Ces rapports devront être exploitables dans le cadre du pilotage opérationnel, de la recette de la solution et des audits ultérieurs.

4.3.7 Traçabilité et explicabilité

La solution devra garantir un niveau élevé de traçabilité et d'explicabilité des traitements réalisés afin de permettre la compréhension, la justification et l'audit des résultats produits.

Pour chaque traitement, la solution devra, a minima, conserver les informations suivantes :

- Le document ou le dossier ayant servi de base à l'analyse,
- Le traitement ou la chaîne de traitement exécutée,
- Le modèle ou service utilisé,
- La version du modèle, des référentiels ou des règles applicables,
- La date et l'heure de traitement,
- Le niveau de confiance associé au résultat,
- Les éventuelles alertes ou anomalies détectées,
- Les validations, corrections ou rejets réalisés par les utilisateurs.

Le niveau de conservation des prompts, paramètres d'exécution et éléments de contexte utilisés par les modèles d'IA sera défini lors de la phase de conception, en fonction des exigences de sécurité, de confidentialité, d'audit et de conformité applicables à la CHD.

4.4 Configuration et administration de la solution

4.4.1 Orchestrer les cas d'usages métiers de la CHD

La solution recherchée devra s'appuyer sur un outil basé sur des technologies d'IA permettant d'orchestrer les différents cas d'usage métiers de la CHD. Cette solution devra agir comme une couche de pilotage et de coordination des traitements documentaires, en permettant de configurer, exécuter et superviser des chaînes de traitement composées de plusieurs étapes successives.

La solution devra également permettre de définir des règles d'enchaînement conditionnelles (par exemple en fonction du type de document, du niveau de confiance obtenu, des catégories détectées ou des résultats des contrôles réalisés). Elle devra être en mesure de déclencher des validations humaines, des reprises ciblées ou des traitements complémentaires lorsque les conditions définies sont réunies.

L'outil devra pouvoir piloter l'appel à différents services ou modèles d'IA sans dépendance à une technologie unique. Les mécanismes d'orchestration devront permettre d'intégrer et de remplacer facilement les composants utilisés au fil de l'évolution des besoins ou des orientations technologiques de la CHD, du CTIE ou d'autres acteurs de l'État luxembourgeois.

L'architecture retenue devra privilégier la modularité, l'interopérabilité et le découplage des composants. Chaque fonction devra pouvoir évoluer indépendamment des autres afin d'éviter les développements spécifiques lourds lors de l'ajout de nouveaux traitements, de nouveaux modèles ou de nouveaux cas d'usage.

4.4.2 Administration fonctionnelle

La solution devra proposer une administration fonctionnelle centralisée, ergonomique et adaptée à des administrateurs métier non-développeurs. Elle devra permettre de définir, créer, dupliquer, modifier, importer, exporter, activer, désactiver et archiver les catégories de données sensibles, référentiels documentaires et archivistiques, règles métier, guides rédactionnels, règles de consolidation, prompts, exemples, seuils de confiance, paramètres de validation, profils de projet, workflows et versions des modèles et services mobilisés.

L'ensemble de ces éléments devra être configurable sans modification du code, versionné et réversible. L'administrateur devra pouvoir préparer une évolution dans un environnement isolé, en visualiser les impacts, exécuter les jeux de tests et contrôles de non-régression, soumettre la version à approbation, planifier son activation et revenir à une version antérieure. La solution devra identifier précisément la configuration ayant contribué à chaque résultat.

Les droits d'administration devront être distincts des droits des utilisateurs métiers :

- Les administrateurs fonctionnels habilités devront pouvoir gérer l'intégralité du cycle de vie des paramètres métier, profils, règles, prompts, référentiels, seuils, jeux de tests et workflows, y compris leur simulation, leur validation, leur publication et leur retour arrière
- Les utilisateurs métiers disposeront uniquement des droits nécessaires à l'exécution des traitements, à la consultation des résultats et aux opérations de validation

Cette séparation des rôles devra contribuer à la sécurité et à la gouvernance des traitements d'IA sans limiter l'autonomie opérationnelle de la CHD. L'ergonomie, l'aide contextuelle, les contrôles de cohérence et la documentation devront permettre à un administrateur fonctionnel formé d'effectuer les opérations courantes sans intervention du titulaire.

Critère d'autonomie. Lors de la recette finale, la CHD réalisera elle-même un scénario complet comprenant la modification d'une règle existante, la création d'une nouvelle règle, son affectation à un profil de projet, le rejeu des tests, la publication de la configuration et le retour à la version précédente. La réussite de ce scénario, sans action du titulaire ni modification du code, conditionnera l'admission.

Principe de facilité d'usage. Les opérations d'administration courantes devront être réalisables au moyen d'une interface graphique ou de formulaires guidés, avec des libellés compréhensibles, une validation des saisies, une prévisualisation, des messages d'erreur explicites et une aide en français. Les opérations de masse devront pouvoir être réalisées par import/export ou application de profils prédéfinis.

4.4.3 Gestion multi-LLM

La solution devra permettre l'utilisation de plusieurs LLM et la sélection du modèle le plus adapté à chaque traitement. Elle devra permettre de changer de modèle sans refonte des processus, et d'enchaîner plusieurs traitements successifs au sein d'un même workflow.

La plateforme devra également intégrer des mécanismes de repli permettant, en cas d'indisponibilité d'un modèle ou d'un fournisseur, de basculer automatiquement vers un modèle alternatif autorisé. Les règles de sélection, de bascule et les modèles effectivement utilisés devront être configurables et tracés afin de garantir la continuité de service, la transparence et l'auditabilité des traitements.

5. Architecture, sécurité et exigences non fonctionnelles

5.1 Architecture et infrastructure

5.1.1 Principes d'architecture

La solution devra s'appuyer sur une architecture modulaire et évolutive, fondée sur la séparation des composants fonctionnels et techniques. Chaque brique devra pouvoir être déployée, mise à jour, remplacée ou enrichie indépendamment des autres afin de faciliter les évolutions fonctionnelles et technologiques du dispositif.

L'architecture devra privilégier l'utilisation d'interfaces standardisées et documentées garantissant l'interopérabilité entre les différents composants, services et modèles d'IA. Les choix

techniques devront limiter les dépendances à une technologie ou un fournisseur spécifique afin de réduire les risques de verrouillage technologique et de préserver la réversibilité de la solution.

La solution devra permettre l'intégration progressive de nouveaux fonds documentaires, processus métiers, services, utilisateurs ou modèles d'IA sans remise en cause de l'architecture existante. Les composants retenus devront pouvoir être combinés ou remplacés afin d'accompagner l'évolution des besoins de la CHD.

L'architecture de référence présentée dans le marché constitue un cadre directeur. Son organisation détaillée, le découpage des composants ainsi que les mécanismes d'intégration seront consolidés lors de la phase de cadrage, sur la base des propositions du titulaire et des contraintes techniques et fonctionnelles de la CHD.

5.1.2 Architecture fonctionnelle cible

La solution devra s'appuyer sur une architecture fonctionnelle couvrant l'ensemble du cycle de traitement des données et des documents :

- Prétraitement des données,
- Orchestration des traitements,
- Appel aux LLM,
- Validation humaine.

La solution devra fournir des interfaces de restitution adaptées aux différents profils d'utilisateurs, permettant la consultation, l'analyse, la validation et l'export des résultats. Elle devra également inclure des fonctions d'administration et de supervision couvrant la gestion des utilisateurs, des droits d'accès, des traitements, des modèles d'IA, des journaux d'activité, des performances et du suivi opérationnel de la plateforme.

L'ensemble des flux de données, des interfaces et des échanges entre les différentes briques devra être documenté par le titulaire dans le cadre de la phase de cadrage et mis à jour tout au long de la vie de la solution. Cette documentation devra préciser les responsabilités de chaque composant, les mécanismes d'intégration retenus, les flux entrants et sortants ainsi que les contrôles de sécurité associés.

5.1.3 Scénarios d'architecture

La solution devra pouvoir s'intégrer dans différents scénarios d'hébergement, en fonction des choix retenus par la CHD. Le titulaire devra analyser et documenter les scénarios envisageables selon que l'infrastructure, les modèles et la plateforme sont fournis par le CTIE, la CHD ou un opérateur souverain.

La réponse devra notamment couvrir les cas suivants :

- Infrastructure et service d'IA hébergés par le CTIE,
- Infrastructure et services d'IA hébergés et exploités par la CHD,
- Infrastructure ou services fournis par un opérateur souverain retenu par la CHD.

Au stade de l'offre, le soumissionnaire décrira, sur la base des informations disponibles et des hypothèses qu'il explicitera, la compatibilité de sa solution avec les différents scénarios d'hébergement.

Pour le scénario CTIE, l'analyse détaillée, le dimensionnement et les modalités d'intégration seront consolidés pendant la phase de cadrage avec la CHD et le CTIE.

Pour les scénarios CHD, le soumissionnaire devra préciser les modalités d'intégration, les prérequis techniques, les contraintes d'exploitation, les mécanismes de sécurité, les conditions de réversibilité ainsi que les impacts éventuels sur les performances, les coûts et l'évolutivité de

la solution. Il devra en outre préciser, le cas échéant, quel scénario reste impossible à réaliser avec la solution qu'il propose. Le prestataire s'attachera à détailler son architecture cible en matière de puissance de calcul attendue, de type de cœur utilisée, de mémoire, l'interconnectivité réseau, stockage et IOPS attendus. Une estimation budgétaire de l'hébergement est attendue dans le cadre du présent appel d'offre.

En début de projet, la CHD documentera les responsabilités des différents acteurs qui devront être clairement identifiés, notamment concernant l'hébergement, l'administration des infrastructures, l'exploitation de la plateforme, la gestion des modèles d'IA, la sécurité, la gestion des accès, la supervision, la maintenance, le support et le respect des exigences réglementaires applicables. Une matrice de responsabilités de type RACI sera proposée par la CHD afin de formaliser le partage des rôles entre le titulaire, la CHD, le CTIE et les éventuels opérateurs tiers. Cette tâche commencera après le cadrage du projet et se poursuivra jusqu'au terme du projet.

5.1.4 Hébergement et souveraineté

L'ensemble des données traitées dans le cadre de la solution, y compris les documents sources, les métadonnées, les index, les résultats des traitements, les sauvegardes, les fichiers temporaires et les journaux techniques, devra être hébergé et conservé dans des environnements préalablement validés par la CHD et conformes à ses exigences de souveraineté et de sécurité.

L'ensemble des traitements, les mécanismes d'orchestration ainsi que les appels aux modèles d'IA devront être réalisés exclusivement au sein des infrastructures autorisées par la CHD. Aucun transfert de données, de documents ou de métadonnées vers un environnement, un service ou un fournisseur non expressément autorisé par la CHD ne devra être possible.

Le titulaire devra préciser la localisation géographique des infrastructures utilisées pour l'hébergement, les traitements, les sauvegardes, les mécanismes de reprise d'activité et la conservation des journaux. Il devra démontrer que l'ensemble des flux de données demeure maîtrisé et qu'aucune donnée n'est transférée, stockée, répliquée ou traitée en dehors des périmètres autorisés.

Le titulaire devra enfin garantir que les données de la CHD ne sont ni réutilisées pour l'apprentissage de modèles, ni accessibles à des acteurs non autorisés, et que l'ensemble des traitements respecte les exigences réglementaires, contractuelles et de sécurité applicables.

5.1.5 Infrastructure et environnements

La solution devra être déployée au minimum sur les environnements distincts suivants : développement, test, recette/préproduction et production. Chaque environnement devra répondre à un objectif spécifique du cycle de vie de la solution et être administré selon des règles garantissant la qualité, la sécurité et la maîtrise des mises en production.

Les différents environnements devront garantir la confidentialité des données, la stabilité des services et la maîtrise des changements. Aucun accès direct ou flux non maîtrisé entre les environnements ne devra être autorisé. Les données de production ne pourront être utilisées dans les environnements hors production qu'après accord de la CHD et application des mesures de protection appropriées.

Le titulaire devra proposer un dimensionnement adapté à chaque environnement en tenant compte des volumes documentaires, du nombre d'utilisateurs, des performances attendues et des besoins de montée en charge. Les configurations des environnements devront être documentées et maintenues cohérentes afin de limiter les écarts entre la préproduction et la production.

La solution devra permettre la reproductibilité des déploiements et des configurations au moyen de mécanismes industrialisés. Le titulaire devra décrire les procédures de déploiement, de mise à jour, de retour arrière et de gestion des versions afin de garantir la traçabilité et la fiabilité des mises en production.

5.1.6 Intégration au SI

La solution devra pouvoir s'intégrer aux services transverses mis à disposition par la CHD, le CTIE ou l'opérateur d'hébergement retenu. Cette intégration concernera notamment les services de gestion des identités et des accès, de supervision, de journalisation, de sauvegarde et d'exploitation.

Le titulaire devra décrire les modalités d'intégration de la solution avec ces services ainsi que les éventuelles contraintes associées.

