

Secrétariat général
Service des affaires financières et immobilières

MINISTÈRE DE L'ÉCONOMIE
ET DES FINANCES

MINISTÈRE DE L'ACTION
ET DES COMPTES PUBLICS

Convention de délégation

Entre :

Le secrétariat général des ministères économiques et financiers, représenté par Mme Marie-Anne BARBAT-LAYANI, secrétaire générale, désigné sous le terme de « délégrant »,

Et :

La direction générale de la concurrence, de la consommation et de la répression des fraudes, représentée par Mme Virginie BEAUMEUNIER, directrice générale, en sa qualité de « déléataire »

Le Commissariat à l'énergie atomique et aux énergies alternatives, représenté par Alexandre BOUNOUH, directeur de l'Institut CEA LIST, désigné sous le terme « co-déléataire »,

Vu le décret 2004-1085 du 14 janvier 2004 modifié relatif à la délégation de gestion dans les services de l'Etat et le décret n° 2012-1246 du 7 novembre 2012 relatif à la gestion budgétaire et comptable publique,

Il est convenu ce qui suit :

Contexte

Le projet consiste à concevoir l'environnement de travail numérique des agents de demain, qui prendra notamment la forme d'une interface utilisable en mobilité permettant d'accéder aux données des entreprises ainsi qu'aux remontées des consommateurs. Ces outils permettront de nourrir la remontée de données économiques et de rendre plus efficaces les contrôles. Ils mettront également à disposition des enquêteurs des informations précises sur les entreprises grâce à l'intelligence artificielle afin de mieux cibler les contrôles.

Après une phase de définition collaborative du besoin, le développement débutera par la réalisation de deux prototypes : une interface mobile par des étudiants de l'Ecole Centrale Supélec et une plateforme de recueil et d'analyse des données par le commissariat à l'énergie atomique et aux énergies alternatives (CEA). Les prototypes seront développés sur la base d'expérimentations, prenant en compte les retours des utilisateurs.

Article 1^{er}

Objet de la délégation

La présente convention a pour objet d'autoriser le déléataire et le co-déléataire à consommer des crédits, hors titre 2, de l'UO 0349-CDBU-CEFI « SG ministères économiques et financiers » du BOP « Transformation action publique » du programme 349 « Fonds pour la Transformation de l'action publique », dont le responsable est la directrice du budget. Cette autorisation permet de financer, en tout ou partie, les projets pilotés et suivis par le déléataire et le co-déléataire et retenus dans le cadre du fonds pour la transformation de l'action publique (FTAP).

Le délégrant confie au déléataire et au co-déléataire, en son nom et pour son compte, dans les conditions ci-après précisées, la réalisation de l'ordonnancement des dépenses et des recettes, pour les projets désignés et dans la limite des montants fixés en annexe, de l'UO 0349-CDBU-CEFI « SG ministères économiques et financiers » du BOP « Transformation action publique » du programme 349.

La délégation emporte, du délégrant vers le déléataire et le co-déléataire, la délégation de la fonction d'ordonnateur pour l'engagement et la liquidation des dépenses ainsi que l'émission des titres de recettes.

Le déléataire et le co-déléataire sont en charge des opérations d'inventaires.

Article 2

Obligations du délégant

Le délégant assure le pilotage des autorisations d'engagement (AE) et des crédits de paiement (CP) de l'UO 0349-CDBU-CEFI « SG ministères économiques et financiers » du BOP « Transformation action publique » du programme 349. Le délégant s'engage à fournir, en temps utile, tous les éléments de suivi budgétaire dont le délégataire et le co-délégataire ont besoin.

Article 3

Obligations du délégataire et du co-délégataire

Le délégataire et le co-délégataire assurent ou font assurer les actes de gestion permettant de consommer les AE et les CP, hors titre 2, de l'UO 0349-CDBU-CEFI dans le respect des règles budgétaires et comptables.

Ils s'engagent à rendre compte trimestriellement de sa consommation de l'UO 0349-CDBU-CEFI au délégant. Ils s'engagent à lui fournir toutes les informations qui seraient nécessaires au délégant.

Dans l'hypothèse où les crédits mis à disposition par le délégant ne sont pas entièrement consommés à l'issue de la réalisation du projet validé, ils s'engagent à en informer le délégant dans les meilleurs délais. Le délégant convient de l'utilisation du solde budgétaire ainsi dégagé.

Ils sont chargés, dans le respect des règles de la commande publique en vigueur, de la passation, de la signature et de l'exécution des marchés et conventions qui s'avéreront nécessaires à la réalisation des projets en annexe. Dans ce cadre, il est expressément convenu entre les Parties que les travaux confiés par le délégataire au CEA seront réalisés dans les conditions définies dans la proposition technique référencée A19-00445 figurant en annexe 2 des présentes.

Ils adressent une copie du présent document au contrôleur budgétaire et comptable ministériel près des ministères économiques et financiers dont il relève.

Article 4

Rôles et responsabilités des parties

La DGCCRF pilote et réalise le projet objet de la présente convention dans le respect du contrat de transformation en date du contrat de transformation en date du 14 octobre 2019.

Le CEA en tant que co-délégataire a pour mission, dans le cadre d'une étude, le développement d'un logiciel permettant le recueil et l'analyse de données non-structurées, ayant pour vocation

d'aider la DGCCRF à améliorer sa compréhension de l'économie, les conditions de travail de ses agents et le ciblage de ses enquêtes. Les livrables sont détaillées en annexe 1 de la présente convention.

