



Marchés publics de services

## **Refonte du système d'information relatif aux données de publications scientifiques du Hcéres**

Cahier des clauses techniques particulières (CCTP)

**Numéro de consultation :** 2019-AOO-DSI-02

**Forme du marché :** marché public ordinaire

**Procédure :** La procédure retenue est celle de l'appel d'offre ouvert (AOO) en application de l'article R. 2124-2 du code de la commande publique

# SOMMAIRE

<b>Article 1 - Objet de la consultation</b>	<b>3</b>
<b>Article 2 - Contexte et objectifs</b>	<b>3</b>
2.01 Le Hcéres	3
2.02 Organisation générale du Hcéres	4
2.03 Le système d'information du Hcéres	5
<b>Article 3 - Présentation du projet</b>	<b>8</b>
3.01 Contexte	8
3.02 Démarche projet	9
3.03 Pilotage et suivi	9
3.04 Description du périmètre du projet	10
<b>Article 4 - Description des missions attendues</b>	<b>12</b>
4.01 Éléments communs à toutes les missions	12
4.02 Éléments communs aux missions de conception des référentiels	13
4.03 Mission 1 : Conception du référentiel géographique	14
4.04 Mission 2 : Conception du référentiel institutionnel	18
4.05 Mission 3 : Conception du référentiel des supports de publication	23
4.06 Complément d'information : référentiel disciplinaire, pris en charge par le Hcéres	26
4.07 Mission 4 : Conception du modèle relationnel métier	26
4.08 Mission 5 : Socle technique de développement, intégration et validation	28
4.09 Mission 6 : Mise en œuvre du modèle métier	29
4.10 Mission 7 : Mise en œuvre du référentiel géographique	29
4.11 Mission 8 : Mise en œuvre du référentiel institutionnel	29
4.12 Mission 9 : Mise en œuvre du référentiel des supports de publication	30
4.13 Mission 10 : Mise en œuvre des interfaces utilisateur	30
<b>Article 5 - Exigences techniques</b>	<b>30</b>
5.01 Exigences techniques générales	30
5.02 Exigences techniques relatives aux interfaces utilisateur	31
5.03 Éléments de volumétrie	32
<b>Article 6 - Prestations et livrables attendus</b>	<b>32</b>
<b>Article 7 - Calendrier prévisionnel du projet</b>	<b>34</b>
<b>Article 8 - Garantie</b>	<b>34</b>
<b>Article 9 - Cadre de réponse</b>	<b>34</b>

## ARTICLE 1 - OBJET DE LA CONSULTATION

---

La présente consultation a pour objet de définir, construire et mettre en place une organisation des données et des processus de mise en qualité des données de bibliographie issues des fichiers du Web of Science (WoS) acquis auprès de Clarivate Analytics et livrés au format XML. Ces travaux doivent conduire à rendre l'organisation existante plus efficiente et plus robuste afin de servir au mieux les besoins de production d'indicateurs et d'analyses bibliométriques du département Observatoire des Sciences et Techniques (OST) du Hcéres. Les modèles définis devront garantir la qualité des données et pouvoir s'adapter à l'évolution des besoins.

Le titulaire aura la charge de concevoir et définir avec les équipes du Hcéres l'organisation des données et les processus inhérents à l'intégration de celles-ci à partir des différentes livraisons réalisées par le fournisseur. Cette conception concerne tout le cycle de constitution des référentiels nécessaire à la mise en qualité ainsi que tous les processus de mise en qualité des données sur la base des référentiels et des enrichissements qui peuvent intervenir tout au long du processus global. Le titulaire assurera ensuite la mise en œuvre des solutions définies avec les équipes du Hcéres.

## ARTICLE 2 - CONTEXTE ET OBJECTIFS

---

### 2.01 Le Hcéres

Le Haut Conseil de l'Évaluation de la Recherche et de l'Enseignement Supérieur (Hcéres), créé par la loi du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche, a remplacé l'Agence d'Évaluation de la Recherche et de l'Enseignement Supérieur (AERES) depuis le 17 novembre 2014. Les membres de son collège et son président ont été nommés le 30 octobre 2015 pour un mandat de 4 ans.

Autorité administrative indépendante, le Hcéres a 5 missions :

- évaluer les structures d'enseignement supérieur et de recherche en France ;
- valider les procédures d'évaluation proposées par d'autres instances ;
- évaluer les établissements d'enseignement supérieur étrangers qui en font la demande ;
- produire des indicateurs et des analyses des systèmes de recherche et d'innovation en s'appuyant sur les travaux de son département Observatoire des Sciences et Techniques (OST) ;
- contribuer à la définition d'une politique nationale de l'intégrité, observer les pratiques et accompagner les acteurs dans la mise en œuvre de leurs obligations, en s'appuyant sur son département Office Français de l'Intégrité Scientifique (Ofis).

Inscrivant son action dans le processus de Bologne<sup>1</sup>, le Hcéres participe à la construction de l'espace européen de l'enseignement supérieur et fonde son action sur les meilleures pratiques internationales et européennes en matière d'assurance qualité. À ce titre, il est membre de l'association européenne des agences d'assurance qualité pour l'enseignement supérieur, Enqa<sup>2</sup>, et est inscrit au registre européen, Eqar<sup>3</sup>.

Le Hcéres travaille en coopération avec les acteurs français du domaine : Cti, CEFDG, CCN-IUT<sup>4</sup>.

Dans sa mission de service public, le Hcéres conçoit ses évaluations et ses analyses, formulées par des experts, comme des outils de développement des entités évaluées. Son objectif est d'accompagner, de conseiller et de soutenir la démarche d'amélioration de la qualité de l'enseignement supérieur et de la recherche en France.

---

<sup>1</sup> <http://www.enseignementsup-recherche.gouv.fr/cid56043/presentation-de-l-e.e.e.s.html>

<sup>2</sup> European Association for Quality Assurance in Higher Education

<sup>3</sup> European Quality Assurance Register for Higher Education

<sup>4</sup> Cti, commission des titres d'ingénieur ; CEFDG, commission d'évaluation des formations et diplômes de gestion ; CCN-IUT, commission consultative nationale des IUT

Quatre valeurs essentielles guident l'action du Hcéres : l'indépendance, l'absence de conflits d'intérêts, la transparence et l'équité. Le Hcéres rend publics sa méthodologie, ses procédures, la composition des comités d'évaluation, les rapports d'évaluation et les analyses. Il assure à chaque entité évaluée un traitement identique dans la procédure, quel que soit l'établissement ou le territoire concerné.

Le Haut Conseil réalise l'ensemble des évaluations dont il a la charge en 5 vagues successives (Vague A, B, C, D et E), au rythme fixé par la politique contractuelle de l'État qui a réparti les établissements concernés en 5 zones géographiques.

## 2.02 Organisation générale du Hcéres

Le Hcéres est administré par un collège composé à parité de 30 membres, dont le président. Le collège arrête le programme annuel d'évaluation et définit les mesures propres à garantir la qualité, la transparence et la publicité des procédures mises en œuvre par le Haut Conseil. À ses côtés, siègent le Conseil d'orientation scientifique de l'OST et le Conseil de l'intégrité scientifique.

Le Hcéres est organisé en 8 départements<sup>5</sup>, dont 4 départements d'évaluation (coordinations territoriales/établissements/recherche/formations), un département Europe et international, un département du système d'information, l'OST et l'Ofis, auxquels s'ajoute un secrétariat général. Ce dernier est composé des services et délégations supports : budget-finances, ressources humaines, logistique, communication, qualité et formation, affaires juridiques.

Le Hcéres compte 225 personnels administratifs et scientifiques, dont une centaine de conseiller(ère)s scientifiques (chercheurs ou enseignants-chercheurs). Travaillant à temps partiel, ces derniers sont en charge de l'organisation scientifique des évaluations. Sous l'autorité du directeur de leur département, ils participent à la constitution des comités d'experts chargés de l'évaluation et suivent le déroulement de celle-ci. L'évaluation est principalement conduite par des pairs (chercheurs et enseignants-chercheurs) et par des experts d'autres horizons (étudiants, professionnels issus du secteur privé ou public, etc.). La composition des comités d'experts varie en fonction de la nature spécifique de l'entité évaluée. Chaque année, le Hcéres fait appel en moyenne à 3 500 experts français et étrangers pour conduire les évaluations.

### (a) L'Observatoire des Sciences et Techniques (OST)

Créé en 1990, l'Observatoire des Sciences et Techniques est un département du Hcéres depuis 2015. Ses analyses nourrissent la réflexion stratégique des parties prenantes de l'enseignement supérieur et de la recherche et contribuent à l'évaluation de l'impact des politiques publiques.-

Ses travaux comprennent : la production régulière d'indicateurs et d'analyses sur la recherche scientifique et technologique, la réalisation d'analyses en appui à des évaluations du Hcéres, la conduite de projets spécifiques et la réponse à des commandes externes. Afin de remplir ses missions, l'OST fiabilise et enrichit des données issues de diverses sources (Web of Science de Clarivate Analytics, Patstat database de l'Office européen des brevets, base e-corda de la Commission européenne).-

Les indicateurs scientométriques et les analyses des systèmes de recherche et d'innovation contribuent notamment à :

- o la connaissance des activités de recherche et d'innovation en France et dans le monde ;
- o l'évaluation des activités de recherche et de transfert ;
- o l'évaluation des politiques en faveur de la recherche et de l'innovation.

L'OST a participé à la production de plusieurs documents récents disponibles en ligne :

- Le Bilan de l'évaluation intégrée de la Comue Hésam Université, annexe « Le profil scientifique de Hésam Université » <sup>(6)</sup> ;
- Caractérisation de la production scientifique- Ministère de la Culture – OST, mai 2018 <sup>(7)</sup> ;
- Indicateurs de production des établissements de recherche universitaire (IPERU) – Groupe 1<sup>8</sup> ;
- La position scientifique de la France dans le monde, 2000-2015 <sup>(9)</sup> ;-

<sup>5</sup> Organigramme en ligne à l'adresse suivante : <http://www.hceres.fr/PRESENTATION/Organisation/Organisation-generale>

<sup>6</sup> <https://www.hceres.fr/fr/BilanIntegre-HESAM>

<sup>7</sup> [http://www.culture.gouv.fr/content/download/212952/2237222/version/1/file/HCERES-OST-Caract%C3%A9risation+de+la+recherche+culturelle\\_Rapport+FINAL.pdf](http://www.culture.gouv.fr/content/download/212952/2237222/version/1/file/HCERES-OST-Caract%C3%A9risation+de+la+recherche+culturelle_Rapport+FINAL.pdf)

<sup>8</sup> [https://www.hceres.fr/sites/default/files/media/downloads/2018\\_IPERU\\_positionnement\\_G1.pdf](https://www.hceres.fr/sites/default/files/media/downloads/2018_IPERU_positionnement_G1.pdf)

<sup>9</sup> <https://www.hceres.fr/fr/actualites/la-position-scientifique-de-la-france-dans-le-monde-2000-2015>

- Dynamics of scientific production in the world, in Europe and in France, 2000-2016 (à venir prochainement).

L'amélioration de la qualité des données est une priorité pour l'OST. Elle est couplée au développement de méthodes à l'état de l'art, par exemple pour mener des analyses textuelles du contenu des publications scientifiques ou adapter ses indicateurs.

Afin de garantir la qualité scientifique des travaux de l'OST, le décret n°2014-1365 du 14 novembre 2014, relatif à l'organisation du Hcéres, dote l'Observatoire d'un Conseil d'orientation scientifique (COS) dont les missions sont :

- o Donner son avis sur le programme prévisionnel de travail de l'OST, notamment au regard des moyens à engager ;
- o Évaluer la qualité scientifique et la pertinence des projets de recherche et de développement envisagés ;
- o Examiner en cours d'année l'avancement des travaux et les résultats obtenus ;
- o Proposer au président du Hcéres, le directeur ou la directrice de l'OST, nommé(e) pour un mandat de 4 ans, renouvelable 1 fois.