5.1.7 Fonctionnement filesystem et interopérabilité

Dans sa version initiale (V1), la solution s'appuiera principalement sur des fichiers, des arborescences de répertoires (file system) et des inventaires sous format Excel fournis par la CHD. Les traitements devront pouvoir être réalisés à partir de ces sources sans nécessiter la mise en place d'interfaces spécifiques avec des applications tierces.

Toutefois, l'architecture devra prévoir des mécanismes d'intégration ouverts et documentés permettant, à terme, le raccordement de nouvelles sources de données, applications ou services. Les interfaces, protocoles et formats d'échange utilisés devront être documentés afin de faciliter de futures intégrations, sans que celles-ci ne constituent une exigence de mise en œuvre dans le cadre de la V1.

5.2 Architecture et gouvernance des modèles d'IA

5.2.1 Architecture multi-LLM

La solution devra permettre l'utilisation de plusieurs modèles d'IA (LLM) au sein d'une même architecture. Les mécanismes d'orchestration devront permettre de sélectionner le modèle le plus adapté en fonction du cas d'usage, de la nature des données traitées, des performances attendues, des contraintes de confidentialité ou des droits d'accès applicables.

La solution devra être en mesure d'exploiter différents types de modèles, qu'ils soient hébergés par la CHD, le CTIE, un opérateur souverain ou tout autre environnement autorisé. Elle devra permettre l'ajout, le remplacement ou la coexistence de plusieurs modèles sans modification majeure des processus métiers ni des composants applicatifs.

Le titulaire devra décrire les mécanismes de sélection, de routage et de gestion des modèles ainsi que les modalités permettant de faire évoluer le portefeuille de modèles utilisés au cours du marché, tout en garantissant la continuité de service et la maîtrise des données.

5.2.2 Utilisation du service Mistral

La solution devra pouvoir s'appuyer sur un service Mistral si le choix de ce LLM est retenu lors de la phase de cadrage du projet. Le titulaire devra démontrer sa capacité à intégrer ce service au sein de l'architecture proposée, sans remettre en cause les principes de modularité et de multi-LLM définis dans le présent marché.

Le cas échéant, le titulaire devra être en mesure d'identifier et documenter les prérequis techniques associés à l'utilisation du service Mistral, notamment les modalités d'accès, les contraintes d'intégration, les conditions d'exploitation, les niveaux de service attendus ainsi que

les éventuelles limitations fonctionnelles ou techniques susceptibles d'avoir un impact sur les cas d'usage de la CHD.

5.2.3 Gouvernance technique des modèles

La solution devra intégrer des mécanismes de gouvernance permettant de recenser et de gérer l'ensemble des modèles d'IA autorisés dans le périmètre de la CHD. L'identification, sa version, son fournisseur, ses cas d'usage autorisés, ses principaux paramètres de fonctionnement, ainsi que ses conditions d'hébergement et d'exploitation devront être documentés pour chaque modèle.

Le titulaire devra garantir la traçabilité des modèles mobilisés dans les différents traitements et permettre le suivi de leur cycle de vie, notamment lors des mises à jour, changements de version ou remplacements de modèles.

Tout changement significatif susceptible d'avoir un impact sur les performances, les résultats produits, la sécurité, la conformité ou les conditions de traitement des données devra faire l'objet d'une analyse préalable et pouvoir déclencher une nouvelle évaluation technique, fonctionnelle ou de sécurité avant sa mise en production.

5.2.4 Maîtrise des appels aux modèles

La solution devra intégrer des mécanismes de pilotage des appels aux modèles d'IA afin d'optimiser les performances, les coûts, la qualité des résultats et l'utilisation des ressources disponibles. Elle devra permettre le routage des requêtes vers le modèle le plus adapté en fonction de la nature du traitement, du niveau de complexité attendu, des exigences de confidentialité, des performances recherchées et des autorisations applicables. Le recours à un modèle donné devra rester proportionné au besoin métier traité.

La solution devra également mettre en œuvre des mécanismes de contrôle des consommations, de limitation des usages, de gestion des quotas et de suivi des coûts associés aux traitements réalisés. Des fonctions de supervision devront permettre de suivre les volumes de requêtes, les temps de réponse, les taux d'erreur et l'utilisation des différents modèles.

5.3 Protection des données et contrôle des accès

5.3.1 Cycle de vie des données

La CHD transmettra les informations sur le cycle de vie des données manipulées dans le cadre des traitements. Elle précisera les règles de conservation, d'archivage et de suppression applicables à chaque catégorie de données, ainsi que les mécanismes permettant d'assurer leur purge sécurisée lorsqu'elles ne sont plus nécessaires. L'outil devra mettre en œuvre une suppression définitive et intégrale des copies des données traitées une fois l'inventaire produit.

5.3.2 Confidentialité et non-réutilisation

Aucune donnée, aucun document, prompt, résultat ou métadonnée de la CHD ne pourra être utilisé pour entraîner, réentraîner, ajuster ou améliorer un modèle d'IA externe.

Cette interdiction s'applique à l'ensemble des services, plateformes, API, modèles et sous-traitants mobilisés dans le cadre du marché. Le titulaire devra documenter les garanties techniques, contractuelles et organisationnelles mises en œuvre pour assurer le respect de cette exigence. Toute évolution de ces garanties devra être soumise à validation préalable de la CHD.

5.3.3 Identités et habilitations

La solution devra s'intégrer aux mécanismes d'authentification de la CHD et prendre en charge, lorsque disponible, les dispositifs de fédération d'identité et d'authentification unique (SSO).

La solution devra permettre de distinguer au minimum les profils suivants :

- Utilisateurs : accès aux fonctionnalités métier et aux données relevant de leur périmètre d'habilitation,
- Administrateurs fonctionnels : gestion des paramètres fonctionnels, référentiels, workflows et habilitations métier,
- Exploitants : supervision des traitements, suivi opérationnel et gestion des incidents,
- Administrateurs techniques : administration de la plateforme, des composants techniques, des configurations et des interfaces.

Le titulaire devra décrire les mécanismes d'authentification, de gestion des rôles, d'attribution des droits et de traçabilité des actions réalisés au sein de la solution.

5.3.4 Chiffrement et gestion des clés

La solution devra assurer le chiffrement des données en transit et au repos conformément aux bonnes pratiques de sécurité en vigueur. Cette exigence s'applique notamment aux documents, métadonnées, index, journaux, sauvegardes et échanges entre composants de la plateforme.

Le titulaire devra préciser les mécanismes de gestion des clés de chiffrement ainsi que la répartition des responsabilités relatives à leur création, leur conservation, leur rotation, leur renouvellement et leur révocation. Les modalités retenues devront être compatibles avec les exigences de sécurité de la CHD et avec le scénario d'hébergement retenu.

5.3.5 Journalisation et audit

La solution devra assurer la traçabilité des opérations réalisées au sein de la plateforme. Les mécanismes de journalisation devront permettre d'enregistrer au minimum les événements relatifs aux authentifications et connexions, aux dépôts de données et de documents, aux traitements exécutés, aux consultations et modifications des informations, aux validations et décisions humaines ainsi qu'aux opérations d'administration et d'exploitation.

Les journaux devront être horodatés, protégés contre toute altération ou suppression non autorisée et conservés selon les règles de rétention définies par la CHD. Ils devront permettre la reconstitution des opérations réalisées à des fins d'audit, de sécurité et d'exploitation.

5.4 Cybersécurité et conformité

5.4.1 Sécurisation de la plateforme

La solution devra respecter les bonnes pratiques de sécurisation des systèmes et applications. Le titulaire devra mettre en œuvre les mesures de durcissement nécessaires. Les exigences de durcissement seront définies en fonction du type de plateforme concernée (Cloud ou On-Premise) : pour les systèmes nécessitant le déploiement d'un système d'exploitation (Windows, Linux, etc.), la référence de sécurité applicable sera le CIS Benchmark correspondant.

Le titulaire devra également définir les processus de veille de sécurité, de traitement des vulnérabilités et de maintien en condition de sécurité de la solution. Les exigences applicables seront précisées en référence aux politiques, standards et référentiels de sécurité de la CHD, du CTIE et de l'environnement d'hébergement retenu.

5.4.2 Sécurité de l'IA générative

La solution devra intégrer des mesures de sécurité et de qualité adaptées aux risques spécifiques des modèles d'IA : prompt injection, exfiltration ou réutilisation non autorisée de données, empoisonnement des contextes ou jeux de référence, contournement des règles, génération de contenus erronés, non étayés ou trompeurs (hallucinations), omission d'informations sensibles, biais de traitement et dérive de performance. Tout résultat généré devra être rattaché à des éléments de preuve issus des sources ; en l'absence de preuve suffisante, la solution devra s'abstenir, signaler l'incertitude et orienter le résultat vers une validation humaine.

Des mécanismes de filtrage, de contrôle de cohérence, de validation humaine, de tests adversariaux, de traçabilité, de supervision et de détection des dérives devront réduire ces risques. Aucun résultat à enjeu sensible ou non étayé ne pourra être intégré automatiquement dans l'inventaire. Les incidents de qualité devront être détectables, quantifiables, explicables et reproductibles à partir de la version du corpus, des règles, des prompts, des paramètres, du workflow et du modèle utilisés.

Les mesures retenues devront s'appuyer sur les bonnes pratiques reconnues du domaine, notamment les référentiels *OWASP Top 10 for LLM Applications*, *OWASP GenAI Security Project* et *OWASP Agentic Applications for 2026*, ainsi que *WFD (Westminster Foundation for Democracy) AI guidelines for parliaments*, ou leurs versions en vigueur au moment de la mise en œuvre de la solution.

5.4.3 Conformité réglementaire

La solution devra être conçue et exploitée dans le respect des exigences réglementaires et normatives applicables à la CHD, notamment en matière de protection des données personnelles, de cybersécurité et d'IA. À ce titre, le titulaire respectera les dispositions du RGPD, de l'*AI Act*, de la directive *NIS2*, ainsi que les règles, politiques et principes définis dans la Charte IA de la CHD.

Selon le périmètre retenu et les risques identifiés, le titulaire devra contribuer à la réalisation des analyses requises, notamment la qualification réglementaire de la solution, l'analyse des risques de sécurité et de conformité, l'analyse d'impact relative à la protection des données ainsi que toute documentation de conformité exigée par les textes applicables.

Toute évolution significative de la solution, des modèles utilisés ou des cas d'usage déployés devra pouvoir faire l'objet d'une réévaluation de la conformité et des analyses associées.

Projet #1165 - DataQuest

Service SSI - Exigences non fonctionnelles complémentaires

Catégorie	Exigence	Détails
MUST	La proposition doit permettre d'être en conformité avec les textes suivants.	- Réglementation IA Act : https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai - Loi NIS2 : https://www.ilr.lu/secteurs-activites/niss/nis-2/ - Charte IA CHD : https://www.chd.lu/sites/default/files/2024-07/Charte_IA_CHD.pdf
SHOULD	La proposition devrait tenir compte des guides de bonnes pratiques suivants.	OWASP (Open Worldwide Application Security Project) • GenAI Security Project : https://genai.owasp.org/ • Top 10 for LLM Applications 2025 : https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/ • Top 10 for Agentic Applications for 2026 : https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/ WFD (Westminster Foundation for Democracy) • AI guidelines for parliaments : https://www.wfd.org/ai-guidelines-parliaments

5.4.4 Développement et DevSecOps

Le standard de développement exigé est l'OWASP ASVS niveau 2. La solution devra être conçue, développée et exploitée conformément à ce standard.

Le titulaire devra mettre en œuvre des mécanismes garantissant la traçabilité, la sécurité et la maîtrise des livraisons. Les exigences applicables seront adaptées lors du cadrage en fonction de la nature réelle de la solution proposée, du niveau de développement spécifique requis et des composants effectivement mis en œuvre.

5.5 Exploitabilité et qualité de service

5.5.1 Performance et capacité

La solution devra être dimensionnée pour traiter un ensemble de 100 000 documents en huit heures au maximum sur le corpus de performance et l'infrastructure de référence validés lors du cadrage, hors délais de validation humaine. La mesure couvrira la chaîne automatisée depuis l'acceptation du l'ensemble jusqu'à la disponibilité des résultats, rapports et statuts. Le titulaire précisera et justifiera les hypothèses de taille, formats, langues, qualité OCR, complexité, concurrence d'usage et ressources nécessaires. Les essais devront mesurer le débit, le temps total et par étape, les files d'attente, les taux d'erreur et de reprise, la taille maximale des documents, le nombre d'utilisateurs simultanés et la capacité de montée en charge, sans dégradation des seuils de qualité définis au marché.

5.5.2 Disponibilité, résilience et continuité

La solution devra présenter un niveau de disponibilité et de résilience cohérent avec la criticité des services rendus à la CHD. Le titulaire devra dimensionner son outil afin de ne pas dépasser une interruption de service d'un maximum cumulé de trois jours ouvrés par mois et de restauration de la configuration de la solution (objectif de récupération du paramétrage de la solution) d'un maximum d'une journée ouvrée.