Article 5

Modification du document

Toute modification des conditions ou des modalités d'exécution de la présente délégation, définie d'un commun accord entre les parties, fait l'objet d'un avenant de droit dont un exemplaire est transmis au contrôleur budgétaire et comptable ministériel.

Article 6

Durée, reconduction et résiliation du document

Le présent document prend effet lors de sa signature par l'ensemble des parties concernées. La délégation est valable trois ans à compter de sa signature.

Fait le 15 janvier 2020.

Pour le secrétariat général
des ministères économiques
et financiers :

Le chef du bureau SAFI 2E,
DENIS JANKOWIAK

Pour le commissariat à l'énergie
atomique et aux énergies
alternatives :

Le directeur du List,
Institut de CEA Tech,
ALEXANDRE BOUNOUH

Pour la direction générale
de la concurrence,
de la consommation
et de la répression des fraudes :
La sous-directrice,
CORALIE OUDOT

ANNEXE 1

DESCRIPTIFS DES LIVRABLES LIÉS AU DÉVELOPPEMENT DU PROTOTYPE PERMETTANT LE RECUEIL ET L'ANALYSE DE DONNÉES NON-STRUCTURÉES

L'appui du CEA comprend les livrables suivants :

- D1.1 Rapport décrivant des fonctionnalités envisageables avec les briques technologiques CEA pour la plateforme de recueil et d'analyse de données (avec des démonstrations réalisées durant les réunions de travail) (T0+1) ;
- D1.2 Rapport contenant les spécifications des besoins de la DGCCRF en termes de traitement de données non-structurées (T0+4) ;
- D2.1 Données issues de la première collecte (T0+9) ;
- D2.2 Rapport contenant les spécifications de la collecte de données et de préconisations pour l'exploitation (T0+15) ;
- D3.1 Mise à disposition de LIMA et de l'outil de configuration CLIMA non-adaptés aux données DGCCRF (T0+6) ;
- D3.2 V1 de l'outil de configuration CLIMA et de LIMA configuré par le modèle V1 (T0+9) ;
- D3.3 V2 de l'outil de configuration CLIMA et LIMA configuré pour les données DGCCRF par le modèle V2 (T0+12) ;
- D3.4 V3 de l'outil de configuration CLIMA et LIMA configuré pour les données DGCCRF par le modèle V3 (T0+15) ;
- D4.1 Mise à disposition d'AMOSE (T0+9) ;
- D4.2 V1 AMOSE adaptée aux données DGCCRF (T0+12) ;
- D4.3 V2 AMOSE adaptée aux données DGCCRF (T0+15) ;
- D5.1 Rapport d'évaluation et description d'intégration dans la plateforme DGCCRF (T0+18).

Le tableau plus bas indique la durée des tâches et les dates prévues pour la livraison des livrables. Les livrables D2.1, D3.1, D3.2, D3.3, D3.4, D4.1, D4.2 et D4.3 seront constitués des briques technologiques à intégrer dans la plateforme DGCCRF. Les livrables D1.1, D1.2, D2.2 et D5.1 sont des rapports, décrivant pour D1.1 les fonctionnalités envisageables avec les briques technologiques du CEA pour de l'analyse de données, les besoins utilisateurs pour D1.2, D2.2 contenant les spécifications de la collecte de données et de préconisations pour l'exploitation et D5.1 décrivant le processus d'intégration, d'évaluation et des préconisations pour des évolutions futures.

Les travaux de réalisation des livrables D1.1, D3.1 et D5.1 déclenchent la possibilité pour le CEA, dans les conditions définies à l'Article 1 de la Convention, de consommer les crédits, hors titre 2, de l'UO 0349-CDBU-CEFI « SG ministères économiques et financiers » du BOP « Transformation action publique » du programme 349, tels que définis dans l'annexe financière à la présente convention (Annexe 3), selon la programmation suivante :

- 50 000 euros pour la réalisation du livrable D1.1 qui est un rapport décrivant des fonctionnalités envisageables avec les briques technologiques CEA pour la plateforme de recueil et d'analyse de données (avec des démonstrations réalisées durant les réunions de travail)
- 250 000 euros pour la réalisation du livrable D3.2, V1 de l'outil de configuration CLIMA et de LIMA configuré par le modèle V1 (T0+ 9)
- 300 000 euros pour la réalisation du livrable D5.1, Rapport d'évaluation et description d'intégration dans la plateforme DGCCRF (T0+18)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Spécification des besoins des agents	D1.1			D1.2														
Collecte de données									D2.1						D2.2			
Traitement sémantique des données non-structurées						D3.1			D3.2			D3.3			D3.4			
Moteur de recherche sémantique									D4.1			D4.2			D4.3			
Evaluation et intégration dans la plateforme																		D5.1

A N N E X E II

PROPOSITION TECHNIQUE CEA RÉFÉRENCÉE A19-00445

**Analyse de données non-structurées pour les agents de la
DGCCRF**

*PROPOSITION TECHNIQUE
D'ETUDE*

REFERENCE : A19-00445

REVISION : V6

DATE : 07/11/2019

CLIENT : DGCCRF

CHARGE D'AFFAIRE : BIANCA VIERU

RESPONSABLE TECHNIQUE : BERTRAND DELEZOIDE

DEPARTEMENT/LABORATOIRE : DIASI/SIALV/LASTI

Signature de la personne habilitée à engager l'organisme :

Date :

Tampon :



EXPRESSION DU BESOIN

Le ministre de l'économie a fixé comme feuille de route à la directrice générale de « transformer la DGCCRF ».