#### (b) Le département du système d'information (DSI)

Le département du système d'information du Hcéres est transversal à l'ensemble des activités du Haut Conseil. Il assure le développement et le maintien en conditions opérationnelles des applications métiers collaboratives et des outils de communication dont il dispose. Il est composé de 13 personnes et dirigé par un directeur.

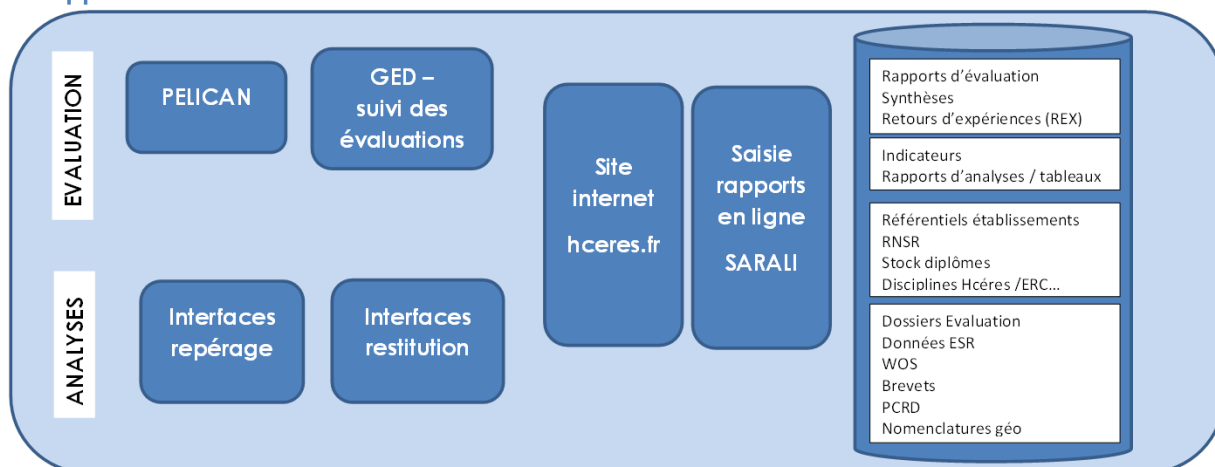
Le département a notamment pour mission :

- d'élaborer une stratégie d'évolution du système d'information et un schéma directeur en adéquation avec le plan stratégique 2016-2020 du Hcéres ;
- d'organiser les ressources du système d'information ;
- de définir une politique des données dans une optique de rationalisation et d'optimisation de leur collecte et de leur mise à disposition ;
- de concevoir et de faire évoluer, en lien avec les parties prenantes de l'évaluation, les supports applicatifs du Hcéres ;
- de faciliter la conception, la mise à disposition et l'exploitation des productions du Hcéres ;
- d'accompagner la participation du Hcéres aux évolutions du système d'information global de l'enseignement supérieur et de la recherche, en France et en Europe.

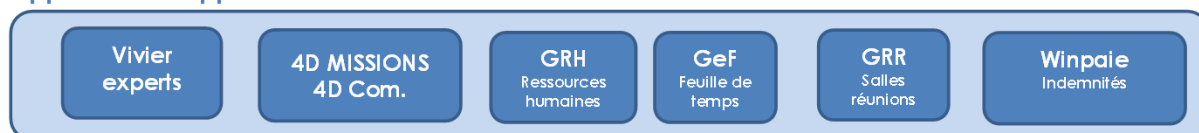
### **2.03 Le système d'information du Hcéres**

Le département maintient le système d'information du Hcéres, composé d'applications de gestion d'une part et d'un système d'information décisionnel d'autre part.

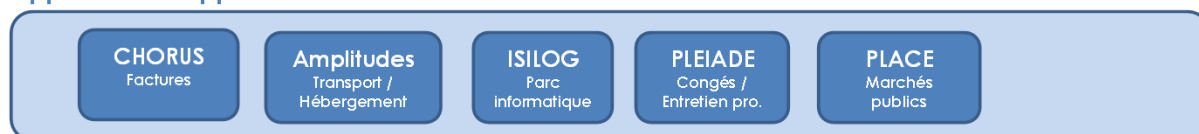
## Applications métiers



## Applications support Hcéres



## Applications support externalisées



Le système d'information de gestion est organisé autour d'un ensemble d'applications en support de l'ensemble des activités du Hcéres, telles que le suivi des évaluations du Hcéres (environ 1 500 par an) et de l'organisation des comités d'experts réalisant les évaluations, la gestion administrative des experts (3 500 par an), la gestion des missions et déplacements des experts et personnels du Hcéres (8 000 par an), la gestion des indemnités des experts mobilisés par le Hcéres, la gestion des réservations de salles, la gestion des ressources humaines, la gestion financière et comptable, etc. Ces applications sont pour partie hébergées sur des systèmes locaux et pour partie sur des infrastructures externalisées auprès d'un hébergeur retenu dans le cadre d'un marché public.

Le système d'information décisionnel (SID) est essentiellement dédié à l'activité des études et des analyses menées par le département Observatoire des Sciences et Techniques (OST). Le SID contient des données bibliographiques (publications scientifiques et brevets, projets européens), ainsi que des référentiels (à compléter et à améliorer) nécessaires à l'analyse des données (disciplinaires, géographiques, etc..) et des indicateurs calculés à partir de ces données.

L'architecture du système d'information décisionnel s'organise autour de deux processus principaux :

- le processus d'intégration de données dans l'infocentre qui aboutit à la construction des informations mobilisées pour les activités d'analyses et d'indicateurs ;
- le processus de restitution des informations.

Ces processus sont organisés selon un ensemble de principes et règles :

- implication de l'expertise fonctionnelle tout au long de la chaîne (intelligence humaine) ;
- automatisation des processus récurrents dès lors qu'il est possible d'énoncer des règles de traitement ;
- maîtrise des mises à jour des données (mises à jour annuelles des données), des consolidations et des mises à disposition des données ;
- rigueur de la qualité des informations contrôlée par une expertise fonctionnelle (expertise bibliométrique, expertise géographique, ...) ;
- accessibilité de l'information à toutes les étapes de la chaîne d'intégration de données ;

- modularité des traitements dans la diversité des données et des outils de restitution.

L'infocentre décisionnel est organisé dans un SGBD (Oracle Enterprise) structuré dans une architecture classique de système d'information décisionnel organisé en DSA, ODS, MDM, DWH et DATAMARTs.

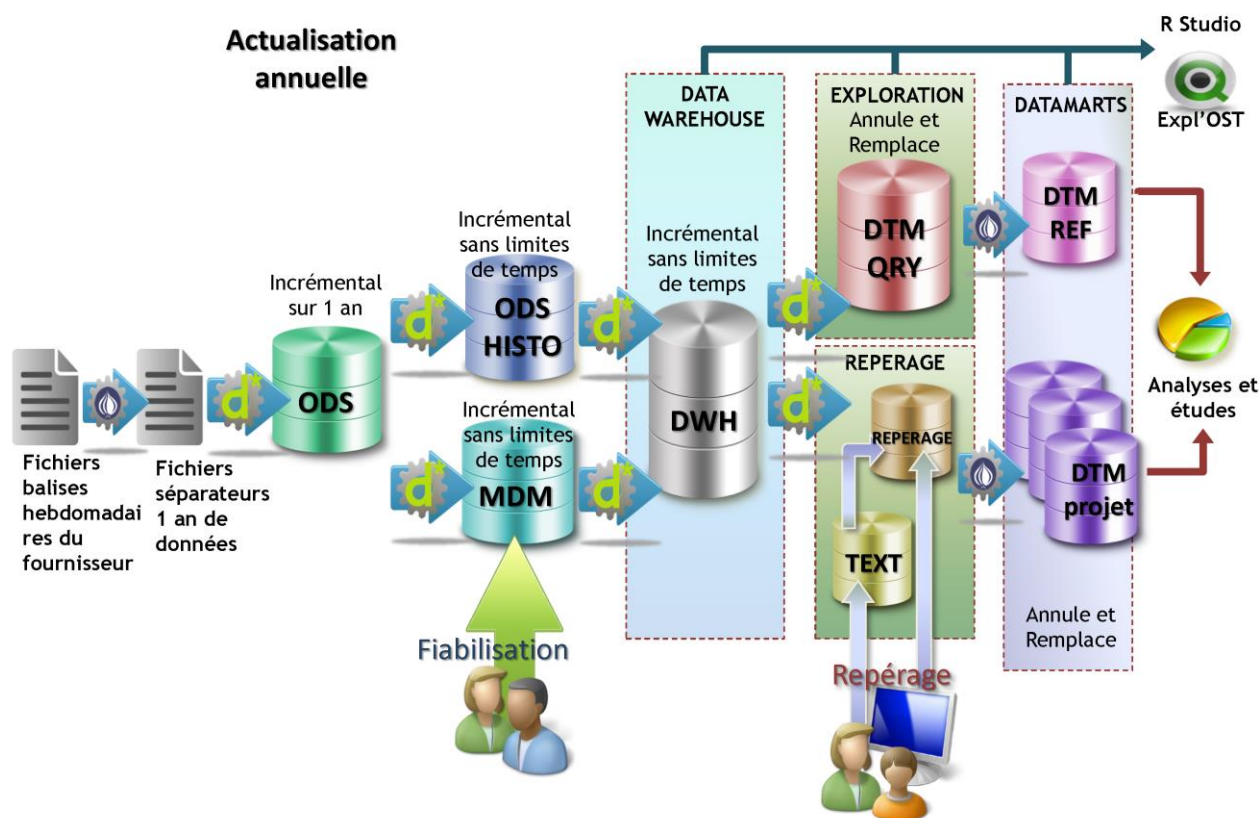
Un ETL (Talend version TIS v 6) permet d'organiser et automatiser le chargement des données. Des applications de restitution pour explorer les données ou restituer des indicateurs soit consolidés à la volée, soit chargés à partir de DATAMARTs, ont été développées sous Qlikview (v 11).

Des outils de traitement de données sont également mobilisés pour explorer les données, réaliser des traitements statistiques, élaborer des analyses basées sur les données de l'infocentre (SQL Developer, STATA, R Studio, etc.).

L'ensemble est mis en œuvre sur des infrastructures externalisées auprès du même hébergeur que le système d'information de gestion.

Le schéma ci-dessous décrit l'organisation du SID concernant les publications scientifiques.

## Architecture du SID - publications scientifiques



Les données du fournisseur Clarivate Analytics sont récupérées chaque semaine sur l'espace FTP mis à disposition par le fournisseur et copiées dans l'infocentre du Hcéres. Une fois par an, l'actualisation des données est réalisée dans le SID. Elle consiste en un ensemble d'opérations qui alimente les différents schémas du SID.

Tout d'abord, les fichiers hebdomadaires sont interprétés par l'intermédiaire d'un script qui transforme les fichiers sources en autant de fichiers qu'il y a de tables dans le schéma ODS. Chaque table contient l'ensemble des données issues des fichiers hebdomadaires, soit 1 an de données, organisées par dimensions (auteurs, adresses, articles, journaux, etc.). Avant chaque nouvelle actualisation, les données de l'ODS sont sauvegardées et l'ODS est vidé pour accueillir les nouvelles données.

Une fois les contrôles techniques réalisés pour vérifier la bonne intégration des données dans l'ODS, les données sont chargées dans le schéma ODS\_HISTO de façon incrémentale. Ce schéma n'est jamais supprimé ou remplacé, les données sont « empilées » afin de garder l'historique des données reçues. Les données intégrées dans l'ODS\_HISTO ne subissent aucune transformation, si ce n'est une réorganisation des



données en tables et l'ajout d'informations permettant la traçabilité des données (dates d'intégration, fichier source, etc.).