Les mécanismes techniques mis en œuvre devront limiter les interruptions de service et garantir la continuité des traitements en cas d'incident. La solution devra prévenir toute perte de données ou de traitement en cours et permettre la reprise des opérations dans des conditions maîtrisées.

5.5.3 Sauvegarde et restauration

La solution devra permettre la sauvegarde des données nécessaires à son fonctionnement et à la continuité d'activité, notamment les configurations, référentiels, métadonnées, index, journaux, données de traitement et tout autre composant requis pour la restauration du service.

Le titulaire devra préciser les données couvertes par les sauvegardes, les fréquences de sauvegarde, les durées de conservation ainsi que les mécanismes de restauration associés. Des tests de restauration devront être réalisés périodiquement afin de vérifier l'intégrité des sauvegardes et la capacité de reprise de la solution.

5.5.4 Supervision et observabilité

La solution devra fournir des capacités de supervision permettant de suivre l'état de fonctionnement des composants techniques, des traitements exécutés, des files d'attente, des flux de données, des appels aux modèles d'IA, des erreurs et de la consommation des ressources.

Le titulaire devra prévoir des mécanismes de surveillance, d'alerte et de diagnostic permettant d'identifier rapidement les incidents et d'en faciliter l'analyse. Des indicateurs d'exploitation et des tableaux de bord devront être disponibles pour le suivi opérationnel de la plateforme.

La supervision devra permettre de distinguer clairement les incidents liés :

- À l'application et à ses composants,
- Aux données, documents ou traitements métier,
- À l'infrastructure d'hébergement,
- Aux services ou modèles d'IA (LLM).

5.5.5 Ergonomie, langues et accessibilité

La solution devra proposer une interface utilisateur ergonomique, cohérente et adaptée aux différents profils. Sa prise en main devra être rapide pour les archivistes et les administrateurs fonctionnels : navigation simple, parcours guidés, vocabulaire métier, prévisualisation des actions, aide contextuelle, modèles de configuration et messages explicites permettant de comprendre et corriger les situations rencontrées. Les tâches courantes d'utilisation et de paramétrage ne devront pas nécessiter de compétence de développement.

La solution devra respecter les exigences d'accessibilité applicables aux outils numériques utilisés par la CHD. Le titulaire précisera le niveau de conformité atteint ainsi que les éventuelles limitations connues.

Le titulaire devra également préciser les langues disponibles pour les interfaces utilisateurs, les messages système, l'administration et la documentation (a minima : français et anglais).

5.5.6 Sobriété environnementale

La solution devra permettre de privilégier l'utilisation de modèles de taille et de capacité adaptées au besoin, afin d'éviter le recours systématique à des modèles surdimensionnés.

Des indicateurs de suivi devront permettre de mesurer l'utilisation des modèles, les volumes traités, les temps de réponse, les consommations de ressources et les coûts associés. La solution devra également permettre la mise en place de règles de contrôle, de limitation ou de routage afin d'optimiser l'usage des ressources et de garantir une utilisation proportionnée des différents modèles disponibles.

5.6 Pérennité de la solution

5.6.1 Licences et maîtrise des coûts

Le titulaire devra détailler l'ensemble des licences, souscriptions et droits d'usage nécessaires au fonctionnement de la solution. Cette description devra couvrir notamment la plateforme principale, les connecteurs, les modèles d'IA, les moteurs OCR ainsi que tout composant tiers requis.

Le titulaire devra également identifier de manière explicite les coûts récurrents liés à l'exploitation de la solution, notamment les coûts de maintenance, d'abonnement, d'hébergement, de support et d'évolution. Lorsque des modèles d'IA sont utilisés, les coûts d'inférence, de consommation ou d'usage devront être précisés, ainsi que les éventuelles limitations de volume, de quota, de capacité ou de performance applicables.

L'ensemble des hypothèses de dimensionnement et de consommation utilisées pour l'estimation des coûts devra être documenté afin de permettre à la CHD d'évaluer le coût complet de possession de la solution.

5.6.2 Portabilité et réversibilité

La solution devra garantir un niveau élevé de réversibilité et de portabilité afin de préserver la liberté de choix de la CHD tout au long du cycle de vie du marché. Le titulaire devra prévoir des mécanismes permettant l'export complet des configurations, workflows, prompts, référentiels, journaux, métadonnées, résultats de traitement et autres éléments nécessaires à l'exploitation de la solution, dans des formats ouverts, documentés et réutilisables.

Le titulaire devra décrire les modalités de réversibilité, les éléments exportables, les formats utilisés ainsi que les éventuelles contraintes ou dépendances résiduelles. Toute limitation à la portabilité ou à la réversibilité devra être explicitement identifiée dans l'offre.

5.6.3 Documentation technique

Le titulaire devra fournir et maintenir à jour l'ensemble de la documentation nécessaire à l'installation, à l'exploitation, à l'administration, à la sécurisation et à la réversibilité de la solution.

La documentation devra comprendre :

- Les dossiers d'architecture fonctionnelle et technique,
- Les dossiers d'installation et de configuration,
- Les procédures de déploiement et de mise à jour,
- Les procédures d'exploitation et d'administration,
- Les procédures de supervision et de gestion des incidents,
- Les procédures de sauvegarde et de restauration,
- Les procédures de continuité et de reprise d'activité,
- La documentation relative à la sécurité et à la gestion des habilitations,
- La documentation des interfaces, flux et mécanismes d'intégration,
- Les guides utilisateurs et administrateurs.

L'ensemble de cette documentation devra être remis dans des formats exploitables par la CHD. Toute évolution significative de la solution, de son architecture, de ses composants ou de ses procédures devra donner lieu à une mise à jour de la documentation correspondante pendant le projet.

6. BUILD : démarche, gouvernance, livrables et planning

6.1 Stratégie de réalisation

6.1.1 Démarche incrémentale

Le projet sera conduit selon une démarche incrémentale permettant d'obtenir rapidement des résultats concrets tout en maîtrisant progressivement la complexité des traitements et des fonds documentaires pris en charge. Chaque incrément devra produire un résultat démontrable, évaluable au regard d'indicateurs définis avec la CHD. L'exécution du marché s'articulera autour de cinq phases successives : une phase de cadrage suivie de quatre incréments de réalisation.

Chaque incrément devra produire un résultat démontrable, évaluable et potentiellement exploitable. Cette organisation permet de respecter le calendrier global du projet tout en conservant une logique de résultats progressifs. Elle vise à produire des bénéfices visibles dès les premiers mois, à sécuriser les choix techniques par des validations successives et à augmenter progressivement la valeur apportée aux métiers avant l'industrialisation complète de la solution.

6.1.2 Organisation Générale du BUILD

La phase BUILD sera structurée autour de quatre axes principaux : l'audit et le cadrage, la mise en place de la plateforme, l'intégration progressive des périmètres documentaires et la stabilisation finale de la solution. Afin de respecter les délais du projet, plusieurs activités pourront être menées en parallèle lorsque leur dépendance fonctionnelle le permet.

- Audit et cadrage. Cette phase permettra de constituer les jeux de référence, de définir les indicateurs de qualité et de valider l'architecture cible ainsi que les modalités d'exploitation de la solution.
- Mise en place de la plateforme. En parallèle du cadrage, le titulaire mettra en œuvre les environnements, les composants techniques, les mécanismes d'orchestration, les fonctions de supervision, les dispositifs de sécurité et les premières capacités d'intégration avec les modèles d'IA retenus.
- Incréments documentaires. Les traitements seront ensuite déployés progressivement sur des périmètres documentaires de complexité croissante, afin d'obtenir rapidement des résultats exploitables, de mesurer la qualité des traitements et d'ajuster les paramétrages avant l'extension aux fonds les plus complexes.
- Stabilisation et industrialisation. La phase finale aura pour objectif de consolider les performances, d'achever les recettes, de finaliser la documentation, de sécuriser l'exploitation et de préparer la généralisation de la solution sur l'ensemble du périmètre retenu.

Certaines activités de configuration, de préparation des données, de mise en place des environnements, de documentation et de conduite du changement pourront être réalisées en parallèle des incréments fonctionnels afin d'optimiser le calendrier global du projet.

6.1.3 Principes de progression

La mise en œuvre de la solution devra suivre une logique de progression maîtrisée, fondée sur l'intégration successive de périmètres documentaires de complexité croissante. Les premiers incréments devront privilégier les fonds les plus simples à traiter afin d'obtenir rapidement des résultats mesurables et d'ajuster les paramétrages de la solution.

L'élargissement du périmètre devra ensuite s'effectuer progressivement par l'intégration de nouveaux formats documentaires, de volumes plus importants, de langues supplémentaires ainsi que de documents présentant des niveaux de complexité ou de sensibilité plus élevés.

6.2 Initialisation et mise en place du socle

6.2.1 Cadrage technique et fonctionnel

Le projet débutera par une phase de cadrage technique et fonctionnel destinée à consolider la compréhension des besoins, à sécuriser les choix d'architecture et à préparer les conditions de mise en œuvre de la solution. Ce cadrage portera notamment sur les fonds documentaires concernés, les processus métier, l'architecture existante, les exigences de sécurité, les infrastructures et services mis à disposition par la CHD ou le CTIE, ainsi que les prérequis nécessaires au déploiement de la solution.

À l'issue du cadrage, les hypothèses, risques, contraintes et points d'arbitrage devront être formalisés. Les sujets restants devront être traduits en décisions, plans d'action, responsabilités et échéances, afin de disposer d'une trajectoire de mise en œuvre claire et partagée avant le démarrage des incréments de réalisation.

6.2.2 Finalisation de l'architecture

Au cours de la phase de cadrage, le titulaire devra élaborer puis faire valider par la CHD l'architecture cible de la solution, tant sur le plan fonctionnel que technique.

La CHD devra notamment confirmer les modalités d'hébergement retenues, les services et modèles d'IA mobilisés, les flux de données et d'échanges entre composants, l'organisation des environnements de développement, de test, de recette et de production, ainsi que les interfaces avec les services existants de la CHD et du CTIE.

La phase de cadrage devra également formaliser la répartition des responsabilités entre les différents acteurs concernant l'exploitation, l'administration, la supervision, la maintenance, la sécurité, les sauvegardes et le support de la solution. Les choix validés seront documentés dans les dossiers d'architecture générale et détaillée qui constitueront la référence pour la mise en œuvre des incréments suivants.

6.2.3 Acquisition des licences et installation

Pendant la phase de cadrage, le titulaire assurera la fourniture des licences nécessaires au fonctionnement de la solution, son installation complète sur les infrastructures retenues, sa configuration initiale ainsi que son paramétrage conformément aux besoins de la CHD.

La solution devra pouvoir être déployée « from scratch » sur les environnements prévus dans le cadre du marché. Le titulaire assurera également la connexion aux services autorisés, notamment les services d'identité, de supervision, de journalisation, de sauvegarde et les services d'IA retenus.

6.2.4 Mise en place du socle industriel

Pendant la phase de cadrage, le titulaire devra mettre en place l'ensemble des composants transverses nécessaires à l'exploitation de la solution, notamment les mécanismes de déploiement, de sauvegarde, de journalisation, de supervision et de sécurité.

Ce socle technique devra être configuré, documenté et opérationnel sur les environnements concernés avant toute ouverture d'un premier périmètre fonctionnel en production. Il constituera la base de référence pour le déploiement, l'exploitation, la supervision et le maintien en condition opérationnelle de la plateforme tout au long du projet.

6.3 Gouvernance du projet

6.3.1 Instances de gouvernance

Le titulaire devra mettre en place une gouvernance adaptée à la conduite du projet et à son suivi opérationnel. Celle-ci devra s'appuyer, a minima, sur :

- Des comités de pilotage,
- Des comités projet hebdomadaires pour assurer le suivi,
- Des ateliers de travail dédiés aux sujets fonctionnels et techniques.

Pour chaque instance, le titulaire précisera les participants, les responsabilités, les livrables attendus et la périodicité des réunions. Les modalités de suivi de l'avancement, de gestion des risques, de traitement des points bloquants et de validation des livrables devront être définies dès le début de la phase de cadrage.

6.3.2 Répartition des responsabilités

Le titulaire devra proposer et maintenir une matrice de responsabilités (RACI) couvrant l'ensemble des acteurs intervenant dans le projet et son exploitation. Cette matrice devra notamment inclure la CHD, le CTIE, le titulaire, le responsable de la sécurité des systèmes d'information (SSI), le délégué à la protection des données (DPO) ainsi que tout autre acteur concerné.

6.3.3 Gestion des risques et dépendances

Le titulaire devra mettre en place et maintenir un registre de pilotage recensant l'ensemble des risques, hypothèses, décisions, actions et dépendances identifiés tout au long du projet.

Ce registre devra permettre de suivre les risques, leur criticité, les mesures de mitigation associées, les responsables désignés ainsi que les échéances de traitement. Il devra également assurer la traçabilité des hypothèses retenues, des décisions prises et des actions engagées.

6.3.4 Gestion des évolutions et arbitrages

Le titulaire devra mettre en place un processus de gestion des évolutions permettant de qualifier, analyser, estimer, prioriser et soumettre à décision toute demande d'évolution de la solution.