Dans ce but, la DGCCRF souhaite investir dans l'analyse de données en s'appuyant sur l'intelligence artificielle pour améliorer sa compréhension de l'économie, les conditions de travail de ses agents et le ciblage de ses enquêtes.

La DGCCRF souhaite que le CEA développe une plateforme de recueil et d'analyse de données structurées et non-structurées. La plateforme devra permettre de recueillir des données provenant de différentes sources notamment :

- Des sources de données internes : à partir des données structurées ou non structurées de la base métier SORA, des données de signalement des consommateurs issues de l'application *Signalconso* et des analyses d'avis de consommateurs de la base de données *Polygraphe*
- Des sources de données externes : informations publiques sur les entreprises recueillies sur internet et à terme si possible des bases de données d'administrations partenaires

La DGCCRF souhaite avoir un outil qui regroupera et analysera toutes ces sources de données non-structurées. L'objectif de la plateforme est :

- Recueillir la donnée autour des entreprises
- Structurer cette donnée par entreprise ou par secteurs d'activité

Permettre une exploitation à plusieurs niveaux : par les enquêteurs pour avoir de l'information rapide dans le cadre des contrôles, pour les rédacteurs dans le cadre de l'élaboration de la réglementation ou des enquêtes et enfin pour les datascientists pour réaliser du profilage notamment.

La plateforme de recueil et d'analyse de données non-structurées développée par le CEA s'intégrera à terme dans le système d'information de la DGCCRF.

La plateforme de recueil et d'analyse de données développée par le CEA s'adressera à l'ensemble des agents de la DGCCRF. Elle leur permettra d'avoir accès aux données internes et publiques des entreprises.

Elle permettra de trouver les antécédents de l'entreprise, le sujet et le résultat de contrôles précédents, mais aussi les remontés des consommateurs sur cette entreprise. Pour les agents, l'accès à un certain nombre d'informations sur l'environnement économique des entreprises comme leurs filiales ou les différents contacts qu'elles peuvent avoir avec d'autres entreprises, les textes juridiques applicables permettra de pré-remplir les documents devant être remplis durant les contrôles.

Ces données seront structurées et stockées dans une base de connaissance qui facilitera leurs utilisations ultérieures. Par exemple, les documents que l'enquêteur doit rédiger pendant et après le contrôle pourraient être remplis en utilisant l'information extraite de documents en lien avec le contrôle. Les agents en identifiant les éléments les plus pertinents pour un contrôle participeront à la définition des scores d'opportunité de contrôle.

Parmi les besoins des agents de la DGCCRF nous pouvons mentionner :

- Accéder aux documents du contrôle en mobilité et en temps-réel
- Regrouper les données nécessaires à l'enquête : remontées des utilisateurs et de l'application Polygraphe, procès-verbaux, textes juridiques, informations stockées dans des constatations transcrites sur papier, données presse, données d'administrations partenaires ...
- Structurer les données nécessaires à la compréhension de l'environnement des entreprises
- Récupérer, traiter et exploiter les notes prises durant les contrôles à l'aide de la dictée vocale



- Récupérer, analyser, traiter et exploiter des documents texte
- Stocker les données dans une base de connaissances permettant un accès facilité aux données
 - Recherche des données en fonction des différents critères
 - Pré-remplir des gabarits définis au préalable
- Faire des recherches en langage naturel pour trouver des documents et des paragraphes pertinents
- Suggérer des entreprises ayant le même profil qu'une entreprise donnée
- Proposer et utiliser des indicateurs pour mieux cibler les contrôles

DESCRIPTION DES TRAVAUX

L'objectif de ce projet est de fournir un logiciel permettant aux agents de la DGCCRF de réunir rapidement des informations utiles à l'élaboration et au ciblage des enquêtes à partir de sources textuelles non-structurées internes et publiques.

Le CEA propose le développement d'un logiciel permettant le recueil et l'analyse de données non-structurées utilisées dans le cadre des missions de la DGCCRF. Ce logiciel permettra de mettre à disposition des documents sous forme synthétique en rapport avec les entreprises, de filtrer ces données, les structurer et les stocker dans une base de connaissance permettant une réutilisation exploitable et facile

L'intégration des briques technologiques développées par le CEA dans la plateforme DGCCRF permettra d'implémenter plusieurs fonctionnalités de traitement automatique du langage (TAL). Celles-ci seront précisées dans la phase de spécification. Elles pourraient notamment rendre possible pour les agents de :

- Récupérer des données publiques sur internet
- Structurer les données nécessaires aux missions CCRF
- Classifier les documents
- Stocker les documents dans une base de connaissance permettant un accès facilité aux données
- Structurer les rapports
- Accéder par une requête en langage naturel aux documents en lien avec le contrôle en préparation ou en cours
- Filtrer les documents en fonction des différents critères
- Faire des recherches en langage naturel pour trouver des documents et des paragraphes pertinents
- Suggérer des entreprises ayant le même profil qu'une entreprise donnée
- Pré-remplir un certain nombre de documents d'enquête à partir des informations recueillies

OBJECTIFS TECHNIQUES

L'objectif de ce projet est de développer et évaluer un logiciel permettant aux agents de la DGCCRF d'accéder aux données utiles à leurs missions. Ce logiciel se fondera sur l'outil existant de traitement de la langue du CEA LIST, LIMA (Libre Multilingual Analyser) qui sera adapté aux besoins des agents de la DGCCRF. Les travaux de développement consisteront à :