Les données nécessitant une mise en qualité, une normalisation et des enrichissements sont stockées dans le schéma MDM (Master Data Management). Il s'agit essentiellement des adresses d'affiliation des auteurs. La normalisation et les enrichissements sont réalisés en s'appuyant sur les référentiels et dictionnaires de données élaborés par le Hcéres, avec des procédures en partie « automatiques ». Le schéma MDM n'est jamais supprimé ou remplacé, les données issues de chaque actualisation y sont ajoutées.

Une fois, la fiabilisation des données terminée, les données sont intégrées dans le DWH par lots. Les lots sont organisés par dimensions métiers (journaux, documents, adresses, citations, etc.). Le DWH est organisé dans un modèle en flocon, dans des tables très normalisées. Le DWH n'est jamais supprimé ou remplacé, les données y sont ajoutées.

Le schéma DTM\_QUERY regroupe les informations du DWH mais organisées en tables métiers, plus proches des besoins exploratoires pour les personnels réalisant des analyses ou pour les besoins des calculs des indicateurs (scripts SQL). Le schéma DTM\_QUERY est remplacé à chaque actualisation.

Certaines données sont indexées dans le moteur Oracle TEXT, dans le schéma « TEXT » pour permettre les recherches par mots clés de manière performante dans des volumes de données importants, dont les adresses d'affiliation des auteurs. Le schéma TEXT est remplacé à chaque actualisation.

Des enrichissements spécifiques sont réalisés sur la base des adresses d'affiliation, appelés « REPERAGES ». Il s'agit d'associer une institution (laboratoire de recherche, organisme de recherche, université, etc.) à une adresse d'affiliation. Les repérages peuvent être réalisés par des acteurs externes aux Hcéres (les institutions elles-mêmes), au travers d'une interface web qui leur est mise à disposition lors de chaque campagne de repérage.

Le schéma REPERAGE contient les enrichissements institutionnels (lien publication-adresse-institution) et également des enrichissements thématiques (lien publication-thème disciplinaire). Le schéma REPERAGE n'est jamais supprimé ou remplacé. En effet de nouveaux enrichissements pouvant être ajoutés au cours des projets, bien que l'actualisation du DWH soit terminée. Ces ajouts peuvent intervenir plusieurs fois entre deux actualisations et à des moments différents d'une année sur l'autre.

Une fois le schéma DTM\_QUERY alimenté et validé, le calcul des indicateurs « de référence » a lieu. Il s'agit d'un ensemble d'indicateurs mobilisant les dimensions géographiques, disciplinaires et annuelles.

Chaque année un nouveau DTM\_REF est calculé sur l'ensemble des données contenues dans le DWH. Les indicateurs « projets » sont calculés au cours des différents projets menés par l'OST et s'appuient sur une partie des indicateurs de référence et sur les enrichissements institutionnels ou thématiques.

A chaque étape du processus d'actualisation, des validations techniques puis fonctionnelles sont réalisées par les différents acteurs mobilisés autour de ce projet.

Le processus complet d'actualisation (chargement, fiabilisation, enrichissements, repérages, indicateurs) s'étale sur une période d'environ 6 mois.

## ARTICLE 3 - PRESENTATION DU PROJET

---

### 3.01 Contexte

Afin de réaliser des études et analyses portant sur les publications scientifiques, le Hcéres exploite les données bibliographiques de la collection Web of Science, fournies par la société Clarivate Analytics, afin de constituer une base bibliographique et bibliométrique.

Si les fichiers de mises à jour sont reçus du fournisseur de façon hebdomadaire, actuellement l'intégration de ces mises à jour dans la base de données du Hcéres est effectuée annuellement. Ce processus d'actualisation, qui comprend le chargement des données, la mise en qualité et les enrichissements, la recette technico-fonctionnelle, jusqu'au calcul d'indicateurs de référence, est très coûteux en temps pour les équipes fonctionnelles et techniques.

Les données bibliographiques étaient jusqu'à présent fournies par Clarivate Analytics au format « tagged data », qui consiste en des fichiers textes au format ASCII, sans délimiteur, organisés à l'aide de balises.



L'évolution vers le format XML des fichiers livrés par le fournisseur a entraîné une réflexion sur l'ensemble du système. Une analyse d'impact a été confiée à une assistance à maîtrise d'ouvrage afin d'envisager différents scénarios possibles pour l'exploitation de ce nouveau format.

À l'issue de cette étude, il est apparu que ce changement de format constitue l'opportunité d'une refonte plus complète du système, afin notamment :

- de pouvoir prendre en compte les nouvelles informations fournies par les fichiers sources qui ne s'intègrent pas dans le modèle existant ;
- de simplifier et améliorer les performances d'un système qui n'avait pu être finalisé lors de sa mise en place ;
- de se doter d'outils plus performants, permettant d'une part de réduire la charge de travail nécessaire au processus d'actualisation, tant du côté fonctionnel (OST) que technique (DSI), et d'autre part de gagner en qualité sur les résultats des enrichissements :
  - par la constitution de référentiels facilitant l'automatisation de la mise en qualité et des enrichissements, améliorant la qualité des résultats et réduisant le besoin d'intervention manuelle ;
  - par la mise à disposition des fonctionnels de l'OST d'interfaces utilisateur plus ergonomiques et plus efficaces facilitant le suivi du déroulement du processus, les opérations manuelles parfois nécessaires à la mise en qualité et aux enrichissements, mais également l'exploration des données ;
- de mettre à profit ce gain en performance pour augmenter la fréquence des actualisations dans l'année.

Devant l'ampleur de ce projet, il a été décidé de le découper en deux phases :

- une première phase (dénommée ci-après Phase 1) concernant le chargement des données, la constitution des référentiels ainsi que la mise en qualité et les enrichissements et la modélisation de l'ensemble des bases de données et des processus ;
- une deuxième phase (dénommée ci-après Phase 2) concernant l'exploitation des données déjà mises en qualité et enrichies pour l'exploration, le calcul d'indicateurs et la production de rapports.

**La présente consultation concerne uniquement la première phase (Phase 1) du projet.**

### **3.02 Démarche projet**

Le marché est constitué d'un ensemble de missions qui correspondent aux modules du projet phase 1, à l'exception des modules pris en charge par les équipes du Hcéres :

- Mission 1 : Conception du référentiel géographique ;
- Mission 2 : Conception du référentiel institutionnel ;
- Mission 3 : Conception du référentiel des supports de publication ;
- Mission 4 : Conception du modèle relationnel métier et des processus associés ;
- Mission 5 : Socle technique de développement, intégration et validation ;
- Mission 6 : Mise en œuvre du modèle métier ;
- Mission 7 : Mise en œuvre du référentiel géographique ;
- Mission 8 : Mise en œuvre du référentiel institutionnel ;
- Mission 9 : Mise en œuvre du référentiel des supports de publication ;
- Mission 10 : Mise en œuvre des interfaces utilisateur.

### **3.03 Pilotage et suivi**

Pour le Hcéres, le projet est piloté par deux chefs de projets, assurant chacun la responsabilité d'un domaine, fonctionnel ou technique.

Ils assurent la coordination des travaux et activités nécessaires à la réussite du projet et sont les interlocuteurs privilégiés pour le pilotage opérationnel du projet. Ils proposent les orientations qui seront présentées au comité de pilotage. Ils suivent l'avancement du projet pour ce qui concerne le planning et le budget notamment.

La gouvernance du projet s'articulera autour de trois instances : un comité de projet, un comité utilisateur et un comité de pilotage :

- **le comité de pilotage** est en charge d'orienter et de valider les différentes étapes et travaux en cours d'étude. Il est constitué du directeur de projet et du chef de projet pour le titulaire, des chefs de projets et des directeurs DSI et OST pour le Hcéres.
- **le comité de projet** a pour objectif de préciser les besoins fonctionnels, valider les solutions qui seront proposées et mises en œuvre au cours des ateliers concernant chacune des missions. Ce groupe est constitué des chefs de projets et des membres pertinents des équipes DSI et OST. Le comité de projet se réunira systématiquement à l'issue de chacune des étapes du projet et sollicitera si nécessaire la validation du comité de pilotage.
- **le groupe utilisateurs** a pour objectif de participer à la rédaction des spécifications et de réaliser les recettes fonctionnelles. Ce groupe est constitué d'acteurs participant activement aux traitements du SID comme les chargés d'études, les statisticiens et les informaticiens.

Les réunions des comités de projet, de pilotage et du groupe utilisateurs se dérouleront dans les locaux du Hcéres.

Le Hcéres attend du titulaire la mise en place d'une organisation de premier niveau :

- un directeur de projet, responsable contractuel permanent du projet, tant sur les plans fonctionnels, techniques et financiers que sur le plan humain ;
- un chef de projet, interlocuteur privilégié des chefs de projet fonctionnel et technique. Il est responsable du pilotage opérationnel du ou des intervenant(s) ;
- une équipe chargée de réaliser les travaux de chacune des missions.

Le titulaire aura un rôle de coordination et d'animation des ateliers et de production des livrables décrits dans chacune des missions. Il sera par ailleurs intégré au sein de l'équipe projet pour réaliser l'ensemble des missions qui lui seront confiées dans le cadre du projet. Des missions complémentaires seront confiées aux équipes internes du Hcéres.

Par ailleurs, pour permettre au Hcéres de juger au mieux de l'équipe proposée, les candidats doivent fournir :

- la composition exacte de l'équipe mobilisée et le CV de chaque intervenant, développeurs compris (seules les personnes mobilisées sur le projet sont pertinentes) ;
- leur organisation en interne pour la réalisation du projet (recours à un prestataire tiers, département SI, etc.) ;
- la présentation des outils de pilotage et de suivi qui seront mis en œuvre pour la réalisation du projet tels que des outils de suivi de projet, de gestion de tâches, gestion de bugs, espace collaboratif.

### 3.04 Description du périmètre du projet

La phase 1 concernée par le marché peut être découpée en plusieurs modules :

- chargement des données brutes : cette étape, prise en charge par les équipes du Hcéres, est d'ores et déjà opérationnelle pour l'initialisation du système, elle ne fait donc pas partie du présent marché ; des ajustements sur cette étape sont possibles, en particulier si cela s'avère nécessaire compte tenu des processus définis dans le cadre du présent marché ;
- élaboration de référentiels : modélisation, cycle de vie, processus de mise en qualité et de rapprochement avec les données bibliographiques, le tout incluant l'historisation ; les référentiels géographique, institutionnel et supports de publication font partie du présent marché ; l'élaboration d'un quatrième référentiel concernant la nomenclature disciplinaire, sera à la charge des équipes du Hcéres ;
- réalisation d'interfaces utilisateur pour la gestion des référentiels et des processus de mise en qualité et d'enrichissement ;
- conception du modèle relationnel métier ;
- mise en œuvre de l'alimentation de ce modèle à partir des données chargées.