Chaque demande devra faire l'objet d'une analyse de ses impacts fonctionnels, techniques, organisationnels, calendaires, financiers, de sécurité et de conformité. Les modalités d'instruction, de validation et de planification des évolutions seront définies lors de la phase de cadrage.

Le titulaire devra assurer la traçabilité des demandes, des estimations, des décisions et des arbitrages au sein des instances de gouvernance du projet.

6.3.5 Gestion documentaire et qualité projet

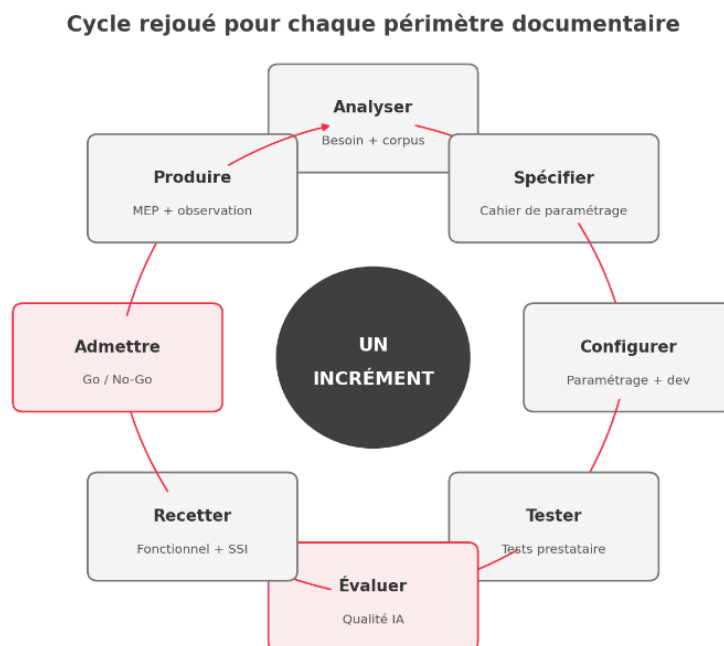
Le titulaire devra centraliser l'ensemble des documents produits auprès du chef de projet qui garantira l'accessibilité, la traçabilité et la conservation des livrables, décisions, comptes rendus et documents de référence. Des règles de gestion documentaire devront être définies dès la phase de cadrage. Chaque livrable devra être identifié de manière unique et faire l'objet d'un suivi de version permettant de retracer son historique de modification et de validation.

6.4 Cycle de réalisation d'un incrément

Chaque ensemble documentaire devra suivre un cycle de mise en œuvre structuré et reproductible, appliqué à chaque nouveau périmètre traité. Ce cycle a pour objectif de sécuriser

les choix réalisés, de mesurer la qualité des résultats obtenus et de garantir l'adéquation de la solution aux besoins métiers de la CHD.

Chaque cycle comprendra les étapes suivantes :



6.4.1 Analyse du besoin et qualification du corpus

Avant tout paramétrage ou traitement, le titulaire devra réaliser une analyse du besoin métier et du corpus documentaire concerné. Cette étape visera à qualifier précisément le périmètre à traiter et à définir les objectifs attendus pour chaque ensemble documentaire.

Le titulaire devra également vérifier que les données nécessaires aux traitements sont accessibles, autorisées, exploitables et représentatives du périmètre ciblé. Les éventuelles limitations liées à la qualité, à la complétude, à la volumétrie ou aux conditions d'accès aux données devront être identifiées et documentées dès cette phase.

6.4.2 Spécifications et cahier de paramétrage

Pour chaque ensemble documentaire, le titulaire devra produire un dossier de spécification décrivant les modalités de mise en œuvre du traitement retenu. Ce dossier constituera la référence de l'incrément et servira de base aux phases de configuration, de test, d'évaluation et de recette.

Il devra notamment préciser :

- les règles de gestion et de traitement applicables,
- les prompts, modèles et composants d'IA utilisés,
- les paramètres et seuils de décision retenus,
- les étapes du workflow et leur enchaînement,
- les contrôles et validations humaines prévus,
- les indicateurs de qualité associés,
- les formats de restitution, d'export et de validation des résultats.

Toute évolution significative des traitements ou des paramètres devra faire l'objet d'une mise à jour de ce dossier afin de garantir la traçabilité et la reproductibilité de l'incrément

6.4.3 Paramétrage et développements

Le titulaire réalisera le paramétrage initial de la solution pour chaque ensemble documentaire, en associant étroitement les administrateurs fonctionnels de la CHD. Les workflows, modèles d'IA, règles de traitement, exemples, seuils, contrôles et interfaces devront être livrés sous une forme lisible, documentée, versionnée et modifiable par la CHD au moyen des fonctions standard d'administration.

Lorsque les fonctionnalités standard de la plateforme ne permettent pas de répondre aux besoins de la CHD, le titulaire réalisera les développements complémentaires nécessaires.

L'ensemble des paramétrages, personnalisations et développements spécifiques devra être documenté et versionné. Les paramétrages devront pouvoir être modifiés, remplacés, exportés, réimportés ou supprimés sans remise en cause du fonctionnement global. Le titulaire devra éviter d'enfermer une règle métier dans le code lorsque celle-ci peut être exposée sous forme de paramètre administrable.

6.4.4 Tests internes du prestataire

Avant toute remise à la CHD, le titulaire devra réaliser les tests nécessaires à la validation de la solution et des paramétrages réalisés. Ces tests comprendront notamment les tests unitaires, les tests d'intégration, les tests de sécurité, les tests de robustesse et les tests de performance.

Les résultats de ces vérifications devront être documentés et transmis à la CHD. Le titulaire communiquera également la liste des anomalies identifiées, leur criticité et les éventuelles mesures de contournement mises en œuvre.

6.4.5 Évaluation de l'IA

Chaque ensemble documentaire devra faire l'objet d'une évaluation fondée sur un corpus de référence validé par la CHD. Cette évaluation visera à mesurer objectivement la qualité des traitements réalisés et des résultats produits.

Les indicateurs utilisés pourront notamment porter sur la précision, le rappel, les taux de faux positifs et de faux négatifs, la complétude des résultats ainsi que la qualité des métadonnées produites ou enrichies. Les seuils d'acceptation applicables à chaque indicateur seront définis avant le démarrage de l'évaluation.

6.4.6 Niveaux de qualité et critères de recevabilité des traitements IA

Les exigences ci-dessous constituent des minima de recevabilité. Elles s'appliquent indépendamment à chaque incrément, à chaque cas d'usage, à chaque catégorie de résultat et à chaque strate représentative du corpus. Une moyenne globale ne pourra pas compenser l'échec d'une catégorie sensible, d'une langue, d'un format ou d'un type de document.

6.4.6.1 Corpus de référence et vérité terrain

La CHD validera un corpus de référence versionné comprenant des cas représentatifs, des cas négatifs, des cas limites, des documents dégradés, des formats et langues variés, ainsi que des exemples couvrant les catégories rares ou à fort enjeu. Les résultats attendus seront établis ou validés par la CHD ; les cas critiques ou litigieux feront l'objet d'une double revue métier.

Les données seront séparées en un jeu de calibration, un jeu de test interne et un jeu d'acceptation tenu à l'écart des réglages. Le titulaire ne pourra utiliser le jeu d'acceptation pour

ajuster la solution. Toute version de corpus, de vérité terrain ou de règle d'annotation devra être identifiée dans le rapport d'évaluation.

6.4.6.2 Niveau de qualité global

Pour chaque opération évaluée, le taux de conformité est égal au nombre d'unités conformes divisé par le nombre total d'unités contrôlées, multiplié par 100. Ce taux devra être au minimum de 98 % pour chaque ensemble et chaque catégorie de résultat, sans préjudice des seuils spécifiques plus exigeants ci-dessous.

6.4.6.3 Conformité technique des sorties

La complétude des livrables, la conformité aux schémas et formats attendus, la présence des identifiants, le rattachement au bon document ou dossier, la conservation de l'ordre, la traçabilité de la source et la lisibilité des résultats sont exigées à 100 %. Tout élément non traité devra apparaître explicitement comme rejet, anomalie ou abstention, avec un motif exploitable.

6.4.6.4 Thematic DataQuest

Les minima suivants s'appliquent indépendamment :

- Complétude des champs obligatoires : au moins 99 %,
- Exactitude des dates, identifiants et autres valeurs structurées : au moins 99,5 % en correspondance exacte,
- Titres et descriptions jugés acceptables sans correction substantielle par les archivistes : au moins 98 %,
- Fidélité factuelle aux documents sources : au moins 99 % ; taux d'hallucination au plus égal à 1 % ; aucune hallucination critique,
- Présence d'un extrait, d'une localisation ou d'un élément de preuve vérifiable pour 100 % des valeurs générées.

6.4.6.5 Sensitive DataQuest

La priorité est de limiter les omissions de données sensibles. Les minima suivants s'appliquent indépendamment :

- Rappel global : au moins 99 % ; rappel d'au moins 99,5 % pour chaque catégorie qualifiée de critique par la CHD, soit un taux de faux négatifs au plus égal à 0,5 %,
- Précision globale : au moins 97 % ; score F1 global : au moins 98 %,
- Exactitude de la catégorie, de la localisation et de l'extrait justificatif : au moins 99 %,
- Mesure séparée pour chaque catégorie de sensibilité ; aucune agrégation ne pourra masquer un seuil non atteint.

6.4.6.6 Hallucinations, abstention et confiance

Une hallucination est toute information générée qui n'est pas soutenue par les documents et métadonnées autorisés, ou qui contredit ceux-ci. Toute hallucination portant sur une identité, une date, une décision, une obligation, une catégorie de sensibilité, une règle de communicabilité ou une restriction d'accès est critique. Les hallucinations critiques ne sont pas admises. La solution devra s'abstenir et solliciter une validation humaine lorsqu'elle ne dispose pas d'éléments suffisants.

Les niveaux de confiance devront être calibrés et contrôlés : aucun résultat inférieur au seuil applicable ne pourra être validé automatiquement ; l'erreur observée dans la tranche de confiance la plus élevée ne devra pas dépasser 0,5 % pour les résultats critiques. Les rapports présenteront les performances par tranche de confiance et la part des résultats orientés vers une revue humaine.

6.4.6.7 Robustesse et équité de traitement

Les indicateurs seront ventilés par langue, format, fonds, type documentaire, qualité OCR, taille de document et catégorie de sensibilité. Aucun segment disposant d'un effectif suffisant ne devra être inférieur aux seuils contractuels. L'écart de rappel entre segments comparables ne devra pas dépasser trois points de pourcentage sans justification acceptée et plan d'action validé par la CHD.

6.4.6.8 Reproductibilité et non-régression

À configuration, corpus et versions identiques, au moins 99 % des valeurs structurées devront être identiques sur trois exécutions successives. Toute évolution de règle, prompt, modèle, seuil, référentiel ou workflow devra déclencher les tests pertinents. Aucune mise en production ne sera admise si un indicateur passe sous son seuil, si la baisse globale dépasse 0,5 point ou si la baisse d'une catégorie dépasse un point.

6.4.6.9 Performance sous contrainte de qualité

La capacité à traiter 100 000 documents en huit heures sera testée sur le corpus et l'infrastructure de référence. Le respect du délai ne pourra être obtenu par une réduction des contrôles, de la traçabilité ou des seuils de qualité. Les erreurs techniques, reprises, files d'attente et ressources consommées devront être mesurées et restituées.

6.4.6.10 Erreurs bloquantes

La constatation d'une seule des situations suivantes entraîne le rejet de l'incrément concerné :

- Omission silencieuse d'un document, d'un dossier ou d'un résultat obligatoire,
- Rattachement d'un résultat, d'une preuve ou d'une qualification au mauvais document ou dossier,
- Altération, écrasement ou perte d'une donnée source sans règle explicite et validation,
- Hallucination critique intégrée ou proposée comme fiable sans signalement ni preuve,
- Divulgaration, transfert, réutilisation ou exposition non autorisée d'une donnée de la CHD,
- Utilisation d'un modèle, d'une version, d'un prompt, d'une règle ou d'une configuration non autorisés, ou impossibilité d'identifier ceux ayant produit le résultat,
- -Absence d'un livrable obligatoire, du rapport de contrôle qualité ou des éléments permettant de reproduire le traitement.

6.4.6.11 Contrôles de la CHD

La CHD effectuera ses propres contrôles, indépendamment de ceux du titulaire. Les contrôles automatiques de structure, de complétude, de rattachement et de traçabilité porteront sur 100 % des résultats. Le contrôle métier portera sur le jeu d'acceptation et sur un échantillon aléatoire stratifié choisi par la CHD représentant au moins 2 % de l'ensemble et au minimum 500 documents ; si l'ensemble documentaire comporte moins de 500 documents, il sera contrôlé intégralement. Les cas critiques, les alertes et les catégories rares pourront être contrôlés à 100 %.