1. Comprendre les besoins métiers des agents de la DGCCRF et rédiger un rapport contenant la spécification des besoins que la technologie IA doit résoudre ou améliorer
2. Déterminer et inventorier les sources de données à analyser, identifier la qualité des données disponibles et déterminer les modalités de récupération des données internes ou externes



3. Choisir la chaîne de traitement optimale pour le traitement des documents textuelles internes et publiques ¹
4. Configurer/apprendre les chaînes de traitement permettant l'extraction et le « rendu » d'information utile et exploitable aussi bien en veille, recherche, que visualisation
5. Mettre en place un processus d'évaluation des chaînes de traitement
6. Mettre en place et évaluer un prototype par rapport à sa conformité aux besoins
7. Développer et mettre en place un prototype
8. Prendre en compte les retours utilisateurs sur le prototype et accompagner la DGCCRF au déploiement du logiciel dans l'environnement cible
9. Réaliser une étude expliquant les choix techniques, les résultats de différentes évaluations et des préconisations pour des évolutions futures.
10. Former et assurer le transfert de connaissances vers les agents de la DGCCRF.

CONTRAINTES

Le succès du projet sera influencé par la disponibilité des utilisateurs tant dans la phase de spécification des besoins que dans les phases de développement. En effet, les utilisateurs doivent contribuer à la définition des fonctionnalités du logiciel afin qu'il corresponde au mieux à leurs besoins.

Les agents de la DGCCRF ont une connaissance fine de l'économie et des entreprises. Dans le cadre du projet, ils pourront être sollicités pour créer l'inventaire de sources de données auxquelles ils souhaitent accéder. Leurs expériences et connaissances seront également utiles pour définir les informations à extraire des documents, et si nécessaire, pour faire des annotations manuelles ou aider à la création de listes de concepts.

Les agents seront sollicités également pour évaluer les différentes versions du logiciel autant du point de vue de l'interface, mais surtout du point de vue des performances des briques technologiques. En fonction des corpus disponibles, cette évaluation sera quantitative (si assez des corpus annotés disponibles) ou qualitative.

Le SICCRF sera sollicité pour définir comment les briques technologiques développées par le CEA communiqueront et s'intégreront avec les bases de données, les logiciels existants et dans la plateforme DGCCRF.

DECOUPAGE EN TACHES

Les travaux proposés dans le cadre de ce projet se découpent en cinq tâches. Après une première phase de spécification qui servira de socle pour les autres, les travaux des tâches de développement se dérouleront en parallèle et seront intégrés dans la plateforme DGCCRF.

SPECIFICATIONS DES BESOINS DES AGENTS

Description des travaux : les spécifications visent à décrire les besoins de la DGCCRF en termes de traitement de données non-structurées et voir comment les développements peuvent répondre aux besoins exprimés par les utilisateurs. Durant cette période, pourront être définis des éléments comme :

- les sources de données à traiter,
- la nature des documents qui doivent être pris en compte pour extraire l'information (remontées d'utilisateurs, procès-verbaux, textes juridiques, informations stockées dans de constatations transcrites sur papier, données presse et données d'administrations partenaires, données issues de la base métier SORA ...)

¹ Cette chaîne pourrait intégrer à terme des traitements pour les données multimédia/images/videos/audios



- les informations qui doivent être extraites de chaque type de document. Par exemple, à partir d'une remontée utilisateur peuvent être extraits le nom de l'entreprise et la constatation, à partir des informations stockées dans de constatations transcrites sur papier les principales caractéristiques de l'entreprise : l'organigramme, responsabilités des dirigeants, activités, fournisseurs, sous-traitants, chiffres d'affaires, factures, typologie et volume des produits mis sur le marché, part dans le CA...
- les gabarits pour les documents à pre-remplir
- les modalités de récupération des données internes et externes

Durant la phase de spécification, le CEA proposera à la DGCCRF un certain nombre de fonctionnalités et de scénarii d'usage envisageables à partir des briques technologiques développées par le CEA. Ces propositions seront détaillées dans le livrable D1.1, qui sera complété par des démonstrations basées sur les technologies du CEA et illustrant leurs avantages et limites pour la plateforme d'analyse de données de la DGCCRF.

Plusieurs cas d'usage seront identifiés avec des agents de l'administration centrale et des services d'enquête. Pour chacun de ce cas seront précisés les résultats attendus par les agents. Ces cas d'usage seront décrits dans le livrable D1.2 et seront utilisés pour l'évaluation des briques technologiques dans la tâche 5.

Cette tâche permettra d'avoir une idée précise de la complexité du problème posé, d'identifier les différentes briques à développer et inclure dans le logiciel, mais aussi voir s'il y a des besoins en termes de R&D plus spécifiques.

Durant cette tâche seront abordées également les pistes d'industrialisation des briques proposés par le CEA et les détails liées à l'intégration de ces briques dans la plateforme DGCCRF.

Elle aboutira à la restitution d'un rapport décrivant les besoins DGCCRF et les techniques nécessaires pour répondre à ces besoins. Le rapport pourrait contenir des premières pistes d'architecture qui seraient explorées pour répondre aux besoins issus de cette étude.