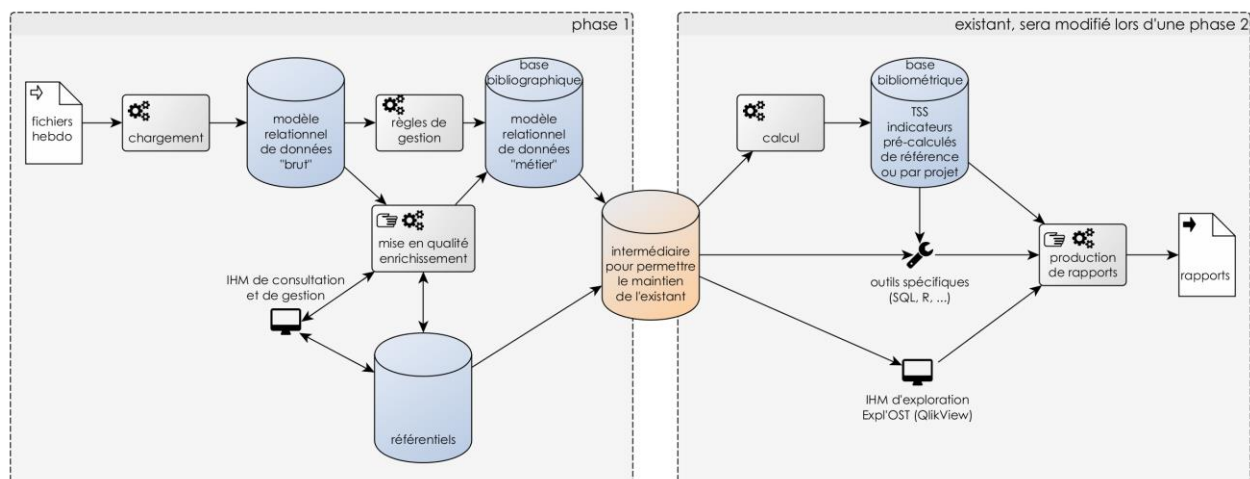


Figure 1: Vue globale du système cible à la fin de la phase 1

Le schéma ci-dessus représente l'organisation globale cible pour la fin de la phase 1.

On peut y voir que les éléments du système d'information concernant l'exploitation des données (indicateurs, exploration, production de rapports...) n'évolueront que dans une phase ultérieure (phase 2) du projet. Afin de pouvoir continuer les activités de production d'indicateurs et de rapports, cœur de métier de l'OST, il faut donc ajouter la réalisation d'une base de données intermédiaire permettant l'interopérabilité entre la nouvelle base bibliographique développée durant la phase 1 et ces éléments existants. Cette partie sera réalisée par les équipes du Hcéres et ne fait pas partie du présent marché.

Pour l'ensemble des missions, le titulaire considérera prioritairement le cycle de vie du système : c'est-à-dire le traitement des données de mises à jour régulièrement transmises par le fournisseur (Clarivate Analytics). La conception des modèles en tiendra compte ainsi que la description des processus.

En ce qui concerne les données bibliographiques, les livraisons provenant du fournisseur sont hebdomadaires, sous forme de fichiers XML. Ces fichiers comportent des ajouts de données, suppressions de données et modifications à appliquer sur les données déjà intégrées. Ces fichiers seront traités soit en masse, à l'occasion des actualisations régulières, soit au fil de l'eau.

Les processus faisant partie du cycle de vie des référentiels seront abordés plus loin dans le document.

Le système développé au cours du présent marché étant destiné à remplacer un système existant, il y aura une étape de reprise de données afin de l'initialiser. Cette étape ne fait pas partie du présent marché et sera prise en charge par les équipes du Hcéres.

Pour information, l'initialisation des données bibliographiques brutes sera réalisée à partir d'un « stock » XML fournis par Clarivate Analytics rassemblant toutes les données historiques depuis 1999. Il est envisagé d'intégrer ces données d'une manière similaire aux ajouts des fichiers hebdomadaires. En ce qui concerne les enrichissements et rapprochements déjà existants, il reste à définir comment les conserver et à quelle étape les introduire dans le nouveau modèle, ce qui sera réalisé par le Hcéres.

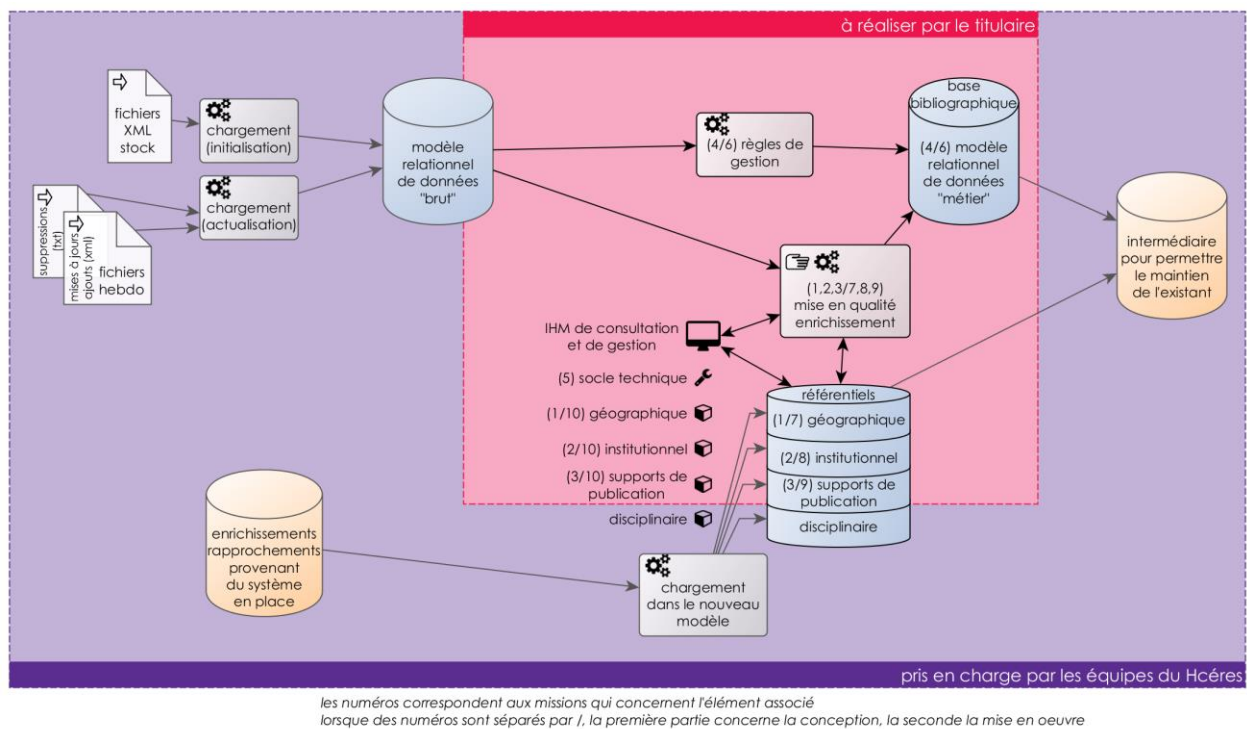


Figure 2: Étapes de la phase 1

La reprise et la maintenance de tout ce qui aura été réalisé dans le cadre du présent marché seront prises en charge par le Hcéres. Le titulaire devra en conséquence s'assurer du transfert de compétences vers les équipes du Hcéres lors de l'exécution des missions.

## ARTICLE 4 - DESCRIPTION DES MISSIONS ATTENDUES

### 4.01 Éléments communs à toutes les missions

#### (a) Évolutivité

Les besoins définis dans le présent document, ainsi que ceux qui seront exprimés au cours des différents ateliers et séances de travail, sont amenés à évoluer au cours de la vie du système.

En effet, des changements peuvent se produire au niveau des données sources fournies par Clarivate Analytics (évolution du modèle des livraisons, ajout de nouvelles informations comme dernièrement open access...).

Les données et la structuration des référentiels sont également susceptibles d'être modifiées.

Enfin, des études futures peuvent faire émerger de nouveaux besoins en matière de processus ou de données à traiter.

La modélisation des données et des processus, la conception des interfaces utilisateur, ainsi que les solutions techniques choisies pour l'implémentation, devront donc être suffisamment souples pour permettre d'intégrer ces évolutions.

#### (b) Maintenabilité

À l'issue du présent marché, les équipes du Hcéres prendront en charge la maintenance du système, et seront amenées à réaliser les éventuelles évolutions mentionnées au paragraphe précédent.

Durant l'exécution du marché, elles assumeront également la conception et l'implémentation de certaines étapes du projet (référentiel disciplinaire, reprise des données existantes).

Le titulaire devra donc s'assurer du transfert de compétence tout au long de l'exécution des missions. De plus, les modèles comme le code produits par le titulaire devront être accompagnés de documentation

technique détaillée, de toute procédure nécessaire à la reconstruction des livrables binaires et à leur déploiement. Les choix techniques pour l'implémentation, ainsi que le style du code produit devront permettre une maintenance et une reprise aisées pour les équipes du Hcéres. Une vigilance particulière sera apportée à la qualité de la documentation.

#### **4.02 Éléments communs aux missions de conception des référentiels**

La constitution d'un référentiel comporte 4 aspects : la modélisation des concepts, le cycle de vie, les méthodes de rapprochement et les interfaces utilisateur.

Sur chacun de ces aspects, un travail conjoint avec les experts métier permettra au titulaire d'établir les spécifications techniques et fonctionnelles. Le titulaire assure ensuite la mise en œuvre des solutions définies.

La modélisation comprend la description des entités du référentiel. Cette description est double. D'une part il s'agit de définir quels sont les éléments que l'on intègre dans le référentiel. D'autre part il s'agit de définir les caractéristiques qui permettent de décrire les éléments intégrés dans le référentiel. Il est à noter qu'il existe des référentiels externes d'autorité qui fournissent à la fois les éléments à intégrer et contribuent à la description de ces éléments.

Chacun des référentiels doit permettre de créer des nouveaux éléments en combinant des éléments préexistants.

Le cycle de vie définit comment évolue le contenu du référentiel. Il expose les règles de gouvernance des données du référentiel ainsi que l'impact d'une évolution du contenu sur le référentiel lui-même et les données qui référencent les entrées concernées. Certaines entrées des référentiels peuvent avoir un caractère limité dans le temps, sous la forme d'une date de début et d'une date de fin de validité. Le référentiel doit conserver une trace des évolutions de son contenu.

Les méthodes de rapprochement décrivent l'ensemble des procédures qui permettent d'associer une entrée du référentiel à une donnée source que l'on souhaite « référentialiser ». L'historique des rapprochements effectués doit être conservé dans le référentiel.

Les interfaces utilisateur à réaliser pour chacun des référentiels, doivent permettre de réaliser des opérations de gestion des référentiels :

- visualisation, exploration des données du référentiel : fonctionnalités de recherche, de tri, de sélection et d'export ;
- ajout / modification / suppression « unitaire » (nombre limité d'éléments, de l'ordre de ce qui est affichable sur une page) ;
- ajout / modification / suppression « massives » (permettant l'intégration dans une chaîne de traitement intégrée) via import/export de fichiers ;
- consultation des rapprochements effectués ;
- outillage de la phase de rapprochement (application séquentielle des règles, mesure de la qualité) ;
- définition/construction de regroupements d'éléments selon différentes typologies.

Ces interfaces utilisateur doivent tenir compte des différents rôles attribués à chacun des utilisateurs et droits associés.

Les référentiels, les processus de gestion associés et les interfaces utilisateur les concernant devront être conçus de façon à garantir l'évolutivité de ceux-ci, comme mentionné au paragraphe 4.01 (a) ci-dessus.

Suivant les référentiels, le Hcéres dispose d'éléments de réflexion plus ou moins avancés sur les différents aspects constituant les référentiels. Le tableau ci-dessous présente l'appréciation que le Hcéres a sur ces aspects.

	Modélisation		Cycle de vie	Rapprochement	Interfaces utilisateur
	liste des éléments	caractérisation			
Référentiel Géographique	avancé	moyen	avancé	avancé	moyen
Référentiel Institutionnel	moyen	faible	faible	moyen	moyen
Référentiel des supports de publications	avancé	avancé	avancé	moyen	moyen

### 4.03 Mission 1 : Conception du référentiel géographique

#### 4.03.1 Objectifs et attendus de la mission

Cette mission consiste à réaliser la conception fonctionnelle et technique du référentiel géographique et des processus associés, tels que décrits dans la partie 4.03.2 ci-dessous. Sont concernés par cette mission les processus de cycle de vie du référentiel mais également ceux relatifs à l'utilisation du référentiel pour la mise en qualité et l'enrichissement des données bibliographiques du modèle métier.