Le titulaire fournira un outil permettant à la CHD de sélectionner les échantillons, afficher côte à côte la source, le résultat attendu et le résultat obtenu, qualifier les écarts, calculer automatiquement les indicateurs et exporter les preuves de contrôle. Les formules de précision, rappel, score F1, taux de faux positifs, taux de faux négatifs, taux de conformité et taux d'hallucination devront être documentées.

6.4.6.12 Rapport d'évaluation

Pour chaque incrément, le titulaire remettra le corpus et la vérité terrain utilisés, les versions de tous les composants, les résultats détaillés, matrices de confusion, intervalles de confiance,

ventilations par strate, liste des erreurs, analyse des causes, impacts, mesures correctives et résultats de non-régression. La CHD devra pouvoir reproduire les calculs.

6.4.6.13 *Décision d'admission*

L'ensemble des seuils applicables et l'absence d'erreur bloquante conditionnent l'admission. En cas d'échec, le titulaire corrigera la non-conformité et retraitera les éléments concernés sans coût supplémentaire dans le cadre du BUILD ou de l'hypercare applicable. Une nouvelle évaluation sera effectuée sur un échantillon indépendant. Deux échecs successifs sur un même critère entraîneront une escalade en comité de pilotage et pourront conduire à une décision de No-Go.

6.4.7 Recette fonctionnelle, technique et SSI

La recette aura pour objet de vérifier la conformité de la solution aux exigences du marché, au dossier de paramétrage validé et aux exigences de sécurité applicables.

Les anomalies identifiées seront enregistrées, qualifiées et classées selon leur niveau de criticité. Elles feront l'objet d'un suivi jusqu'à leur correction, leur requalification ou leur acceptation formelle par la CHD.

6.4.8 Admission et décision de mise en production

Pour chaque incrément, une décision d'admission sera prononcée sur la base des résultats de recette et des critères d'acceptation définis préalablement.

L'admission pourra être prononcée avec ou sans réserve. Les critères de décision de type Go / No-Go seront définis lors du cadrage. L'admission contractuelle d'un incrément devra être distinguée de l'autorisation opérationnelle de mise en production, qui relèvera des instances habilitées de la CHD.

6.4.9 Mise en production et déclenchement de l'hypercare

Après admission, le titulaire procédera au déploiement de l'incrément concerné dans l'environnement de production et réalisera les vérifications post-déploiement nécessaires à la validation du service.

Chaque mise en production de tout ou partie d'un incrément, pour l'un ou l'autre des deux cas d'usage, déclenchera une période d'hypercare de trois mois sur le périmètre effectivement déployé. Cette période court à compter de la vérification post-déploiement, indépendamment de la poursuite du BUILD. Plusieurs périodes d'hypercare pourront donc se chevaucher ; les obligations du titulaire s'appliqueront cumulativement aux périmètres encore couverts.

Avant chaque mise en production et pendant l'hypercare, les cas de tests automatisés et les jeux de référence concernés devront pouvoir être rejoués en masse. La solution comparera les résultats obtenus aux résultats attendus, signalera toute régression et empêchera la publication d'une configuration ne respectant pas les seuils, sauf dérogation formelle de la CHD. Les procédures de désactivation, de retour arrière et de restauration de la configuration devront être testées.

6.5 Trajectoire et planning des incréments

6.5.1 Axes de complexification

La mise en œuvre des incréments devra suivre une trajectoire progressive de complexification des traitements et des corpus documentaires. Chaque incrément devra capitaliser sur les acquis du précédent, réutiliser les composants déjà validés et enrichir progressivement les capacités de la solution.

L'évolution des périmètres pourra notamment porter sur :

- Les formats documentaires pris en charge (documents natifs, PDF numérisés, formats complexes ou atypiques),
- Les volumes de données traités,
- Le nombre de langues supportées,
- La qualité et l'homogénéité des documents,
- Les catégories de données sensibles à détecter,
- Le niveau d'automatisation des traitements et des contrôles.

Cette progression devra permettre de sécuriser les résultats obtenus, de valider les hypothèses de traitement à chaque étape et d'assurer une montée en charge maîtrisée avant la généralisation de la solution à l'ensemble des fonds concernés. Chaque nouveau incrément devra s'appuyer sur les paramétrages, retours d'expérience et indicateurs de qualité issus des incréments précédents.

6.5.2 Premier périmètre et quick-win

Le projet débutera par un premier périmètre documentaire volontairement limité, constitué de documents textuels numériques de faible complexité (format unique, langue unique, mono-caractère, périmètre métier cohérent). Ce premier incrément visera à démontrer rapidement la valeur métier de la solution à travers la production d'inventaires et l'automatisation de traitements documentaires simples. L'identification des données sensibles pourra être limitée à des cas élémentaires et ne visera pas, à ce stade, une couverture exhaustive.

L'objectif de disposer d'un premier périmètre exploitable à mi-février 2027 implique la conduite en parallèle des activités de cadrage, de préparation de l'infrastructure, d'installation de la plateforme et de configuration des premiers traitements.

6.5.3 Incréments intermédiaires

La mise en œuvre fonctionnelle sera organisée en 4 incréments successifs d'environ 60 jours chacun. Chaque incrément devra apporter une valeur métier tangible et étendre les capacités de la solution sur un nouvel axe de complexité.

Les évolutions pourront notamment concerner la diversité des contenus textuels, la prise en charge des images et des traitements OCR, le multilinguisme, la détection de données sensibles, l'augmentation des volumes traités ou l'extension du niveau d'automatisation des traitements. Chaque incrément devra capitaliser sur les composants, configurations et retours d'expérience issus des incréments précédents.

6.5.4 Stabilisation et généralisation

La dernière incrémentation sera également une phase de stabilisation afin de consolider les performances, la sécurité, l'exploitabilité et la maintenabilité de la solution. Cette phase permettra également de finaliser la documentation, de traiter les anomalies résiduelles, d'optimiser les configurations et de valider les conditions d'exploitation cible.

Cette phase préparera la généralisation de la solution et son fonctionnement pérenne dans les environnements de production de la CHD.

6.5.5 Planning prévisionnel

Le titulaire fournira un planning détaillé couvrant la période depuis janvier 2027, dans un délai de 11 mois couvrant les 8 mois d'implémentation du projet et les 3 mois de la dernière phase d'hypercare. Ce planning devra présenter les différentes phases du projet, les incréments prévus, les jalons de validation, les recettes, les admissions, les mises en production ainsi que les principales dépendances entre activités.

Le titulaire devra identifier les chemins critiques, les points de décision attendus de la CHD, du CTIE ou des autres parties prenantes, ainsi que les principaux prérequis conditionnant l'avancement du projet.

Le planning précisera explicitement si les durées des incréments sont exprimées en jours calendaires ou en jours ouvrés.

6.6 Livrables, compétences et préparation de l'exploitation

6.6.1 Livrable par phase et incrément

Le titulaire devra fournir les livrables nécessaires au pilotage, à la validation et à l'exploitation de la solution. Chaque phase et chaque incrément donneront lieu à la production et à la mise à jour des documents correspondants.

Les livrables attendus sont :

- Un document de cadrage fonctionnel et technique complet, en fin de phase de cadrage,
- Un dossier d'architecture fonctionnelle et technique, en fin de phase de cadrage,
- Un dossier de sécurité et de conformité, en fin de phase de cadrage,
- Un cahier de paramétrage (configurations et versions déployées de la solution), pour chaque incrémentation,
- Les rapports de tests, pour chaque incrémentation,
- Les rapports d'évaluation des traitements IA, pour chaque incrémentation,
- Les procès-verbaux de recette, pour chaque incrémentation,
- Un dossier d'admission, pour chaque incrémentation.

Pour le projet entier :

- La documentation d'exploitation et d'administration,
- Les guides utilisateurs,
- Les supports de formation,
- Les dossiers de mise en production et de transfert en exploitation,
- Les supports de présentation pour les COPIL et les procès-verbaux associés.

Si besoin, les livrables projets devront être mis à jour régulièrement afin de refléter l'état réel de la solution.

6.6.2 Formation et transfert de compétences

Le titulaire assurera la formation des différents publics concernés par la solution, notamment les utilisateurs, les administrateurs fonctionnels et les équipes techniques.

Les actions de formation devront permettre à la CHD de prendre rapidement en main la plateforme, d'exploiter et contrôler les résultats, d'administrer les processus courants et de réaliser en autonomie les paramétrages relevant de son périmètre. Elles comporteront des exercices pratiques sur la création et l'évolution de règles, prompts, seuils, référentiels, profils et workflows, le rejeu des tests, la publication et le retour arrière.

Le titulaire organisera un transfert de compétences progressif, fondé sur la co-configuration puis la réalisation supervisée et enfin autonome par la CHD. Les supports, guides pas-à-pas, exemples, modèles, jeux de tests et procédures remis devront permettre de faire évoluer les règles livrées et d'en créer de nouvelles sans dépendance au titulaire.

6.6.3 Accompagnement au changement

Le titulaire devra prévoir les actions nécessaires à l'adoption de la solution par les utilisateurs (archivistes). Cet accompagnement pourra notamment comprendre des démonstrations, ateliers, supports de communication, guides pratiques et séances de prise en main.

Une attention particulière devra être portée à la compréhension des capacités et des limites de l'IA. Les utilisateurs devront être sensibilisés au caractère assisté des traitements réalisés et au maintien d'une validation humaine pour les décisions et résultats engageant la CHD.

6.6.4 Clôture du BUILD et transfert à l'exploitation

La phase BUILD sera considérée comme achevée lorsque les périmètres prévus auront été admis, les anomalies résiduelles traitées ou acceptées, la documentation et les jeux de tests remis, les formations réalisées et les conditions d'exploitation validées. La clôture suppose également la réussite, par des administrateurs de la CHD et sans intervention du titulaire, du scénario d'autonomie défini au présent cahier des charges.

Le passage à l'exploitation fera l'objet d'un transfert formalisé précisant les composants exploités, les responsabilités opérationnelles, les procédures, configurations, exports, jeux de tests et éléments remis aux équipes concernées. Ce transfert n'emporte aucun engagement de maintenance ou de support récurrent du titulaire au-delà des périodes d'hypercare prévues au marché.

7. Transition vers l'exploitation, hypercare, autonomie et réversibilité

7.1 Principes et périmètre

7.1.1 Périmètre de l'après-BUILD

Le présent marché ne comprend pas une prestation récurrente d'exploitation, de support ou de maintenance applicative après le BUILD et les périodes d'hypercare. À l'issue de celles-ci, l'exploitation et les éventuelles maintenances seront assurées par la CHD, le CTIE, un tiers ou un marché distinct.

Le titulaire devra néanmoins documenter la répartition cible des responsabilités entre la CHD, le CTIE, les fournisseurs de services d'IA et les éventuels futurs exploitants. La matrice RACI précisera sans ambiguïté la date de fin de chaque obligation du titulaire et les éléments transférés.

7.1.2 Hypercare après chaque mise en production

Chaque mise en production de tout ou partie d'un incrément déclenchera, pour le périmètre déployé, une période d'hypercare de trois mois. Les périodes sont indépendantes et peuvent se chevaucher. L'hypercare regroupe l'observation renforcée, l'assistance, la surveillance des performances et de la qualité IA, la gestion prioritaire des incidents, les corrections de non-conformités, les ajustements de paramétrage, le rejeu des tests et l'accompagnement à l'autonomie de la CHD.

La période débute après les vérifications post-déploiement et se clôt séparément par procès-verbal lorsque les critères de sortie sont atteints. Elle ne constitue ni une prestation de RUN ni un engagement de maintenance applicative récurrente au-delà des trois mois applicables au périmètre concerné.

7.1.3 Obligations pendant l'hypercare

Pendant chaque période d'hypercare, le titulaire assurera la surveillance renforcée du périmètre déployé, le contrôle des traitements et indicateurs IA, le diagnostic, la correction des non-conformités, les ajustements de paramétrage, l'assistance aux utilisateurs et la tenue d'un registre des incidents, décisions et actions.

Un point de suivi au minimum hebdomadaire présentera les volumes, performances, erreurs, faux positifs, faux négatifs, hallucinations, corrections utilisateur, incidents, causes racines, délais de traitement et actions préventives. Les éléments nécessaires à la reproduction et à la vérification des incidents devront être conservés.

7.2 Assistance et niveaux de service pendant l'hypercare

7.2.1 Canaux, disponibilité et escalade

Le titulaire mettra à disposition, pendant les plages ouvrées de la CHD définies au cadrage, un interlocuteur joignable directement par téléphone, complété par une adresse électronique et un outil de suivi. Chaque demande recevra un identifiant, un accusé de réception, une criticité, un responsable, un historique, une échéance et une confirmation de clôture par la CHD.

7.2.2 Gestion des incidents et problèmes pendant l'hypercare

Un incident critique, entraînant une indisponibilité totale, un risque pour les données, une erreur bloquante ou l'impossibilité d'utiliser le périmètre déployé, devra être pris en charge sous 30 minutes ouvrées. Un contournement ou rétablissement sera visé sous quatre heures ouvrées et des informations d'avancement seront fournies au moins toutes les heures.