Jalons de remise de livrables Dij:

D1.1 Rapport décrivant des fonctionnalités envisageables avec les briques technologiques CEA pour la plateforme de recueil et d'analyse de données (avec des démonstrations réalisées durant les réunions de travail)

D1.2 Rapport contenant les spécifications de besoins de la DGCCRF en termes de traitement de données non-structurées (T₀+4)

COLLECTE DE DONNEES

Description des travaux : sur la base des spécifications développées dans la tâche précédente, cette tâche a pour objectifs :

1. La collecte effective de l'ensemble des données nécessaires à l'évaluation et à la démonstration du prototype d'analyse de données de la DGCCRF.
2. L'évaluation de la faisabilité de l'automatisation de ces collectes en exploitation.

Pour une part de cette tâche, le CEA fera appel à un fournisseur extérieur d'outils et de prestation de Crawl, probablement KBCrawl. Les conditions de contractualisation sont encore à définir. Le financement de la prestation est prévu dans l'enveloppe de cette tâche. Le fournisseur mettra en place ces outils, a minima, pour les données externes disponibles en sources ouvertes (payant ou non).

Jalons de remise de livrables Dij:

D2.1 Données issues de la première collecte (T₀+9)



D2.2 Rapport contenant les spécifications de la collecte de données et de préconisations pour l'exploitation (T₀+15)

TRAITEMENT SEMANTIQUE DES DONNEES NON-STRUCTUREES

Description des travaux : les briques technologiques développées dans cette tâche permettront de faire l'extraction d'information et la classification de documents. Pour cela, elles se baseront sur un processus de traitement sémantique des données par annotation automatique de documents. L'annotation consiste à associer automatiquement de l'information aux documents afin de les structurer pour faciliter leur analyse par un outil de gestion des connaissances (veille, recherche, visualisation). Le besoin d'annoter des entités, relations, événements, thèmes, opinions sera déterminé pendant la phase de spécification.

Dans un premier temps, ces briques se baseront sur une ontologie généraliste. Pendant le projet, cette ontologie sera adaptée au domaine des contrôles financiers en utilisant les données définies dans la tâche de spécification.

Les travaux mis en place dans le cadre de cette tâche incluent :

- l'adaptation à un domaine spécialisé : cette adaptation permettra de prendre en compte dans l'extraction d'information (concepts, relations, événements) les connaissances existantes au sein de la DGCCRF. Elle pourrait être faite par extraction par règles et/ou par génération de règles à partir d'annotations manuelles ou de listes pour les concepts. Cette adaptation se basera et s'intégrera dans la brique d'analyse linguistique (LIMA) qui permettra de transformer les données non-structurées en données structurées utiles et exploitables.
- la mise en place d'un configurateur : permettre aux utilisateurs de configurer eux-mêmes l'extraction d'information en fonction de leurs besoins (CLIMA). De manière continue, ce processus d'aide à la création d'une ontologie sera mis en place pour le domaine du projet. Il intégrera sur la durée de la tâche les besoins de configuration apportés par l'arrivée de nouveaux documents, informations ou fonctionnalités de traitement de données.

D'autres briques technologiques pourraient être ajoutées aux briques technologiques mentionnées si les besoins définis dans l'étape de Spécification le demande ou si durant les évaluations menées auprès des agents de la DGCCRF un nouveau besoin apparaîtra.

Les briques technologiques développées dans le cadre de cette tâche seront mises à disposition en vue de leurs intégrations dans la plateforme DGCCRF régulièrement selon les jalons donnés plus bas.

Jalons de remise de livrables Dij :

D3.1 Mise à disposition de LIMA et de l'outil de configuration CLIMA non-adaptés aux données DGCCRF (T₀+6)

D3.2 V1 de l'outil de configuration CLIMA et de LIMA configuré par le modèle V1 (T₀+9)

D3.3 V2 de l'outil de configuration CLIMA et LIMA configuré pour les données DGCCRF par le modèle V2 (T₀+12)

D3.4 V3 de l'outil de configuration CLIMA et LIMA configuré pour les données DGCCRF par le modèle V3 (T₀+15).

MOTEUR DE RECHERCHE SEMANTIQUE

Description des travaux : dans le cadre de cette tâche sera mis en place un moteur de recherche sémantique en langage naturel. Ce moteur de recherche permettra d'obtenir des documents multimédia répondant à un besoin de l'utilisateur défini par une requête en langage naturel. Il se



fondera sur le traitement des documents et sur l'ontologie de connaissance pour proposer à l'utilisateur des documents pertinents rangés en classes thématiques définies dans l'ontologie.

Le CEA adaptera son moteur de recherche AMOSE aux besoins de la DGCCRF. Le moteur de recherche se basera sur une base de connaissance qui sera alimentée par des documents issues des bases de données internes à la DGCCRF, des documents publics et internes provenant des sources identifiées lors de la tâche de spécification. Chaque document sera ensuite annoté par l'outil de traitement de données de la tâche précédente et sera indexé pour être intégré dans la base de connaissance.

Les briques suivantes seront intégrées dans la plateforme de recherche sémantique :

- La plateforme LIMA développée durant la tâche « Traitement sémantique des données non-structurées »
- La plateforme d'indexation/recherche AMOSE. Cette brique permettra par une requête en langage naturel de rechercher des informations au sein de l'index réalisé par la brique précédente.

Comme pour la tâche précédente, d'autres briques technologiques pourraient être ajoutés en fonction des spécifications définis.

Les briques technologiques développées dans le cadre de cette tâche seront mises à disposition en vue de leurs intégrations dans la plateforme DGCCRF régulièrement selon les jalons donnés plus bas.