Le titulaire organisera les ateliers nécessaires à l'élaboration des spécifications fonctionnelles et techniques du référentiel géographique, en collaboration avec l'équipe du Hcéres dédiée.

Il élaborera la cartographie des processus de mise à jour du référentiel, de même que les spécifications fonctionnelles détaillées qui comprendront notamment les règles de gestion des rapprochements pour la mise en qualité des données et les enrichissements, ainsi que la logique de mise à jour du modèle métier.

Il élaborera ensuite les spécifications techniques détaillées en cohérence avec les autres référentiels (missions 2 et 3), le modèle métier (mission 4) et le socle technique défini lors de la mission 5.

L'implémentation du référentiel géographique et des processus associés sera effectuée lors de la mission 7 sur la base des spécifications définies au cours de cette mission. L'implémentation de l'interface utilisateur sera réalisée lors de la mission 10 sur la base des spécifications définies au cours de cette mission.

#### 4.03.2 Description fonctionnelle du référentiel géographique

Le référentiel géographique a pour finalité la structuration des données et productions pérennes de l'OST.

##### (a) Concept décrits

En fonction de la zone géographique considérée, le degré de précision attendu et donc l'échelle de calcul des indicateurs est différent :

- La France : niveau commune ;
- L'Europe (y compris la Suisse, la Norvège, la Turquie, le Liechtenstein et la Macédoine) : niveau région NUTS ;
- Les Etats-Unis : niveau Etat ;
- Le reste du monde : niveau pays.

(b) Sources de données

- Périmètre « MONDE » : Pays - Codes ISO

Le code ISO utilisé par le Hcéres correspond à la classification des pays au sens de la norme ISO 3166-1 ([https://fr.wikipedia.org/wiki/ISO\\_3166-1](https://fr.wikipedia.org/wiki/ISO_3166-1)) en 3 lettres, mise à jour au 16 avril 2017. La table de référence interne (Schéma NOMENC) s'appuie sur 202 codes ISO, correspondant aux pays « officiels » sur 249 définis par la norme ISO (ex : la Guadeloupe dispose d'un code ISO propre - « GUA », mais est ici rattachée à la France - « FRA »)

- Périmètre « UE » élargie : Régions - Codes NUTS

L'office de statistiques européen Eurostat a mis en place une nomenclature - NUTS, qui s'appuie principalement sur les découpages administratifs dans chaque pays. Les pays couverts par cette nomenclature sont les pays de l'union européenne (UE28). Elle est toutefois étendue à des pays ne faisant pas partie de l'union (<https://ec.europa.eu/eurostat/web/nuts/background>).

Eurostat propose les 5 niveaux de découpage suivants :

- NUTS 0 : Pays
- NUTS 1 : « Grandes régions »
- NUTS 2 : Régions
- NUTS 3 : Départements
- LAU : « Local Authority Unit »

Le référentiel NUTS a été développé à des fins de comparaisons statistiques avec la constitution de régions de tailles équivalentes. Il s'appuie ainsi sur des systèmes de remontées de données qui peuvent s'avérer hétérogènes d'un pays à l'autre, tant sur les unités représentées que sur la fiabilité des découpages, qui dépendent des pays. Pour la France, le LAU est équivalent à la commune (environ 36 000) alors qu'en Grande Bretagne, il équivaut aux autorités administratives locales telles qu'elles ont été définies par la réforme « local government act », 1992.

Eurostat propose plusieurs formats de données qui peuvent être directement exploitables :

- Gisco, le SIG de la commission européenne, recense l'ensemble des classifications et nomenclatures développées par Eurostat :

<https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/>

et propose aussi des fichiers au format cartographique (.shp) :

<https://ec.europa.eu/eurostat/fr/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts>

- Ramon est une plate-forme qui permet de disposer d'information sur la constitution des nomenclatures utilisées par Eurostat :

[https://ec.europa.eu/eurostat/ramon/index.cfm?TargetUrl=DSP\\_PUB\\_WELC](https://ec.europa.eu/eurostat/ramon/index.cfm?TargetUrl=DSP_PUB_WELC)

et notamment les codes NUTS

[https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST\\_CLS\\_DLD&StrNom=NUTS\\_2016L&StrLanguageCode=EN&IntCurrentPage=1&StrLayoutCode=LINEAR](https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&StrNom=NUTS_2016L&StrLanguageCode=EN&IntCurrentPage=1&StrLayoutCode=LINEAR)

- Ainsi que d'autres supports qui peuvent être utiles à mobiliser :
  - o Les tables de correspondance « LAU » / NUTS3 :  
<https://ec.europa.eu/eurostat/fr/web/nuts/local-administrative-units>
- Les tables de correspondance entre deux versions des NUTS (ex : 2013/2016)
  - o En version « tableau » : <https://ec.europa.eu/eurostat/fr/web/nuts/history>
  - o En version « carte » : <https://ec.europa.eu/eurostat/web/nuts/nuts-maps-.pdf>
- Les tables de correspondance « Codes postaux » / NUTS3 :  
<http://ec.europa.eu/eurostat/tercet/flatfiles.do>

- Périmètre « France » : Codes INSEE villes

L'INSEE fournit le référentiel des découpages administratifs français (communes), qui correspondent pour grande partie aux LAU Eurostat pour la France.

Ces référentiels sont disponibles à partir des liens ci-dessous



	Nombre de communes	Date de mise à jour	
Métropole et DOM	35.885	22/09/2017	<a href="https://www.insee.fr/fr/information/2114819">https://www.insee.fr/fr/information/2114819</a>
COM	94	01/03/2017	<a href="https://www.insee.fr/fr/information/2028040">https://www.insee.fr/fr/information/2028040</a>

<https://www.insee.fr>, onglet « géographie administrative et d'étude »

Au total, ce référentiel comprend à ce jour 35 979 entités (communes françaises), identifiables par le couple « code commune » + « nom commune », le code commune faisant office d'identifiant unique, à l'inverse du libellé de commune, qui peut présenter des homonymies.

(c) Cycle de vie :

Périmètre « MONDE » : Pays - Codes ISO

La nomenclature des pays est relativement stable. Certains codes ISO peuvent néanmoins être modifiés.

Périmètre « UE » élargie : Régions - Codes NUTS

La nomenclature Eurostat est ainsi revue tous les trois ans pour tenir compte des évolutions administratives de chaque pays. D'une mise à jour à l'autre, les opérations suivantes peuvent être réalisées sur les entités géographiques, de tout niveau :

- Fusion : 2 entités ou plus se regroupent en une seule ;
- Séparation : 1 entité est découpée en 2 entités ou plusieurs ;
- Changement de frontières : des LAU (équivalent de communes) changent d'entité de rattachement.

Les fichiers doivent donc être vérifiés avant intégration des référentiels.

Il est nécessaire de procéder à une vérification/fiabilisation des découpages avant intégration.

Les opérations sont les suivantes :

- Cohérence LAU / NUTS3 / NUTS 2 ;
- Cohérence ancien NUTS / nouveau NUTS.

Périmètre « France » : Codes INSEE ville

Le découpage du territoire français en communes est assez stable dans le temps mais reste soumis à des évolutions, avec notamment, des fusions de communes qui se traduisent par :

- Un regroupement ou un changement du nom des deux communes, voire la disparition du nom de l'une des deux qui est intégrée à la seconde ;
- Un rattachement du territoire des deux communes au code INSEE unique de l'une des deux communes d'origine.

(d) Rapprochement

La classification des publications consiste à attribuer, en fonction du périmètre, un code ISO, un code NUTS et un code INSEE/LAU à chacune des adresses de ses auteurs.

**Tableau 1 : Codes attribués selon le périmètre**

	Code ISO	Code NUTS	Code INSEE
Monde	X	-	-
UE	X	X	-
France	X	X	X

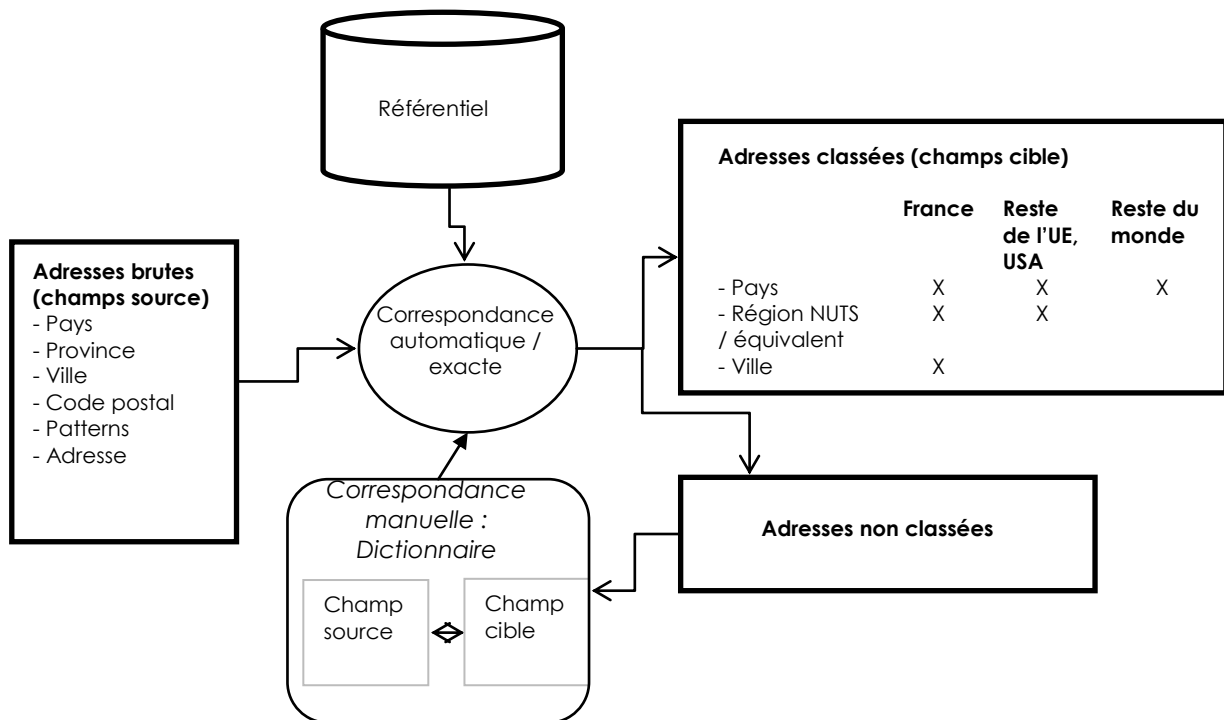
D'une manière générale, cette attribution est déterministe : les informations contenues dans l'adresse issue des données sources permettent sans ambiguïté cette affectation.

Certaines adresses sont toutefois ambiguës et des procédures de traitement de données peuvent être nécessaires. Elles s'appuient, dans la mesure du possible, sur des règles automatiques et/ou des tables de rapprochement (dictionnaires) qui permettent de minimiser les ambiguïtés, mais doivent être complétées par une vérification et une fiabilisation manuelle.

La liste (non exhaustive) des ambiguïtés observées dans les données est la suivante :

- Ambiguïtés sémantiques / formes équivalentes : la donnée source est mal orthographiée, représente une zone particulière (ex : zone d'activité) mais peut être attribuée à une valeur du référentiel sans ambiguïté ;
- Ambiguïtés sémantiques / non attribuables : la donnée source fait potentiellement référence à plusieurs valeurs du référentiel, et une information complémentaire est nécessaire pour classer l'adresse ;
- Ambiguïtés locales : la donnée source permet de situer localement l'adresse, mais fait référence à une zone géographique plus large.