Un incident majeur affectant une fonction essentielle ou un nombre significatif d'utilisateurs devra être pris en charge sous deux heures ouvrées, avec contournement visé sous un jour ouvré. Les autres incidents devront être pris en charge sous un jour ouvré et faire l'objet d'un plan de résolution sous cinq jours ouvrés.

Une analyse de cause racine et un plan d'action seront exigés pour tout incident critique ou récurrent. À la demande de la CHD, le titulaire interviendra sur site au plus tard le jour ouvré suivant lorsqu'un incident critique n'est pas rétabli sous quatre heures, lorsqu'une même cause se reproduit au moins trois fois sur dix jours ouvrés, ou lorsqu'au moins cinq incidents affectent un même périmètre sur cinq jours ouvrés.

7.2.3 Mesure des niveaux de service

Les délais seront mesurés dans l'outil de suivi depuis le signalement ou la détection automatique jusqu'à la prise en charge, au rétablissement et à la résolution. Toute suspension de délai devra être justifiée et acceptée par la CHD.

Le tableau de bord d'hypercare présentera le respect des délais, l'âge des incidents, les réouvertures, les récurrences, la disponibilité du périmètre, les volumes en échec, les résultats des tests automatisés et les indicateurs de qualité IA.

Le titulaire informera sans délai la CHD de toute dégradation significative, risque de non-conformité ou incident affectant les données. Les modalités d'escalade nominatives seront définies avant la première mise en production.

7.3 Limites contractuelles de l'hypercare

7.3.1 Prestations incluses pendant l'hypercare

Sont incluses uniquement les corrections nécessaires pour rendre conforme le périmètre livré aux exigences, spécifications et critères d'acceptation, ainsi que les ajustements de configuration requis pour stabiliser son fonctionnement et sa qualité pendant l'hypercare.

Ces interventions couvrent le diagnostic, la correction, les tests, le redéploiement maîtrisé, le rejeu des traitements concernés, la mise à jour de la documentation et l'assistance nécessaire à la validation par la CHD. Elles ne constituent pas une maintenance récurrente au-delà de la période applicable.

Toute intervention sera testée sur un environnement adapté, soumise aux tests automatisés et de non-régression, approuvée selon le processus de changement et assortie d'une procédure de retour arrière.

7.3.2 Prestations exclues du marché

Sont exclues du présent marché, après admission et hors correction d'une non-conformité couverte par l'hypercare, les prestations de maintenance corrective, préventive, adaptative et évolutive, le support récurrent, l'exploitation quotidienne et le maintien en condition opérationnelle ou de sécurité.

Toute nouvelle fonctionnalité, adaptation à un besoin apparu après validation, intégration supplémentaire, changement de périmètre ou intervention après l'expiration de l'hypercare relèvera de la CHD, du CTIE, d'un tiers ou d'un contrat distinct. Le titulaire identifiera séparément, dans son offre, les prérequis et coûts récurrents éventuels sans les inclure dans le prix du présent marché.

7.3.3 Clôture de chaque hypercare

Chaque hypercare sera clôturée séparément pour le périmètre concerné après remise d'un bilan, transfert des connaissances, mise à jour des configurations, documentations et jeux de tests, et vérification des critères de sortie.

La clôture suppose l'absence d'incident critique ouvert, la résolution ou l'acceptation formelle des autres anomalies, le respect des seuils de qualité sur une période de quatre semaines, la réussite des tests de non-régression et la capacité de la CHD à administrer le périmètre.

La fin de l'hypercare sera actée par un procès-verbal. Elle mettra fin aux obligations de support du titulaire sur le périmètre, sans affecter les éventuelles autres périodes d'hypercare encore actives.

7.4 Autonomie de la CHD après le projet

7.4.1 Administration des modèles, prompts, règles et workflows

La solution et les livrables devront permettre à la CHD d'administrer les modèles autorisés, prompts, règles, référentiels, paramètres, seuils, profils et workflows sans dépendance au titulaire. Les opérations courantes seront réalisables dans l'interface d'administration et couvertes par les formations et guides remis.

Toute évolution sera versionnée, testée, validée et réversible. La CHD pourra appliquer des profils en masse, différencier les configurations par projet de traitement, simuler l'impact d'un changement, rejouer les jeux de référence et publier une version selon un circuit d'approbation configurable.

7.4.2 Suivi autonome de la qualité de l'IA

La solution devra fournir à la CHD les tableaux de bord et exports nécessaires au suivi autonome de la précision, du rappel, des faux positifs, des faux négatifs, des hallucinations, des abstentions, des corrections utilisateur, des performances par strate et des dérives. Les seuils d'alerte seront paramétrables par la CHD.

7.4.3 Suivi des consommations et coûts

La solution devra permettre à la CHD de suivre les licences, volumes, appels aux modèles, consommations de tokens, ressources d'infrastructure et coûts. Ces fonctions de mesure sont incluses dans la solution ; leur analyse récurrente après hypercare incombera à la CHD ou à l'exploitant qu'elle désignera.

Des tableaux de bord exportables devront permettre d'identifier les dérives, de définir des quotas et alertes, et d'optimiser le routage vers les modèles autorisés sans intervention du titulaire.

7.4.4 Supervision technique et fonctionnelle

La solution devra fournir les mécanismes, tableaux de bord, alertes et journaux nécessaires pour que la CHD ou son exploitant suive la disponibilité, les traitements, erreurs, performances, files d'attente et principaux indicateurs fonctionnels et techniques.

Les droits d'accès, seuils d'alerte, vues et exports devront être administrables par la CHD. Les procédures de diagnostic et d'escalade seront remises avant la clôture du BUILD.

7.5 Sécurité, continuité et réversibilité

7.5.1 Responsabilités après hypercare

Après l'expiration de l'hypercare, le maintien en condition opérationnelle et de sécurité ne relève pas du présent marché. La solution devra toutefois exposer les mécanismes, alertes, inventaires de composants et procédures permettant à la CHD, au CTIE ou à un tiers d'assurer ces activités.

Pendant le BUILD et l'hypercare, le titulaire informera la CHD des vulnérabilités et risques affectant les composants qu'il livre, proposera les mesures correctives et appliquera les corrections nécessaires à la conformité du périmètre couvert.

7.5.2 Incidents de sécurité pendant le BUILD et l'hypercare

Pendant le BUILD et l'hypercare, le titulaire appliquera les procédures de détection, qualification, notification, investigation, confinement et remédiation des incidents de sécurité affectant son périmètre, avec la même exigence d'escalade que pour un incident critique.

Les responsabilités respectives de la CHD, du CTIE, du titulaire et des éventuels fournisseurs de services ou de modèles d'IA devront être clairement définies, notamment en cas de violation de données, d'incident affectant un modèle d'IA ou de compromission d'un composant tiers.

Les modalités d'escalade, de communication et de notification des parties concernées devront être documentées.

7.5.3 Sauvegarde et restauration

Le titulaire configurera et testera, pendant le projet, les mécanismes de sauvegarde couvrant les configurations, workflows, référentiels, journaux, paramètres, métadonnées et résultats nécessaires à la restauration. Il remettra les procédures permettant leur exécution ultérieure par l'exploitant désigné.

Un test de restauration sera réalisé avant chaque première mise en production d'un périmètre et avant la clôture du BUILD. Les opérations récurrentes ultérieures ne sont pas incluses dans le marché.

7.5.4 Continuité et reprise d'activité

Le titulaire documentera et testera pendant le projet les procédures de continuité et de reprise applicables à la solution, en précisant les responsabilités, dépendances et conditions de redémarrage.

Les objectifs de reprise et de disponibilité seront vérifiés avant admission. Leur suivi récurrent après hypercare incombera à la CHD ou à l'exploitant désigné.

7.5.5 Réversibilité et fin du marché

Le titulaire devra garantir la réversibilité complète de la solution à l'issue du marché ou en cas de changement de prestataire. Il assurera le transfert des éléments nécessaires vers la CHD, le CTIE ou tout nouveau titulaire désigné par la CHD.

La réversibilité devra couvrir notamment les données, référentiels, configurations, workflows, prompts, modèles paramétrés, historiques, journaux, documentations, bases de connaissances et tout autre élément nécessaire à la poursuite de l'exploitation.

Le titulaire fournira les procédures, exports et informations permettant de reprendre l'exploitation de la solution dans des conditions opérationnelles satisfaisantes, sans reconstruction complète des paramétrages ou des processus mis en œuvre durant le marché.

7.6 Documentation, transfert et clôture

7.6.1 Documentation et base de connaissances transférées

Le titulaire maintiendra à jour, jusqu'à la fin de la dernière période d'hypercare, l'ensemble de la documentation fonctionnelle, technique, d'administration, d'exploitation, de sécurité et d'architecture nécessaire à l'utilisation autonome de la solution.

Les incidents résolus, causes racines, procédures, paramétrages, jeux de tests, exemples et retours d'expérience alimenteront une base de connaissances exportable et intégralement remise à la CHD.

7.6.2 Formation et validation de l'autonomie

Les formations seront réalisées pendant le BUILD et complétées au besoin pendant les hypercares. Elles viseront l'autonomie des utilisateurs et administrateurs fonctionnels ; aucune formation continue récurrente après ces périodes n'est incluse dans le marché.

Les supports, guides d'administration, exercices, corrigés, profils modèles et jeux de tests seront mis à jour jusqu'à la fin de la dernière hypercare et remis dans des formats modifiables.

7.6.3 Bilan de clôture

À la clôture de chaque hypercare, le titulaire présentera un bilan couvrant incidents, qualité IA, performances, consommations, sécurité, corrections, évolutions de paramétrage, compétences transférées et risques résiduels.

Le bilan final consolidera l'ensemble des périodes d'hypercare, confirmera les éléments transférés et identifiera les activités récurrentes qui devront être reprises hors du présent marché.

8. Déroulement de la procédure d'attribution du marché

8.1 Dépôt des offres

La remise des offres a lieu aux jour et heure fixés sous la section 1.7 « Date limite de réception des offres ». Lorsque l'opérateur économique constate que son offre contient une erreur avant que le délai de remise ne soit dépassé, il peut remettre une deuxième offre.

En cas de remise électronique de plusieurs offres, seule l'offre remise le plus récemment est prise en considération. Les autres offres sont détruites à l'issue de la procédure. L'opérateur économique veillera donc à inclure l'entièreté des documents à chaque dépôt.

Il n'est tenu compte que des offres arrivées ou remises dans le délai fixé.

Aucune séance d'ouverture n'est organisée.

L'offre avec tous les documents demandés par le présent cahier des charges devra être regroupée dans un fichier zip et envoyée obligatoirement par mail à archives@chd.lu avec copie à data@digital.etat.lu.

8.2 Examen de l'offre

Le pouvoir adjudicateur examine et vérifie les différentes offres quant à leur conformité avec le présent cahier des charges.

8.3 Délai de validité de l'offre

L'offre sera valable au moins cinq (5) mois après la date d'ouverture des offres. En cas de circonstances exceptionnelles, la CHD se réserve la possibilité de prolonger la durée de validité de 5 (cinq) mois. Le cas échéant, les opérateurs économiques ayant soumis une offre en seront informés.

8.4 Procédure d'attribution et contrat

Le déroulement de la procédure restreinte sans publication d'avis comprendra les étapes suivantes :

- Envoi restreint du cahier des charges ;
- Remise des offres par les candidats ;
- Ouverture des offres, sélection, analyse et réduction du nombre de candidats en fonction des critères décrits en 8.6 et 8.7. Le pouvoir adjudicateur retiendra un maximum de 5 candidats pour la phase suivante ;
- Présentations des offres et démonstrations optionnelles des outils : sur base des offres et des critères d'attribution décrits au 8.7, les 5 candidats retenus réalisent une présentation obligatoire de leur offre et de l'approche proposée. Pendant cette présentation, il sera considéré comme un avantage si les candidats peuvent faire une démonstration de l'outil qu'ils envisagent (optionnelle) ;
- Les présentations et les offres seront évaluées par un panel fonctionnel et technique de la CHD et du ministère de la Digitalisation ;
- Après évaluation des offres finales et des présentations selon les critères d'attribution, le pouvoir adjudicateur sélectionne le candidat final ;
- La conclusion du contrat a lieu par l'apposition de la signature du pouvoir adjudicateur sur le document de soumission remis par l'adjudicataire, après contre-signature du cahier des charges et du bordereau de soumission par la CHD.