Jalons de remise de livrables Dij:

D4.1 Mise à disposition d'AMOSE (T_0+9)

D4.2 Première version d'AMOSE adaptée aux données DGCCRF (T_0+12)

D4.3 Deuxième version d'AMOSE adaptée aux données DGCCRF (T_0+15)

EVALUATION ET INTEGRATION DANS LA PLATEFORME

Description des travaux : l'objectif de cette tâche est de préparer les briques technologiques d'analyse linguistique, LIMA, le configurateur CLIMA et le moteur de recherche, AMOSE, pour qu'ils soient intégrables dans la plateforme DGCCRF. L'intégration de ces briques dans la plateforme DGCCRF commencera à partir de T_0+9 avec la livraison d'une première version de LIMA/CLIMA. La livraison de plusieurs versions est prévue afin de s'assurer que les développements techniques sont en accord avec les besoins utilisateurs. Tout au long du projet, le CEA accompagnera la DGCCRF dans l'intégration des briques technologiques au sein du système d'information de la DGCCRF.

Les développements seront évalués tout au long du projet. Les performances seront évaluées selon les cas d'usage définis lors de la phase de spécification en étroite liaison avec les utilisateurs. Cela permettra de prendre en compte les retours utilisateurs dans la phase de développement et ces évaluations indiqueront les progrès fait.

Pendant le projet il y aura des réunions régulières permettant d'assurer la formation et le transfert de connaissance vers les utilisateurs. Ces informations figureront aussi dans la documentation finale.

Le processus d'intégration des briques technologiques dans la plateforme DGCCRF implémenté, les évaluations de ces briques et de préconisations pour des adaptations futures seront décrites dans le rapport final d'évaluation et d'intégration.

Jalons de remise de livrables Dij:

D5.1 Rapport d'évaluation et description d'intégration dans la plateforme DGCCRF (T_0+18)



LIMITES DE FOURNITURE

La fourniture comprend :

D1.1 Rapport décrivant des fonctionnalités envisageables avec les briques technologiques CEA pour la plateforme de recueil et d'analyse de données (avec démonstrations réalisées lors des réunions de travail) (T₀+1)

D1.2 Rapport contenant les spécifications des besoins de la DGCCRF en termes de traitement de données non-structurées (T₀+4)

D2.1 Données issues de la première collecte (T₀+9)

D2.2 Rapport contenant les spécifications de la collecte de données et de préconisations pour l'exploitation (T₀+15)

D3.1 Mise à disposition de LIMA et de l'outil de configuration CLIMA non-adaptés aux données DGCCRF (T₀+6)

D3.2 V1 de l'outil de configuration CLIMA et de LIMA configuré par le modèle V1 (T₀+9)

D3.3 V2 de l'outil de configuration CLIMA et LIMA configuré pour les données DGCCRF par le modèle V2 (T₀+12)

D3.4 V3 de l'outil de configuration CLIMA et LIMA configuré pour les données DGCCRF par le modèle V3 (T₀+15).

D4.1 Mise à disposition d'AMOSE (T₀+9)

D4.2 V1 AMOSE adaptée aux données DGCCRF (T₀+12)

D4.3 V2 AMOSE adaptée aux données DGCCRF (T₀+15)

D5.1 Rapport d'évaluation et description d'intégration dans la plateforme DGCCRF (T₀+18)

Le tableau plus bas indique la durée des tâches et les dates prévues pour la livraison des livrables. Les livrables D2.1, D3.1, D3.2, D3.3, D3.4, D4.1, D4.2 et D4.3 seront constitués des briques technologiques à intégrer dans la plateforme DGCCRF.

Les livrables D1.1, D1.2, D2.2 et D5.1 sont des rapports, décrivant pour D1.1 les fonctionnalités envisageables avec les briques technologiques du CEA pour de l'analyse de données, les besoins utilisateurs pour D1.2, D2.2 contenant les spécifications de la collecte de données et de préconisations pour l'exploitation et D5.1 décrivant le processus d'intégration, d'évaluation et des préconisations pour des évolutions futures.

ListD/Q/FA7 Rev5 du 20 septembre 2016

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Spécification des besoins des agents	D1.1			D1.2														
Collecte de données									D2.1						D2.2			
Traitement sémantique des données non-structurées						D3.1			D3.2			D3.3			D3.4			
Moteur de recherche sémantique									D4.1			D4.2			D4.3			
Evaluation et intégration dans la plateforme																		D5.1



La fourniture ne comprend pas :

Le CEA ne fournit pas le service d'OCR des textes scannés. Les briques technologiques développées prennent en entrée des textes numérisés dans un format standard (type XML/utf8, xslx par exemple).

DELAI DE REALISATION DES ETAPES CLEFS

Les travaux décrits ci-dessus pourraient être réalisés dans un délai de 18 mois selon l'échéancier suivant :

Action	Délai (mois)
Rapport décrivant des fonctionnalités envisageables avec les briques technologiques CEA pour la plateforme de recueil et d'analyse de données (avec des démonstrations réalisées lors des réunions de travail)	T ₀ + 1
Rapport contenant les spécifications des besoins de la DGCCRF en termes de traitement de données non-structurées	T ₀ + 4
Version de base de LIMA, CLIMA	T ₀ + 6
Deuxième version de LIMA, CLIMA et AMOSE	T ₀ + 12
Troisième version de LIMA, CLIMA et AMOSE	T ₀ + 15
Rapport d'évaluation et description d'intégration dans la plateforme DGCCRF	T ₀ + 18

T₀ : date de notification de la commande



CONNAISSANCES ANTERIEURES UTILISEES POUR CETTE ETUDE

(brevets, savoir-faire secrets, logiciels, algorithmes,...)