Pour une information cible donnée (pays / région / ville), le processus de rapprochement peut donc être schématisé de la façon suivante :



La procédure de rapprochement est réalisée en deux étapes :

- Une 1<sup>ère</sup> étape d'affectation automatique ;
- Une 2<sup>nde</sup> étape d'enrichissement / correction manuelle pour les adresses non classées ou les classements ambigus (au moins deux modalités pour un champ dans la classification cible) mobilisant les dictionnaires / tables de rapprochement.

#### Dictionnaires / tables de rapprochement

Les dictionnaires sont des outils qui permettent de corriger les adresses ambiguës ou non classées pour leur attribuer un classement unique et ainsi minimiser le taux de non classés. Pour tenir compte de l'hétérogénéité des données sources et aboutir au classement le plus précis possible, plusieurs niveaux de correspondance doivent pouvoir être faits, entre les différents niveaux de granularité dans les données sources et les données cible, tels que décrits ci-dessous.

**Tableau 2 : Liste des correspondances possibles entre données source et données cible**

Source \ Cible	Ville	Région NUTS	Pays
Adresse	X	X	X
« Patterns » / expressions régulières d'adresses	X	X	X
Ville	X	X	X
Code postal		X	X
Pays			X

Exemple :

- le champ « ville » contient « Saint-Etienne » qui est identifiée comme information ambiguë, la correspondance « expression régulière dans l'adresse/ville » doit être mobilisée pour déterminer la ville correcte sans ambiguïté ;
- le champ « ville » contient « Bretagne » et aucune information au sein de l'adresse ne permet de connaître plus précisément la ville, la correspondance « Ville/Région NUTS » doit être mobilisée pour déterminer la région NUTS bien que la ville ne soit pas connue.

(e) Interface utilisateur

En complément des fonctionnalités communes aux interfaces utilisateur associées aux référentiels, celle relative au référentiel géographique doit permettre :

- De décrire explicitement les règles de gestion / priorité : affichage de contenus textuels à destination des utilisateurs ;
- De décrire les processus / logiques d'enrichissement manuels : affichage de contenus textuels à destination des utilisateurs ;
- D'exposer les adresses des publications avec le résultat du rapprochement avec le référentiel.

4.03.3 *Livrables*

- modélisation conceptuelle des données du référentiel géographique ;
- modélisation et cartographie des processus (mise à jour du référentiel et des rapprochements) ;
- règles de gestion applicables lors des processus (mise à jour du référentiel et des rapprochements) ;
- spécifications fonctionnelles détaillées ;
- maquette des interfaces utilisateurs (rendu IHM, ...) ;
- spécifications techniques détaillées ;
- plan de test ;
- documentations, comptes rendus d'ateliers.

**4.04 Mission 2 : Conception du référentiel institutionnel**

4.04.1 *Objectifs et attendus de la mission*

Cette mission consiste à réaliser la conception fonctionnelle et technique du référentiel institutionnel et des processus associés, tels que décrits dans la partie 4.04.2 ci-dessous. Sont concernés par cette mission les processus de cycle de vie du référentiel mais également ceux relatifs à l'utilisation du référentiel pour la mise en qualité et l'enrichissement des données bibliographiques du modèle métier.

Le titulaire organisera les ateliers nécessaires à l'élaboration des spécifications fonctionnelles et techniques du référentiel institutionnel, en collaboration avec l'équipe du Hcéres dédiée.

Il élaborera la cartographie des processus de mise à jour du référentiel, de même que les spécifications fonctionnelles détaillées qui comprendront notamment les règles de gestion des rapprochements pour la mise en qualité des données et les enrichissements, ainsi que la logique de mise à jour du modèle métier.

Il élaborera ensuite les spécifications techniques détaillées en cohérence avec les autres référentiels (missions 1 et 3), le modèle métier (mission 4) et le socle technique défini lors de la mission 5.

L'implémentation du référentiel institutionnel et des processus associés sera effectuée lors de la mission 8 sur la base des spécifications définies au cours de cette mission. L'implémentation de l'interface utilisateur sera réalisée lors de la mission 10 sur la base des spécifications définies au cours de cette mission.

#### 4.04.2 Description fonctionnelle du référentiel institutionnel

Le référentiel institutionnel comporte principalement deux catégories d'entités : l'opérateur de recherche et la structure de recherche. Pour les besoins fonctionnels des études réalisées, il doit être possible de combiner les entités des deux natures pour constituer des périmètres ad hoc.

##### (a) Concepts décrits

L'opérateur de recherche s'entend comme un acteur de l'enseignement supérieur et de la recherche comme une université, une école (ingénieur, commerce, architecture, art, ...), un organisme de recherche, un CHU, etc. ;

La structure de recherche correspond à une unité de recherche, à un département de recherche (éventuellement à un collectif d'auteur).

Une troisième nature d'acteur, nommée combinaison, s'appuie sur des entités des deux natures décrites dans cette section. Il peut s'agir de la combinaison de plusieurs opérateurs, de plusieurs structures ou d'un mélange d'entités de ces deux natures.

Une structure de recherche peut être liée à un ou plusieurs opérateurs de recherche, que l'on appelle dans ce cas la (les) tutelle(s). Une combinaison de structures de recherche peut permettre de reconstituer un opérateur de recherche. Il y a donc une certaine hiérarchie dans les entités référencées, sans que la présence de cette hiérarchie soit obligatoire.

Des acteurs qui ne sont ni des opérateurs de recherche, ni des structures de recherche (ou des combinaisons) doivent également être pris en compte afin de pouvoir identifier l'implication d'une entreprise (ou d'un acteur en dehors de la recherche) dans une publication (voir section rapprochement). Il n'est pas envisagé d'être exhaustif pour les acteurs de ce type.

La dimension institutionnelle est une dimension sur laquelle des travaux sont conduits au sein du Hcéres à partir de données dépassant les seules publications. La caractérisation des acteurs de ce référentiel devra assurer une bonne compatibilité du référentiel construit pour la base de publications avec les autres bases (brevets et projets européens) exploitées par le Hcéres. La présente mission pourra bénéficier des travaux réalisés dans le cadre de la typologie des déposants de brevets.

##### (b) Sources de données

Opérateurs de recherche :

Il n'existe pas de source unique faisant autorité en matière d'opérateurs de recherche à l'échelle mondiale bien que différentes initiatives peuvent servir à alimenter le référentiel. La liste des entrées du référentiel s'appuie sur différentes sources qu'il convient de combiner.

Niveau France :

Dans le cadre de ses activités d'évaluation, le Hcéres maintient en interne une liste des opérateurs de recherche concernés. Pour l'essentiel il s'agit d'opérateurs de recherche relevant de l'ESR français. Certains de ces opérateurs font l'objet d'une attention particulière afin de produire des rapports d'indicateurs, en particulier ceux relevant des programmes budgétaires 150, 159, 172, 190 et 193 de la LOLF.

Une source de données officielle existe pour le niveau opérateur, « l'Application de Consultation et Cartographie des Etablissements du système éducatif français », ou ACCÉ. Cette application est consultable à l'adresse ci-dessous :

<https://www.education.gouv.fr/acce/search.php>

Le champ des établissements recensés via cette application présente néanmoins des écarts avec les besoins du Hcéres : tous les établissements du primaire et du secondaire y figurent tandis que, par exemple, les centres hospitaliers universitaires (CHU) ou les centres de lutte contre le cancer (CLCC) sont inégalement présents.

Des combinaisons d'opérateurs sont déjà constituées pour les besoins de l'accompagnement de l'évaluation des coordinations territoriales. L'appartenance d'un opérateur à une combinaison peut être caractérisée pour tenir compte de la diversité des liens d'associations entre opérateurs.

Environ 500 opérateurs de recherche français sont concernés par l'évaluation et/ou la production de rapports d'indicateurs.

Niveau Européen :

A des fins de comparaisons dans le cadre de projets, un groupe de 23 établissements européens fait l'objet d'un travail spécifique, au sein du Hcéres, en matière de calcul d'indicateurs. Il s'agit des opérateurs de la « League of European Research Universities » (ou LERU). La liste de ces opérateurs est accessible à l'adresse suivante :

<https://www.leru.org/members>

Plus largement, le projet « European Tertiary Education Registry » (ou ETER) constitue un référentiel des établissements d'enseignement supérieur concernant près de 2,800 institutions de 37 pays (UE28 + Albanie, Islande, Liechtenstein, Monténégro, Norvège, Serbie, Suisse, Turquie et Macédoine). Le site officiel du projet fournit toute la documentation utile (gestion de la « démographie » des institutions référencées, description des métadonnées associées, mode d'emploi de l'API, ...). Ce site est accessible à l'adresse suivante :

<https://eter-project.com/#/home>

Le Hcéres est engagé dans un projet Européen qui s'appuie sur ETER pour la partie référentiel institutionnel et collabore régulièrement avec les porteurs du projet.

Niveau monde :

A des fins de comparaisons dans le cadre de projets, un groupe de 62 établissements nord-américains fait l'objet d'un travail spécifique, au sein du Hcéres, en matière de calcul d'indicateurs. Il s'agit des établissements membres de « l'Association of American Universities » (ou AAU). La liste de ces opérateurs est accessible à l'adresse suivante :

<https://www.aau.edu/who-we-are/our-members>

*Clarivate Analytics*, éditeur des données sources de la base de publications de l'OST, maintient une liste de noms d'organisations normalisés. Cette normalisation permet de rassembler les différentes graphies d'une organisation observées dans les adresses sous un intitulé unique. Les données au format XML fournissent les associations identifiées par *Clarivate Analytics* entre une adresse et une organisation présente dans cette liste. Cette information est habituellement désignée sous le nom « organization enhanced ».

La liste des organisations ainsi que la documentation associée peuvent être consultés à l'adresse ci-dessous :

<http://help.prod-incites.com/inCites2Live/dataAndSubscriptionNotificationsGroup/dataAndSubsNotice.html>

La liste des « organization enhanced » n'est toutefois pas exhaustive.

Structures de recherche :

Le niveau structures de recherche concerne « exclusivement » la France, dans le sens où ces structures sont liées à au moins un opérateur de recherche français. Dans le cadre des politiques de coopérations scientifiques à l'international, certaines structures sont géographiquement implantées à l'étranger et peuvent impliquer des opérateurs de recherche étrangers. Toutefois, leur nombre est limité par rapport à l'ensemble des structures de recherche présentes dans le « Répertoire National des Structures de Recherche », ou RNSR, qui doit servir de base au référentiel. Ce répertoire est maintenu par le ministère en

charge de l'enseignement supérieur et de la recherche. Ce répertoire peut être consulté à l'adresse ci-dessous :

<https://appliweb.dgri.education.fr/rnsr/ChoixCriteres.jsp?PUBLIC=OK>

L'ordre de grandeur pour les structures de recherche se situe autour de 3,000 entités. Jusqu'ici, les structures de recherche n'apparaissent pas en tant que telles dans les rapports d'indicateurs. Elles sont utilisées par combinaison pour attribuer des publications à des opérateurs de recherche. Elles peuvent également permettre d'identifier l'implication des opérateurs de recherche auxquelles elles sont liées dans les publications.

Afin de pouvoir, le cas échéant, mobiliser plus largement le niveau structure de recherche pour des travaux de production d'indicateurs, il est nécessaire de les intégrer dans le référentiel.

La relation hiérarchique existante entre les structures de recherche et les opérateurs de recherche doit être caractérisée pour tenir compte des différentes situations observées. Des relations d'appartenance peuvent exister entre structures de recherche.

Tableau récapitulatif

	opérateurs		structures	
	source de données	ordre de grandeur	source de données	ordre de grandeur
France - rapports d'indicateurs	liste interne	500	(140 utilisées dans des combinaisons)	
France - évaluations Hcéres	Référentiel entités évaluées Hcéres, (ACCÉ)	500, (non pertinent)	RNSR, (Référentiel entités évaluées)	3000
hors France - indicateurs / comparaison : LERU, AAU	liste interne	23+62		
Hors France - UE	ETER	2800, 37 pays		
Hors France - Reste du monde	Clarivate Analytics - Organization Enhanced	12000 (Academic, Academic system, Research Org, Health)		

Les ensembles présentés dans le tableau ci-dessus ne sont pas nécessairement disjoints. Ils sont distingués car le niveau de détail nécessaire à leur description et/ou les besoins en termes de rapprochement diffèrent.