8.5 Calendrier prévisionnel estimatif de la procédure

Phases	Dates ou période
Publication du cahier des charges	07/09/2026
Date limite de réception des questions et incohérences et ambiguïtés sur le cahier des charges	16/09/2026, 12h00
Réponse aux questions	25/09/2026
Réception des offres	07/10/2026, 12h00
Analyse des offres selon les critères de sélection et d'attribution à l'exception du critère d'attribution n° 4	07/10/2026-28/10/2026
Notification concernant les opérateurs économiques sélectionnés pour les présentations	29/10/2026
Fin des présentations	06/11/2026
Analyse des présentations selon le critère d'attribution n°4 pour attribution de la note finale.	13/11/2026
Notification de l'opérateur économique retenu (après signature de la convention entre la CHD et le Ministère)	18/11/2026
Période de Standstill (15 jours calendaires)	18/11/2026-02/12/2026
Signature du contrat	14/12/2026
Démarrage des prestations	11/01/2027

8.6 Critères de sélection

Pour qu'une offre soit considérée comme conforme, elle doit pouvoir justifier des éléments suivants :

1. Le candidat devra justifier d'au moins trois références de projets similaires. Les projets présentés décriront le même type de technologie. Ils porteront soit sur l'analyse de corpus documentaires hétérogènes soit sur des organismes publics.
2. Les composants de l'architecture du candidat seront interchangeables et l'effort d'interchangeabilité sera quantifié dans son offre.
3. Le candidat présente l'infrastructure cible et l'évalue financièrement.
4. L'offre couvre toutes les exigences fonctionnelles.
5. L'offre respecte-t-elle le calendrier SpeedUp (8 mois d'implémentation et 3 mois d'hypercare) de livraison.
6. Le candidat suit les recommandations de contenu et de format de réponse énoncées dans la section 8.9 *Format et contenu de l'offre*.

Seuls les candidats ayant soumis une offre conforme seront évalués sur les critères d'attribution.

8.7 Critères d'attribution

8.7.1 Critère 1 – Expérience

L'offre devra comporter une brève présentation de l'entreprise comprenant ses principaux domaines d'activité et son expérience pertinente.

Un historique des projets antérieurs de nature similaire (même type de technologie), idéalement au moins trois références, accompagné d'une brève description du contexte, de la portée, des technologies utilisées, des livrables et du rôle de l'opérateur économique.

Le candidat sera évalué sur la qualité des trois références, sur la nature du corpus documentaire (exemple : présence de données sensibles au sens de la [Loi u 17 août 2018 relative à l'archivage](#)) et sur sa capacité à avoir développé un outil performant utilisant des technologies d'IA fiables.

8.7.2 Critère 2 – Approche méthodologique et fonctionnelle

Le candidat décrira l'approche proposée pour la réalisation de la mission, en précisant les modalités d'implication des équipes métier et leur intégration aux différentes phases du projet. Il présentera notamment la méthodologie retenue pour la préparation, l'exécution, le suivi et la validation des tests, ainsi que les modalités de prise en compte des retours des utilisateurs métier.

En outre, le candidat présentera un planning détaillé et réaliste des différentes phases de la mission, en précisant les principales échéances, dépendances, livrables et jalons de validation. Il quantifiera l'effort nécessaire à la réalisation des prestations, notamment en termes de charge de travail, de profils mobilisés et de répartition de l'effort dans le temps.

Il s'attachera également à préciser l'étendue de la couverture fonctionnelle proposée au regard des besoins et exigences du présent cahier des charges. Il identifiera les fonctionnalités couvertes, en distinguant, le cas échéant, les éléments pris en charge de manière standard, les adaptations nécessaires et les éventuelles limites ou exclusions.

Le candidat justifiera de manière claire et structurée la couverture fonctionnelle proposée. Il établira notamment le lien entre les exigences du cahier des charges et les fonctionnalités, solutions ou prestations prévues, afin de permettre à l'adjudicateur d'évaluer de manière objective le niveau de couverture et la pertinence des choix proposés.

Le candidat présentera sa démarche qualité ainsi que les dispositifs prévus pour assurer la conformité, la fiabilité et la maîtrise des prestations. Il détaillera notamment les modalités d'identification, d'évaluation, de suivi et de traitement des risques, ainsi que les mesures préventives et correctives envisagées en cas de dérive, d'anomalie ou de difficulté rencontrée au cours de la mission.

Enfin, le candidat présentera la composition de l'équipe fonctionnelle proposée ainsi que les compétences, qualifications et expériences des profils mobilisés. Il précisera le rôle et le niveau d'intervention de chaque profil, en mettant en évidence leur adéquation avec les exigences de la mission, les domaines fonctionnels concernés et les enjeux spécifiques du projet. Les profils des membres de l'équipe seront présentés dans un CV.

Le candidat devra proposer un ou plusieurs profils fonctionnels évalués selon les critères suivants :

- Diplôme de niveau BAC+5 minimum dans un domaine pertinent pour la mission, notamment informatique, systèmes d'information, gestion, sciences économiques, ingénierie ou domaine équivalent,
- Au moins 4 ans d'expérience professionnelle significative, avec une expérience éprouvée dans la conduite d'entretiens avec les parties prenantes, le recueil et l'analyse des besoins, la formalisation

des exigences et la réalisation d'analyses fonctionnelles, notamment dans le cadre de projets faisant intervenir des solutions basées sur les technologies d'intelligence artificielle,

- Le ou les profils devront démontrer une capacité avérée à analyser les processus métier, identifier et formaliser les besoins, traduire les besoins métier en exigences fonctionnelles et contribuer à la définition et à la validation de solutions, en assurant l'interface entre les équipes métier et les équipes techniques. Une expérience dans des projets intégrant des technologies d'IA constituera un atout,
- Le ou les profils devront faire preuve d'excellentes qualités rédactionnelles, notamment pour la production de spécifications fonctionnelles, comptes rendus d'entretiens, dossiers d'analyse, cahiers de tests, procédures, rapports et autres documents nécessaires à la mission,
- La maîtrise du français ou de l'anglais est obligatoire (langue maternelle ou certificat CERCL niveau C2 en expression écrite).

L'équipe fonctionnelle devra :

- Se déplacer sur le site de la CHD ou du ministère de la Digitalisation afin de participer aux réunions du comité de pilotage ou lors de certaines réunions nécessitant des échanges plus approfondis comme des ateliers ou des comités de projet spécifiques.
- Être flexible et s'adapter dans la mesure du possible au calendrier politique dont dépendent les équipes métier et technique de la CHD.

Le prestataire devra également être en mesure de proposer des profils équivalents et de garantir la continuité de service grâce à un dispositif de remplacement en cas d'absence prolongée d'un des profils retenus.

8.7.3 Critère 3 – Approche technique

Le candidat présentera l'architecture technique proposée et démontrera sa cohérence avec les besoins fonctionnels et techniques du présent cahier des charges. Il justifiera les principaux choix d'architecture et précisera leur faisabilité au regard du contexte existant, des contraintes techniques, des exigences de sécurité, d'intégration, d'exploitation et d'évolutivité.

Le candidat détaillera le dimensionnement de l'architecture proposée au regard des besoins identifiés. Il justifiera les hypothèses retenues, notamment en matière de capacité, de performance, de disponibilité, de volumétrie et de montée en charge, et précisera les marges prévues afin de garantir la pérennité de la solution.

Le candidat présentera et justifiera les coûts associés à l'architecture proposée, en distinguant autant que possible les coûts d'acquisition, de licences, d'infrastructure, d'intégration, d'exploitation et de maintenance. Il mettra en évidence les principaux facteurs de coût ainsi que les éventuelles options ou leviers permettant d'optimiser le coût global de la solution sur l'ensemble de son cycle de vie.

Le candidat présentera la composition de l'équipe technique proposée, ainsi que les compétences, qualifications et expériences des profils mobilisés. Il précisera les responsabilités de chaque profil et leur adéquation avec les exigences techniques de la mission. Le candidat décrira également les dispositions prévues pour assurer la continuité de service, notamment en cas d'absence, de turnover ou d'indisponibilité d'un intervenant clé, ainsi que les modalités de transfert et de conservation des connaissances.

Le candidat devra proposer un ou plusieurs profils techniques évalués selon les critères suivants :

- Diplôme de niveau BAC+5 minimum dans un domaine pertinent pour la mission, notamment informatique, ingénierie informatique, systèmes d'information, télécommunications ou domaine technique équivalent.,
- Au moins 4 ans d'expérience professionnelle significative dans le domaine des technologies de l'information, avec une expérience avérée dans la conception, l'architecture, l'intégration, le

développement ou l'exploitation de solutions informatiques, idéalement dans des environnements complexes ou à fortes exigences de disponibilité, de sécurité et de performance,

- Le ou les profils devront démontrer une maîtrise des domaines techniques nécessaires à la réalisation de la mission, notamment en matière d'architecture des systèmes d'information, d'intégration de solutions, d'interopérabilité, de sécurité, de performance et d'exploitation. Les compétences spécifiques requises par la solution proposée devront être clairement identifiées et justifiées,
- Le ou les profils devront faire preuve d'excellentes qualités rédactionnelles, notamment pour la production de documents d'architecture, de dossiers techniques, de spécifications, de procédures, de rapports et de toute autre documentation nécessaire à la mission,
- La maîtrise du français ou de l'anglais est obligatoire (langue maternelle ou certificat CERCL niveau C2 en expression écrite).

L'équipe technique devra :

- Se déplacer sur le site de la CHD ou du ministère de la Digitalisation afin de participer aux réunions du comité de pilotage ou lors de certaines réunions nécessitant des échanges plus approfondis comme des ateliers ou des comités de projet spécifiques.
- Être flexible et s'adapter dans la mesure du possible au calendrier politique dont dépendent les équipes métier et technique de la CHD.

Le prestataire devra également être en mesure de proposer des profils équivalents et de garantir la continuité de service grâce à un dispositif de remplacement en cas d'absence prolongée d'un des profils retenus.

8.7.4 Critère 4 – Présentation

À l'issue de l'évaluation des offres sur les critères 1, 2, 3 et 5 (85 points), les 5 candidats ayant obtenu les meilleurs scores seront invités à réaliser une présentation devant l'adjudicateur. Cette présentation a pour objectif de permettre à ce dernier d'apprécier concrètement la qualité de l'approche proposée, ainsi que la capacité des équipes technique et fonctionnelle à présenter et défendre leur proposition.

Les candidatures sont classées en ordre selon les résultats de l'évaluation et le pouvoir adjudicateur les convoque pour les présentations dans cet ordre.

Lors des présentations, les candidats seront invités à présenter leur offre et éventuellement à faire une démonstration de l'outil qu'ils envisagent, ce qui constituerait un avantage auprès des évaluateurs sous la forme de 5 points supplémentaires accordés au critère 4.

La présentation des offres est une obligation et la non-représentation d'un candidat à cette réunion constitue un motif d'exclusion.

La présentation sera organisée sur une durée maximale d'une heure (1h00) et devra respecter les critères suivants :

- Respect de la durée : le candidat devra respecter la durée impartie de 1h00, incluant la présentation (30 mn) et les échanges avec l'adjudicateur (30 mn).
- Respect des participants et répartition du temps de parole : la présentation devra être assurée par deux participants au maximum, représentant respectivement les volets technique et fonctionnel. Le candidat devra démontrer une répartition équilibrée et pertinente du temps de parole entre les deux profils.
- Respect des langues : la présentation devra être réalisée en français ou en anglais, conformément aux exigences linguistiques définies dans le présent cahier des charges.
- Qualité de la présentation : l'adjudicateur évaluera la clarté, la structure, la pertinence et la qualité pédagogique de la présentation, ainsi que la capacité des intervenants à présenter de manière synthétique et convaincante les éléments essentiels de leur proposition.

- Qualité des réponses apportées : l'adjudicateur évaluera la pertinence, la précision et la maîtrise des réponses apportées aux questions posées au cours de la présentation, ainsi que la capacité des intervenants à justifier les choix proposés et à apporter des réponses adaptées aux enjeux fonctionnels et techniques du marché.

8.7.5 Critère 5 – Prix

L'évaluation du prix se fait en fonction de l'offre la plus économiquement avantageuse. Ainsi la note du candidat pour ce poste est donnée par la formule :

$$20 * (\text{Meilleur prix parmi les candidats}) / (\text{Prix du candidat})$$

8.8 Notation

La pondération des critères d'attribution est la suivante :

Critères	Points sur 100
Critère 1 – Expérience	5
Critère 2 – Approche méthodologique et fonctionnelle	35
Critère 3 – Approche technique	25
Critère 4 – Présentation	15
Critère 5 – Prix	20

Les prestataires seront notés par un panel d'experts pour chaque domaine de compétence (hors prix : en proportion du prestataire le moins cher).

8.9 Format et contenu de l'offre

En soumettant une offre, le soumissionnaire reconnaît avoir recueilli toutes les informations nécessaires à l'élaboration d'une offre valable, c'est-à-dire qu'il a pris connaissance des difficultés et des particularités des solutions à fournir et qu'il en tient compte lors de la préparation de son offre.

Toute modification apportée au cahier des charges par le soumissionnaire est considérée comme nulle et non avenue (Règlement grand-ducal du 8 avril 2018 portant exécution de la loi du 8 avril 2018 sur les marchés publics, art. 60 (1)).