Pour le CEA :

- Logiciel CEA - Moteur de recherche multimédia : AMOSE R16857-20121108 (LIMA 01.370085.00.S.P.2007.000.10600) 002 du 15/12/2012 : dépôt du 20/03/2013 à l'Agence pour la Protection des Programmes sous le numéro IDDN.FR.001.12007.000.S.C.2013.000.20900
- Logiciel CEA - Analyseur linguistique LIMA : LIMA R16857-20121108 du 15/12/2012 : dépôt du 20/03/2013 à l'Agence pour la Protection des Programmes sous le numéro IDDN.FR.001.370085.001.S.P.2007.000.10600
- Jorge García, Flores Olivier Ferret, and Gaël de Chalendar. Summarizing through sense concentration and Contextual Exploration rules: the CHORAL system at TAC. In Proceedings of the Second Text Analysis Conference (TAC 2009), November 2009.
- Regroupement automatique de documents (clustering) : Logiciel de clustering Ref : DIASI/13-853 et DIASI/15-768
- Méthodes Statistiques pour la Reconnaissance d'Entités Nommées et l'Adaptation Rapide à un Nouveau Domaine, Rapport de stage de Wei Wang, 2009.
- Extraction non supervisée de signatures thématiques structurées à partir de textes. Thèse de Ludovic Jean Louis, 2011.
- Extraction d'information générique à partir de textes fondée sur une analyse linguistique profonde. Thèse de Wei Wang, 2012.
- Rapport DIASI 2016-387 Outil d'extraction d'entités à partir d'ontologie existantes (UMLS).
- Rapport technique DIASI 2016-589 présentant le fonctionnement de CLIMA pour l'adaptation automatique de l'extraction d'information à de nouveaux domaines métier.



Conditions générales d'exécution d'une étude

ARTICLE 1 - OBJET ET CHAMP D'APPLICATION

Le présent document a pour objet de déterminer les conditions auxquelles sera soumise l'étude effectuée par le CEA pour le compte du délégataire (ci-après la ou les partie(s)).

ARTICLE 2 - CONCLUSION DU CONTRAT - PIÈCES CONTRACTUELLES

Le contrat est conclu et entre en vigueur à la date de signature de la présente convention.

ARTICLE 3 - DURÉE - DELAIS D'EXECUTION

Le Contrat prendra fin à l'exécution du dernier livrable, sauf résiliation anticipée au titre de l'article 7 ci-après.

Les articles 4, 8, 9 et 10 survivront à l'expiration du Contrat ou à sa résiliation, pour quelle que cause que ce soit et pour la durée qui leur est propre.

En tout état de cause, les engagements du CEA relatifs aux délais s'entendent sous réserve du respect par le délégataire de ses propres obligations, notamment en ce qui concerne la fourniture en temps utile des documents, renseignements ou produits nécessaires à l'exécution de l'étude.

ARTICLE 4 - CONFIDENTIALITE

4.1 Chaque partie s'engage à considérer comme confidentielles toutes les informations ou donnée financière, commerciale, technique, juridique ou de toute autre nature (ci-après désignées par Informations Confidentielles), communiquée par la partie divulgateur à la partie récipiendaire dans le cadre de la dite étude, sous forme écrite, ou orale, ou résultant de visites de locaux, sous forme d'échantillons, de dessins, de modèles, de programmes informatiques ou sous toute autre forme, ayant été expressément identifiée comme confidentielle par la partie divulgateur au moment de la divulgation et par l'apposition de la mention 'confidentiel' sur le support matériel de l'information divulguée ou, si l'information a été communiquée oralement ou visuellement, ayant été désignée comme confidentielle au moment de la divulgation et confirmée comme telle par écrit par la partie divulgateur dans un délai de trente (30) jours à compter de ladite divulgation étant entendu que l'information sera considérée comme une Information Confidentielle pendant cette période de trente (30) jours.

La partie réceptrice des Informations Confidentielles devra prendre toutes les précautions nécessaires pour qu'elles ne soient communiquées à aucun tiers et à ne transmettre à son personnel que la partie des Informations Confidentielles qui lui est strictement nécessaire pour l'exécution de l'étude.

4.2 Chaque partie s'engage à ne pas communiquer d'Informations Confidentielles à des tiers, sans l'accord préalable et écrit de l'autre partie, et à prendre toutes les dispositions nécessaires pour faire respecter par son personnel, et les tiers éventuellement autorisés, la confidentialité des dites Informations.

4.3 En tout état de cause, chaque partie s'engage à ne pas utiliser ces Informations Confidentielles pour quelque usage que ce soit, autre que celui pour lequel la partie émettrice les a communiquées.

4.4 La présente obligation de confidentialité ne s'appliquera pas à toute Information Confidentielle ou élément d'Information Confidentielle dont une partie est en mesure d'établir :

- qu'elle était accessible au public à la date de sa communication par l'une des parties, ou a été rendue accessible au public en l'absence de toute faute imputable à la partie l'ayant reçue,
- qu'elle était déjà connue de la partie réceptrice au moment de la communication,
- qu'elle lui a été transmise sans obligation de confidentialité par un tiers la détenant légitimement,
- qu'elle a été obtenue par la partie réceptrice par des développements indépendants entrepris de bonne foi par des membres de son personnel n'ayant pas eu accès aux Informations Confidentielles.

4.5 La présente obligation de confidentialité survivra à l'expiration du Contrat ou à sa résiliation anticipée pendant une durée de cinq (5) ans.