#### (c) Cycle de vie

Opérateurs de recherche :

Les opérateurs de recherche peuvent disparaître, fusionner, se scinder ou être créés. La survenue de ces événements doit être enregistrée dans le référentiel à des fins d'historisation.

Dans le cas des opérateurs de recherche français publics, ces événements font l'objet d'une notification légale par le biais de la publication d'un décret.

De manière comparable, les évolutions de certaines combinaisons d'opérateurs font également l'objet d'une notification légale par le biais de la publication d'un décret.

D'autres combinaisons, constitués pour les besoins de projets spécifiques ne sont pas signalées par une publication légale.

Structures de recherche :

Les structures de recherche connaissent également des évolutions. Celles-ci sont bien plus fréquentes que les évolutions relatives aux opérateurs de recherche. On peut estimer qu'environ 5% des structures connaissent une évolution chaque année. En effet, 20% des structures font l'objet d'une évaluation chaque année (soit entre 500 et 600), et parmi celles-ci, 20% ne sont pas renouvelées à l'identique. Les différents types d'évolution que peuvent connaître les structures de recherche sont pris en compte dans le RNSR.

(d) Rapprochement

L'association entre une publication et un acteur (opérateur ou structure de recherche) se fait à partir des adresses des publications. Au sein du Hcéres, cette association est habituellement appelée « repérage ». **Une même adresse peut être associée à plusieurs acteurs.**

Une partie des opérateurs français, ceux pour lesquels des indicateurs sont produits, sont sollicités chaque année afin d'identifier les adresses des publications auxquelles ils sont associés. Le cas échéant, un ensemble d'adresses situées dans le périmètre géographique d'implantation de l'opérateur lui est présenté. Pour d'autres opérateurs, ces associations sont établies en recherchant la présence d'éléments distinctifs de cet acteur, ou « patterns », dans les adresses. Au fil du temps, le Hcéres a constitué un dictionnaire de « patterns », qui comporte 982 entrées.

En cumulant ces deux méthodes un peu plus de 20 millions de liens d'association entre une adresse et un opérateur ont été établis. Cela correspond à 2,3 millions de publications distinctes associées à 93,000 adresses distinctes.

D'autres méthodes de rapprochement sont à envisager, en particulier, le champ « organization enhanced » présent dans les données XML associe une adresse à une entrée de la liste des « organization enhanced ». Cette information, dont la qualité reste à apprécier, peut permettre de construire des indicateurs sur des institutions pour lesquelles le Hcéres ne dispose d'aucun élément à l'heure actuelle (pas d'évaluation ni d'indicateurs produits).

En conservant l'approche interactive avec les acteurs, un rapprochement massif peut être envisagé sur la base des publications identifiées par l'opérateur par le biais de son archive institutionnelle.

Cas des entreprises :

Un dictionnaire de « patterns » spécifique à certains types d'institutions, dont les entreprises, a été construit au fil du temps. La présence d'une de ces « patterns » dans une adresse est conservée en renseignant une caractéristique (flag 0/1) au niveau de cette adresse. En effet, le lourd travail d'identification précise d'une entreprise donnée, et l'harmonisation des dénominations trouvées dans les adresses n'est pas fait.

(e) Interface utilisateur

En complément des fonctionnalités communes aux interfaces utilisateur associées aux référentiels, celle relative au référentiel institutionnel doit permettre :

- de consulter les adresses des publications qui sont associées à une entrée du référentiel (structure, opérateur ou combinaison) ;
- de gérer les « patterns » intervenant dans la procédure de rapprochement : ajout / modification / suppression unitaires (via une page) ou massives (via un import / export de fichier) ; pour les « patterns » entreprise comme pour celles correspondant aux opérateurs et structures de recherche ;
- d'identifier les liens d'associations entre une adresse et un opérateur qui permettrait d'alimenter la procédure de rapprochement géographique (cas d'une adresse ne contenant pas assez d'informations pour la rapprocher d'une entrée du référentiel géographique mais pour laquelle un opérateur à qui on peut affecter une entrée du référentiel géographique est identifié).

4.04.3 *Livrables*

- modélisation conceptuelle des données du référentiel institutionnel ;
- modélisation et cartographie des processus (mise à jour du référentiel et des rapprochements) ;



- règles de gestion applicables lors des processus (mise à jour du référentiel et des rapprochements) ;
- spécifications fonctionnelles détaillées ;
- maquette des interfaces utilisateurs (rendu IHM, ...) ;
- spécifications techniques détaillées ;
- plan de test ;
- documentations, comptes rendus d'ateliers.

#### **4.05 Mission 3 : Conception du référentiel des supports de publication**

##### *4.05.1 Objectifs et attendus de la mission*

Cette mission consiste à réaliser la conception fonctionnelle et technique du référentiel des supports de publication et des processus associés, tels que décrits dans la partie 4.05.2 ci-dessous. Sont concernés par cette mission les processus de cycle de vie du référentiel mais également ceux relatifs à l'utilisation du référentiel pour la mise en qualité et l'enrichissement des données bibliographiques du modèle métier.

Le titulaire organisera les ateliers nécessaires à l'élaboration des spécifications fonctionnelles et techniques du référentiel des supports de publication, en collaboration avec l'équipe du Hcéres dédiée.

Il élaborera la cartographie des processus de mise à jour du référentiel, de même que les spécifications fonctionnelles détaillées qui comprendront notamment les règles de gestion des rapprochements pour la mise en qualité des données et les enrichissements, ainsi que la logique de mise à jour du modèle métier.

Il élaborera ensuite les spécifications techniques détaillées en cohérence avec les autres référentiels (missions 1 et 2), le modèle métier (mission 4) et le socle technique défini lors de la mission 5.

L'implémentation du référentiel des supports de publication et des processus associés sera effectuée lors de la mission 9 sur la base des spécifications définies au cours de cette mission. L'implémentation de l'interface utilisateur sera réalisée lors de la mission 10 sur la base des spécifications définies au cours de cette mission.

##### *4.05.2 Description fonctionnelle du référentiel des supports de publication*

Les supports de publication représentent une dimension importante des données bibliographiques. En bibliométrie ils sont porteurs de spécialités (catcodes<sup>10</sup>) et peuvent être utilisés comme base de normalisation pour certains indicateurs, d'où la nécessité de les identifier rigoureusement.

Le référentiel sera composé d'une liste contenant les métadonnées unifiées et historisées pour chaque support de publication et des dictionnaires mettant en correspondance les données brutes et les entrées de la liste unifiée.

##### (a) Concepts décrits

On distinguera trois types de supports de publication dans le WoS : les revues (Journals), les livres en série (Book series) et les livres (Books). Ces types sont clairement identifiés dans les données XML. Ces trois types de support présentent des métadonnées de structure et de qualité différentes. Ils doivent être traités séparément.

#### **Revues**

Les métadonnées à retenir pour identifier une revue sont suivantes :

- Identifiant interne
- Au moins un titre. Les variantes de titre sont :
  - Titre abrégé
  - Titre long
  - Titres alternatifs (facultatif)
- Au moins un ISSN, parmi les possibilités suivantes :

---

<sup>10</sup> Voir section « complément d'information : référentiel disciplinaire, pris en charge par le Hcéres pour des détails sur les 'catcodes' »

- ISSN-print : édition imprimé
- e-ISSN : édition en ligne
- ISSN-L : ISSN de liaison (en général, un de deux précédents)
- Editeur
- Pays d'édition
- Langue(s)
- Année de début de parution (le cas échéant, dans la base OST)
- Année de fin de parution (le cas échéant, dans la base WoS)
- Spécialités (Catcodes) associées à la revue

### **Livres en série**

Les métadonnées à retenir pour identifier un livre en série sont les suivantes :

- Identifiant interne
- Un titre de série
- Un titre de volume. Les variantes de titre sont :
  - Titre abrégé
  - Titre long
  - Titres alternatifs (facultatif)
- Au moins un ISSN, parmi les possibilités suivantes :
  - ISSN-print : édition imprimé
  - e-ISSN : édition en ligne
  - ISSN-L : ISSN de liaison (en général, un de deux précédents)
- ISBN
- Editeur
- Pays d'édition
- Langue(s)
- Année de début de parution (le cas échéant, dans la base OST)
- Année de fin de parution (le cas échéant, dans la base WoS)
- Spécialités (Catcodes) associées aux volumes

### **Livres**

Les métadonnées à retenir pour identifier un livre sont les suivantes :

- Identifiant interne
- Un titre
- ISBN
- Editeur
- Pays d'édition
- Langue(s)
- Année de parution
- Spécialités (Catcodes) associées

Les actes de colloques peuvent être édités en tant que numéro spécial d'une revue, en tant que volume d'une série de livres dédiée ou en tant que livre. Dans ce dernier cas, le nom du colloque correspond exactement au titre du livre ou en fait partie.

#### **(b) Sources de données**

L'identification de l'ensemble des métadonnées pourrait se faire à partir de données brutes du WoS, de release notes de changement de titres de Clarivate Analytics, ainsi qu'à l'aide de ressources externes gratuites :

- portail public ISSN <https://portal.issn.org/>,
- NLM catalogue <https://www.ncbi.nlm.nih.gov/nlmcatalog/>,
- WorldCat <https://www.worldcat.org/>
- Sites web des éditeurs.

### (c) Cycle de vie

Pour les revues et livres en séries :

- Changement de titre mineur ne s'accompagnant pas d'un changement d'ISSN.
- Changement de titre majeur s'accompagnant d'un changement d'ISSN, sans pour autant qu'il s'agisse d'une revue différente.
- Changement de support (papier ou électronique) : un nouveau support est accompagné par un nouvel ISSN, sans pour autant qu'il s'agisse d'une revue différente
- Changement d'éditeur et éventuellement du pays, notamment à la suite de fusions-acquisitions des sociétés éditrices.
- Fusion : deux périodiques peuvent fusionner en donnant naissance à un nouveau périodique.
- Scission : un périodique peut éclater en deux ou plus nouveaux périodiques.
- Arrêt : l'éditeur peut arrêter la production d'un périodique.
- Arrêt d'indexation dans la base WoS : Clarivate Analytics peut prendre décision d'exclure un périodique du portefeuille de la base ; il peut revoir cette décision quelque temps après et réintégrer ce périodique.
- Changement de catcode (ou *subject category*) : Clarivate Analytics peut revoir les catcodes attribués à un périodique.

Les livres ne nécessitent pas d'unification car leur parution est ponctuelle et le cycle de vie simple.

Ainsi la mise à jour du référentiel comprendra :

- Les nouvelles entrées (données issues de la base actualisée) ;
- Les modifications résultantes des événements listés précédemment ;
- La correction des erreurs identifiées dans les données.

### (d) Rapprochement

Les métadonnées des supports de publication périodiques présentes dans la base WoS sont parfois lacunaires ou erronées, la constitution d'un référentiel doit donc passer par une étape d'unification (rapprochement de données).

Aucune métadonnée n'est suffisante en soi pour identifier un périodique ; l'unification doit s'appuyer sur une combinaison de métadonnées, dans un ordre défini :

1. Titre(s)
2. ISSN(s)
3. Editeur(s)

Il sera nécessaire d'introduire un identifiant interne pour chaque support de publication.

Il n'est pas toujours possible d'établir la filiation entre périodiques suite aux changements de titre, fusion ou scission faute de source d'information fiable et exhaustive, et plus particulièrement en ce qui concerne l'historique. On pourra considérer les manques dans le suivi de filiations comme un défaut prévisible et acceptable.