L'offre doit être rédigée en langue française ou anglaise et contenir tous les documents indiqués ci-après conformément aux sections 8.6 « Critères de sélection » et 8.7 « Critères d'attribution » selon l'ordre indiqué ci-après :

- 1 Le bordereau de soumission dûment signé et rempli de l'annexe 10.1.
- 2 L'accord de confidentialité société dûment signé et rempli de l'annexe 10.2.
- 3 Un document présentant l'adéquation de la proposition du candidat avec les critères de sélection et d'attribution et comportant les points suivants dans le même ordre que présenté ci-après :
 - 1^{ère} partie : Une brève présentation de l'entreprise,
 - 2^{ème} partie : Les expériences / références en termes de projet,
 - 3^{ème} partie : La démarche de gestion de projet
 - 4^{ème} partie : Le calendrier et les jalons correspondants,
 - 5^{ème} partie : La réponse aux exigences fonctionnelles,
 - 6^{ème} partie : Démarche qualité et gestion des risques

- 7^{ème} partie : La réponse aux exigences techniques,
- 8^{ème} partie : Une brève présentation des profils fonctionnels et techniques pressentis,
- 9^{ème} partie : La grille tarifaire présentant les estimations de charge détaillées et le lien avec le prix déclaré sur le bordereau.

Le candidat devra mentionner, au début de chaque partie, à quelles sections du cahier des charges il répond ou fait référence.

4. Les CV des profils proposés détaillant leurs fonctions au cours des trois dernières années, en mettant l'accent sur toute expérience pertinente.

5. Le cas échéant un document décrivant l'approche pour la sous-traitance ou si une offre collective est envisagée.

Le candidat portera une attention toute particulière au niveau de détail qui doit permettre à la CHD d'en évaluer la plausibilité.

Le document visé au point 3 est limité à huit pages A4 pour les parties 1 à 8, hors page de garde et sommaire. La partie 9 relative à la grille tarifaire, les CV visés au point 4 ainsi que, le cas échéant, les documents visés au point 5 relatifs à la sous-traitance ou à l'offre collective ne sont pas comptabilisés dans cette limite.

9. Conditions d'exécution du marché

9.1 Règlementation applicable

Sauf disposition contraire prévue dans le présent document, les lois et règlements suivants, tels que modifiés, s'appliquent :

- La loi modifiée du 8 avril 2018 sur les marchés publics,
- Le règlement grand-ducal d'exécution du 8 avril 2018 de la loi modifiée du 8 avril 2018 sur les marchés publics,
- Le règlement grand-ducal modifié du 27 août 2013 relatif à l'utilisation des moyens électroniques dans les procédures de marchés publics et les procédures d'attribution de contrats de concession,
- La loi du 23 juillet 1991 ayant pour objet de réglementer la sous-traitance,
- La loi du 10 novembre 2010 instituant les recours en matière de marchés publics,
- La loi modifiée du 8 juin 1999 sur le budget, la comptabilité et la trésorerie de l'État,
- Les dispositions du Code civil,
- Le règlement (UE) n° 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données),
- La loi du 1^{er} août 2018 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel en matière pénale ainsi qu'en matière de sécurité,
- La loi modifiée du 31 juillet 2006 portant introduction d'un code du travail.

9.2 Régime tarifaire

Le marché sera attribué sur la base d'un prix forfaitaire. Le prix total proposé par l'opérateur économique doit être global et ne peut dépasser 140.000 EUR hors TVA. L'offre doit indiquer clairement le prix forfaitaire total hors TVA, la contingence de 10% du prix forfaitaire global hors TVA, le montant de la TVA applicable et le prix total TVA comprise.

La grille tarifaire doit inclure une proposition de facturation sous la forme d'un plan de paiement lié au calendrier du projet, aux étapes clés et/ou aux livrables. Les paiements seront effectués

conformément au plan de facturation accepté par l'autorité contractante et sous réserve de la validation des étapes clés ou des livrables correspondants.

Tous les coûts nécessaires à la bonne exécution du contrat doivent être inclus dans le prix forfaitaire, y compris la gestion de projet, les réunions, la documentation, le développement, les tests, l'assistance au déploiement et tous les autres coûts nécessaires à la fourniture de la solution, sauf mention contraire expresse.

Le prestataire envoie les factures électroniques via PEPPOL. Les factures détailleront les montants à payer par article. Les factures sont payables endéans un délai de trente (30) jours.

La pouvoir adjudicateur refuse d'ores et déjà le paiement d'acomptes.

9.3 Continuité d'activité

La mission devra être effectuée par le ou les profils présentés dans l'offre. En cas d'indisponibilité du ou des profils proposés dans l'offre, le titulaire devra présenter un profil équivalent à la CHD. Cette dernière se réserve le droit de refuser le profil après évaluation.

9.4 Droit d'inspection

Afin de faciliter à la fois le transfert de connaissances vers le pouvoir adjudicateur et le contrôle de qualité, le pouvoir adjudicateur s'octroie un droit d'inspection sur l'ensemble des livrables, méthodes et pratiques utilisés par l'opérateur économique lors de la production des résultats prévus dans le présent cahier des charges. La solution et l'approche proposées doivent tenir compte du fait qu'à l'issue du contrat, le pouvoir adjudicateur souhaite être en mesure de garantir la maintenance du projet lui-même. Au cours de la phase de mise en œuvre du projet, l'opérateur économique accorde un droit d'inspection au pouvoir adjudicateur afin de lui permettre de surveiller la qualité et de contrôler l'avancement des travaux.

9.5 Résiliation

Le contrat est résilié de plein droit et sans indemnité dans les cas suivants :

- Manquement aux conditions du marché,
- Non-respect des délais impartis,
- Faute grave dans l'exécution de la prestation.

Les manquements et fautes graves ci-avant énoncés visent notamment les cas suivants :

- Lorsque le prestataire ne respecte pas les clauses essentielles du contrat,
- En cas d'interruption ou de suspension de la prestation sans accord de la CHD,
- Lorsqu'après deux avertissements écrits, le prestataire continue à fournir un travail de qualité insuffisante,
- En cas de détérioration des biens appartenant à la CHD,
- Lorsque le prestataire contrevient à ses obligations légales et notamment à celles relatives à la législation du travail et à la convention collective applicable,
- En cas de faillite ou de liquidation judiciaire ou conventionnelle ou ouverture d'une autre procédure collective contre le prestataire,
- Lorsqu'il appert que le prestataire ne remplit pas ses obligations relatives à l'assurance responsabilités professionnelles,
- En cas de non-respect des règles de confidentialité ou de sécurité.

La résiliation du contrat se fera par lettre recommandée avec accusé de réception moyennant un préavis d'un mois. En cas de cessation de la relation contractuelle ou à la fin de la mission, le prestataire s'engage à restituer l'ensemble des données, y incluses toute copie et reproduction

de données, et à effacer de manière sécurisée toutes les données du pouvoir adjudicateur de tout système de traitement ou de conservation de données.

9.6 Modification et adaptation du marché

Le marché ne pourra pas être modifié ou adapté.

9.7 Sécurité, confidentialité et protection des données à caractère personnel

Les aspects liés à la sécurité et à la confidentialité sont pris très au sérieux par la CHD. Un accord de non-divulgaration ainsi qu'un accord relatif au traitement des données à caractère personnel doivent être signés par toute personne travaillant sur ce projet.

Les technologies impliquant un transfert des données de la CHD vers des services de cloud privés ne sont pas autorisées dans le cadre du projet. Le service de cloud gouvernemental (*GovCloud*) ou celui de la CHD seront privilégiés.

Le code source sera stocké dans le GitLab du CTIE et les livrables du projet dans le SharePoint dédié au projet de la CHD.

En cas de traitement de données à caractère personnel pour le compte de la CHD, le titulaire s'engage à conclure avec la CHD un accord de traitement des données conforme aux dispositions de l'article 28 du RGPD.

En outre, l'entreprise contractante doit apporter la preuve qu'elle offre des garanties suffisantes pour mettre en œuvre les mesures techniques et organisationnelles appropriées, de manière que le traitement des données à caractère personnel couvert par ce projet réponde aux exigences du RGPD, que la protection des données dès la conception soit respectée et que la protection des droits des personnes concernées conformément au RGPD soit assurée.

Tout au long du projet et au-delà, tous les salariés du titulaire participant au projet s'engagent à préserver la confidentialité de l'ensemble des informations et données relatives au projet et s'interdisent de les utiliser, de les divulguer, de les transmettre ou de les rendre accessibles à toute personne autre que les salariés habilités.

9.8 Propriété de la solution

Le titulaire cède à titre exclusif à la CHD l'ensemble des droits de propriété intellectuelle attachés aux résultats, créations, développements, documents, logiciels, données et, plus généralement, à tout élément protégeable réalisé spécifiquement dans le cadre du présent marché, y compris les droits d'auteur patrimoniaux et moraux, dans les limites prévues par la législation applicable.

Cette cession couvre notamment les codes sources et codes objets, scripts, algorithmes, modèles entraînés ou affinés spécifiquement pour le projet, paramètres et configurations, structures de données, métadonnées générées, schémas, interfaces, documentations, jeux et scénarios de test ainsi que les rapports relatifs au projet.

Le titulaire n'est pas autorisé à exploiter les éléments cédés au titre du présent marché, quels que soient l'usage ou le support d'exploitation.

Chaque partie reste titulaire des droits de propriété intellectuelle relatifs à ses éléments, connaissances et outils préexistants. Lorsque des éléments préexistants du titulaire sont nécessaires à l'exploitation de la solution, celui-ci accorde à la CHD les droits nécessaires à leur

utilisation, leur maintenance, leur modification, leur évolution et leur réutilisation, directement ou par l'intermédiaire d'un tiers.

La CHD peut mettre le socle technique et les composants génériques réutilisables à disposition du CTIE afin qu'ils puissent être exploités et, le cas échéant, mis à disposition d'autres entités de l'État luxembourgeois par l'intermédiaire du CTIE. Cette mutualisation ne concerne pas les données, documents, corpus, métadonnées ni résultats métier propres à la CHD.

Le titulaire identifie, avant leur utilisation, les éléments appartenant à des tiers qu'il envisage d'intégrer à la solution, notamment les logiciels libres, modèles, bibliothèques et données, ainsi que les licences applicables. Il les communique à la CHD. Leur utilisation est soumise à l'accord préalable et écrit de la CHD et doit être compatible avec les exigences d'exploitation, de maintenance, de réversibilité et de mutualisation prévues par le présent marché.

Le titulaire garantit disposer de tous les droits et autorisations nécessaires à la cession et aux licences prévues par la présente clause.

Les droits ainsi cédés sont protégés par les lois en vigueur au Grand-Duché de Luxembourg en matière de propriété intellectuelle et de droits d'auteur, notamment par la loi modifiée du 18 avril 2001 sur les droits d'auteur, les droits voisins et les bases de données, et demeurent la propriété exclusive de la Chambre des Députés.

La présente clause s'applique pour le monde entier et pendant toute la durée légale de protection des droits concernés.

9.9 Communication autour de la solution

Le titulaire n'est pas autorisé, sans l'accord écrit préalable de la CHD, à faire de la publicité pour les services fournis dans le cadre du présent projet. En cas de communication relative au projet, le titulaire doit mentionner, outre la CHD, le ministère de la Digitalisation et l'appel à projets. Tout document ou information échangés entre la CHD et le prestataire dans le cadre de l'exécution du contrat sont strictement confidentiels et ne peuvent donc être divulgués à des tiers. À ce titre, le prestataire devra signer un accord de confidentialité avec la CHD.

9.10 Responsabilité

9.10.1 Responsabilité délictuelle

D'une façon générale, la mise en œuvre de la responsabilité civile délictuelle se fait conformément aux articles 1382 et suivants du Code Civil luxembourgeois. Le prestataire est responsable des fautes, négligences et faits fautifs ou non de son personnel. Le prestataire s'oblige à tenir la CHD quitte et indemne de toute condamnation encourue par celle-ci notamment en vertu de l'article 1384 al 1er du Code Civil, s'il est établi que le prestataire était gardien de la chose ayant causé le dommage.

9.10.2 Responsabilité contractuelle

La mise en œuvre de la responsabilité contractuelle se fait conformément aux dispositions du Code Civil luxembourgeois. Le prestataire est notamment tenu de la garantie de conformité et de la garantie des vices cachés des matériaux livrés. Ces garanties s'appliquent à l'ensemble des livrables. Le prestataire ne peut jamais être considéré comme exécutant incompetent et servile quant aux solutions techniques préconisées dans le présent marché.

9.11 Litige

Tout litige qui pourrait naître du présent marché et qu'il n'est pas possible d'arranger à l'amiable est de la compétence exclusive des juridictions des Cours et Tribunaux de l'arrondissement judiciaire de Luxembourg. La loi applicable est exclusivement la loi luxembourgeoise.

Toute clause contractuelle du prestataire qui serait unilatérale et dérogatoire aux stipulations du présent appel d'offre et de ses annexes est réputée nulle et non avenue.

10. Annexes

10.1 Bordereau de soumission



Bordereau_Soumission.docx

10.2 Accord de confidentialité – société



NDA société XYZ.docx

10.3 Accord de confidentialité – individuel



NDA salarié de la société XYZ.docx

10.4 Accord relatif à la protection des données à caractère personnel



AccordProtectionDonnées_CHD.docx