ARTICLE 5 - PROPRIETE DES LIVRABLES

LIVRABLES : tout élément matériel (notamment rapports, démonstrateur, maquettes, etc.) à l'exclusion de toute CONNAISSANCE, devant être réalisé et fourni par le CEA au délégataire conformément à ce qui a été prévu dans l'annexe technique de la étude.

Les LIVRABLES appartiennent au délégataire. Le délégataire reconnaît expressément que les LIVRABLES, résultant des travaux réalisés, ont une qualité et une finalité de nature expérimentale et n'ont pas été élaborés ni n'ont vocation à être utilisés à d'autres fins que de la R&D. Toute utilisation par le délégataire d'un LIVRABLE, dont la propriété matérielle lui est cédée en vertu du présent article, s'effectue à ses risques et périls. Sauf accord entre les Parties, le CEA n'assurera aucune prestation de maintenance des LIVRABLES.

ARTICLE 6 - DROIT APPLICABLE - JURIDICTION COMPETENTE

L'étude est soumise à la loi française et les litiges s'y rapportant que les parties n'auraient pu résoudre à l'amiable, dans un délai de six (6) mois à compter de la survenance du litige, seront soumis à la juridiction des tribunaux compétents de Paris.

ARTICLE 7 - RESILIATION

Le Contrat pourra être résilié de plein droit par chacune des parties en cas de manquement ou d'inexécution par l'autre partie de l'une quelconque de ses obligations au titre des présentes. Cette résiliation ne deviendra effective que trente (30) jours après l'envoi par la partie non défaillante d'une lettre recommandée avec avis de réception exposant ses motifs à moins que dans ce délai, la partie défaillante n'ait satisfait à ses obligations ou n'ait apporté la preuve d'un empêchement consécutif à un cas de force majeure.

ARTICLE 8 - PROPRIETE DES CONNAISSANCES

8.1 Définitions :

- CONNAISSANCES : désigne toute connaissance brevetable ou non, brevetée ou non, y compris logiciel, savoir-faire, secret de fabrique ainsi que les droits de propriété intellectuelle y afférents. Il est précisé que les CONNAISSANCES s'entendent hors LIVRABLES.

- CONNAISSANCES ANTERIEURES : désigne les CONNAISSANCES nécessaires à la réalisation de l'étude et qui appartiennent à l'une ou l'autre des parties avant la date de démarrage de l'étude. Les CONNAISSANCES ANTERIEURES du/des laboratoire(s) du CEA impliqué(s) dans la réalisation de l'étude ainsi que celles du délégataire sont listées aux présentes.

- CONNAISSANCES NOUVELLES : toutes CONNAISSANCES résultant des travaux effectués en exécution de l'étude.

8.2 Propriété des CONNAISSANCES ANTERIEURES :

Chacune des parties conserve la pleine et entière propriété de ses CONNAISSANCES ANTERIEURES.

8.3 Propriété des CONNAISSANCES NOUVELLES :

Les CONNAISSANCES NOUVELLES sont la propriété du CEA.

ARTICLE 9 - EXPLOITATION DES CONNAISSANCES

ANTERIEURES CEA ET DES CONNAISSANCES NOUVELLES

Avant toute exploitation commerciale et industrielle par le délégataire des CONNAISSANCES NOUVELLES, un contrat de licence devra être conclu par les parties. Sur demande écrite formulée par le délégataire au plus tard dans les trois (3) mois qui suivront la date de fin d'exécution de l'étude tel que prévu en article 3 ci-dessus, le contrat de licence sera négocié de bonne foi, à des conditions commerciales normales et sera non exclusif.

Si nécessaire à l'exploitation visée des CONNAISSANCES NOUVELLES, ce contrat pourra inclure les CONNAISSANCES ANTERIEURES du CEA, sous réserve d'éventuels droits préexistants de tiers et des engagements préexistants du CEA.

Il est précisé que cette option de licence est conditionnée par le respect par le délégataire de ses obligations de paiement telles que prévues à l'article 4 ci-dessus.

A l'issue de ce délai de trois (3) mois, l'engagement susvisé prendra fin.

ARTICLE 10 - LIMITATION DE RESPONSABILITE

10.1 L'engagement du CEA est limité à une obligation de moyens, à savoir la réalisation de l'étude suivant les règles de l'art, en fonction de ses connaissances et de son expérience au moment de l'étude et au regard des spécifications fournies par le délégataire.

10.2 Le délégataire mettra en œuvre les CONNAISSANCES NOUVELLES et les CONNAISSANCES ANTERIEURES du CEA à ses risques et périls. Le CEA ne garantit pas la faisabilité industrielle des opérations découlant de l'étude, ni la performance technique de ses CONNAISSANCES. Le délégataire est seul responsable, notamment vis-à-vis de ses propres délégataires, de ses activités mettant en œuvre lesdites CONNAISSANCES.

CONDITIONS FINANCIÈRES

Direction/Etablissement	Projet	T3	Calendrier			UO	Activité
			2019	2020	2021		
DGCCRF	Transformer les modalités de contrôle de la DGCCRF en s'appuyant sur l'intelligence artificielle et les signalements des consommateurs	AE	0	825	1 180	0349-CDBU-CEFI	034901013601
		CP	0	825	1 180		
CEA	Sous projet : développement du logiciel permettant le recueil et l'analyse de données non-structurées	AE	0	600	0		
		CP	0	300	300		
Total		AE	0	1 425	1 180		
		CP	0	1 125	1 480		