### (e) Interface utilisateur

En complément des fonctionnalités communes aux interfaces utilisateur associées aux référentiels, celle relative au référentiel des supports de publication doit permettre :

- d'afficher les tables de correspondance, avec les possibilités de recherche, de sélection et de tri ;
- d'aider au rapprochement ;
- d'assurer la traçabilité de mises à jour (date, auteur, critères) ;
- d'établir un lien avec le référentiel disciplinaire.

#### 4.05.3 Livrables

- modélisation conceptuelle des données du référentiel des supports de publications ;
- modélisation et cartographie des processus (mise à jour du référentiel et des rapprochements) ;
- règles de gestion applicables lors des processus (mise à jour du référentiel et des rapprochements) ;
- spécifications fonctionnelles détaillées ;
- maquette des interfaces utilisateur (rendu IHM, ...) ;
- spécifications techniques détaillées ;
- plan de test ;
- documentations, comptes rendus d'ateliers.

#### 4.06 Complément d'information : référentiel disciplinaire, pris en charge par le Hcéres

Une brève description fonctionnelle de ce référentiel est donnée ici à titre indicatif.

##### (a) Concepts décrits

Le concept de base du référentiel disciplinaire est celui de « subject category » ou 'catcode'. Ces catcodes sont regroupées suivant différentes nomenclatures utilisées pour les rapports d'indicateurs pour constituer des disciplines.

##### (b) Source de données

Les données sources contiennent le libellé de la catégorie. Il en existe 253 différentes. Elles sont affectées au niveau du document bien que relevant du support de publication de ces documents. Ce décalage provient de la structure même des données WoS au format XML où les documents sont présentés individuellement.

La documentation fournie par l'éditeur en accompagnement des données XML contient la liste des « subject categories ».

Deux nomenclatures sont utilisées régulièrement pour la production des indicateurs.

Les disciplines OST sont au nombre de 11. Les disciplines ERC sont au nombre de 25.

Dans le cas de besoins spécifiques à des projets, des regroupements ad hoc de catégories sont effectués.

##### (c) Cycle de vie

Les événements affectant la liste des « subject categories » sont assez rares (1 suppression et 2 créations depuis 2000). Lorsque cela se produit, la documentation accompagnant les données en fait explicitement mention et les utilisateurs sont avertis en amont de la mise en œuvre de l'évolution.

Les nomenclatures disciplinaires utilisées dans les rapports d'indicateurs sont également relativement stables.

##### (d) Rapprochement

Une table de correspondance associe à chacun de ces libellés un code sur 2 caractères permettant une manipulation plus facile par les chargés d'études.

La ou les discipline(s) au(x)quelle(s) une catégorie est rattachée sont affectées sur la base de tables de correspondances implémentant les nomenclatures disciplinaires (OST et ERC).

#### 4.07 Mission 4 : Conception du modèle relationnel métier

##### 4.07.1 Objectifs et attendus de la mission

Cette mission consiste à réaliser la conception fonctionnelle et technique du modèle métier et des processus associés, tels que décrits dans la partie 4.07.2 ci-dessous. Sont concernés par cette mission les processus du cycle de vie des données bibliographiques ainsi que ceux relatifs aux traitements sans lien avec un référentiel.

Le titulaire organisera les ateliers nécessaires à l'élaboration des spécifications fonctionnelles et techniques du modèle métier, en collaboration avec l'équipe du Hcéres dédiée.

Il élaborera les spécifications fonctionnelles détaillées ainsi que la cartographie des processus de mise à jour du modèle métier sans lien avec un référentiel.

Il élaborera ensuite les spécifications techniques détaillées en cohérence avec les référentiels (missions 1, 2 et 3) et le socle technique défini lors de la mission 5.

L'implémentation du modèle métier et des processus associés sera effectuée lors de la mission 6 sur la base des spécifications définies au cours de cette mission.

#### 4.07.2 Description fonctionnelle du modèle métier

##### (a) Description des données

L'entité principale des données bibliographiques est le document. Cela peut correspondre à un article scientifique, un chapitre de livre, une note...

Le document est caractérisé dans les données sources par un identifiant unique propre à la base de données Web of Science de Clarivate Analytics, seule source utilisée par le Hcéres actuellement. D'autres identifiants peuvent optionnellement figurer dans les données, notamment le Digital Object Identifier (doi) pour les publications en ligne.

Le fournisseur définit 42 types de documents, un document pouvant être de plusieurs types. Les indicateurs actuellement calculés par le Hcéres prennent en compte seulement certains types de documents.

Un document est rédigé en une ou plusieurs langues et peut comporter un ou plusieurs résumés (*abstracts*). Il peut contenir un ou plusieurs mots-clés (*keywords*). La recherche dont le document est l'objet peut avoir bénéficié de financements (*fundings*). Le document peut être ou non disponible en libre accès sur internet (*open access*).

Un document est rédigé par un ou plusieurs auteurs. Chaque auteur peut avoir, dans le cadre de ce document, une ou plusieurs affiliations. Ces adresses constituent le point d'entrée du rapprochement avec les référentiels institutionnel (dont la description fonctionnelle se trouve dans la partie 4.04.2) et géographique (dont la description fonctionnelle se trouve dans la partie 4.03.2).

Un document peut faire référence à d'autres documents, présents ou non dans la base de données du fournisseur. Il peut lui-même faire l'objet de citations de la part d'autres documents. L'éditeur fournit un fichier annexe des données sources contenant le nombre de citations reçues par chaque document. Néanmoins, les indicateurs sont calculés sur des périmètres spécifiques qui diffèrent de celui retenu par l'éditeur.

Un document est publié dans une section de support éditorial (numéro mensuel d'une revue, volume d'une collection...) qui elle-même est incluse dans un support de publication (journal, collection...). Les supports de publication font l'objet d'un référentiel dont la description fonctionnelle se trouve dans la partie 4.05.2.

Dans les données sources, un document est rattaché à une ou plusieurs 'subject category' (ou catcode). C'est au niveau du document que la discipline est utilisée pour le calcul des indicateurs.

Toutefois, l'ensemble des documents associés à une même section de support éditorial sont rattachés aux mêmes 'subject categories'. Les 'subject categories' font l'objet d'un référentiel dont la description fonctionnelle se trouve dans la partie 4.06.

##### (b) Cycle de vie

###### - Ajouts :

De nouveaux documents sont régulièrement publiés. Le fournisseur peut également ajouter à sa base de données des documents publiés précédemment mais qui n'y figuraient pas jusqu'à présent.

###### - Modifications :

Les données concernant un document peuvent subir des modifications, par exemple pour corriger des erreurs ou mettre à jour les identifiants des références. Seul l'identifiant unique propre au Web Of Science ne peut pas faire l'objet de modifications. Les données sources ne distinguent pas les modifications des nouveaux documents : les données concernant le document modifié sont livrées à nouveau en totalité.

- *Suppressions :*  
Des documents sont régulièrement supprimés des données sources. La liste de ces suppressions est livrée dans un fichier annexe, se présentant sous la forme d'une liste d'identifiants.

**Cas des citations : les décomptes de citations reçues par les documents présents dans la base sont recalculés à l'occasion des événements décrits ci-dessus.**

Toutes les étapes de ce cycle de vie feront l'objet d'une historisation dans la base bibliographique.

#### 4.07.3 Livrables

- modélisation conceptuelle des données du modèle métier ;
- modélisation et cartographie des processus ;
- règles de gestion applicables lors des processus ;
- spécifications fonctionnelles détaillées ;
- spécifications techniques détaillées ;
- plan de test ;
- documentations, comptes rendus d'ateliers.

### 4.08 Mission 5 : Socle technique de développement, intégration et validation

#### 4.08.1 Objectifs et attendus de la mission

L'objectif de cette mission transverse est de garantir un système cohérent tout au long du développement et d'assurer la bonne intégration et validation en fin de projet.

Il s'agit tout d'abord de définir et mettre en place l'environnement de développement et les éléments communs aux différentes étapes du projet.

Cela permettra aux équipes du Hcéres de réaliser ses travaux parallèlement au titulaire, tout en facilitant l'intégration des différents modules et en garantissant l'homogénéité du résultat.

Le titulaire travaillera en étroite collaboration avec les équipes du Hcéres afin de définir les éléments suivants :

- le socle technique du système dans son ensemble ;
- les outils de développement : langage(s), bibliothèques ;
- l'environnement de développement : outils logiciels, configuration, procédures de construction des binaires le cas échéant ;
- la méthodologie de suivi de version ;
- les normes et standards de programmation utilisés.

Ces choix devront naturellement prendre en compte les besoins exprimés et les spécifications élaborées lors des missions précédentes, ainsi que les contraintes techniques définies dans l'article 5. Les propositions émises par le titulaire devront faire l'objet d'un argumentaire technique précis. Il est important de noter qu'il existe déjà un environnement technique et qu'il conviendra de le conserver.

Cette mission comprend également l'intégration de l'ensemble des modules pour former le système complet. En amont, le titulaire définira un plan d'intégration et de tests associés. Puis, une fois la mise en œuvre des modules réalisée, il exécutera ces plans pour permettre la validation du résultat.

Les développements devront se faire dans l'environnement technique du Hcéres auquel le titulaire aura pleinement accès. Le titulaire aura en charge l'installation et les moyens d'installer les composants techniques de la solution mise en œuvre sur les serveurs de production.

#### 4.08.2 Livrables

Pour la définition du socle technique :

- architecture technique et logicielle du système ;
- environnement de développement ;
- consignes de développement ;
- méthodologie de suivi de version ;
- fichiers de configuration ;

- code source qui servira de squelette pour les interfaces utilisateur (par exemple, affichage minimaliste de l'écran d'accueil) ;
- procédures de construction des binaires ;
- plan d'intégration ;
- plan de tests d'intégration ;

Pour l'intégration :

- ensemble des programmes et codes source, scripts d'installation ou de modification de la base de données, scripts d'initialisation ;
- cahier de recettes contenant la liste des tests d'intégration effectués, les résultats attendus et les résultats observés ;
- documentation technique finalisée ;
- procédure détaillée d'installation ;
- système en production, avec logs d'installation.

#### **4.09 Mission 6 : Mise en œuvre du modèle métier**

##### *4.09.1 Objectifs et attendus de la mission*

Cette mission consiste à implémenter le modèle métier et les processus associés sur la base des spécifications définies lors de la mission 4. Le développement devra s'appuyer sur le socle technique mis en place lors de la mission 5 et suivre les principes et consignes définis au cours de cette même mission.

##### *4.09.2 Livrables*

- ensemble des programmes et codes source, scripts d'installation ou de modification de la base de données, scripts d'initialisation ;
- cahier de recettes contenant la liste des tests effectués, les résultats attendus et les résultats observés ;
- documentation technique.

#### **4.10 Mission 7 : Mise en œuvre du référentiel géographique**

##### *4.10.1 Objectifs et attendus de la mission*

Cette mission consiste à implémenter le référentiel géographique et les processus associés sur la base des spécifications définies lors de la mission 1, à l'exception des interfaces utilisateur (qui feront l'objet de la mission 10). Le développement devra s'appuyer sur le socle technique mis en place lors de la mission 5 et suivre les principes et consignes définis au cours de cette même mission.

##### *4.10.2 Livrables*

- ensemble des programmes et codes source, scripts d'installation ou de modification de la base de données, scripts d'initialisation ;
- cahier de recettes contenant la liste des tests effectués, les résultats attendus et les résultats observés ;
- documentation technique.

#### **4.11 Mission 8 : Mise en œuvre du référentiel institutionnel**

##### *4.11.1 Objectifs et attendus de la mission*

Cette mission consiste à implémenter le référentiel institutionnel et les processus associés sur la base des spécifications définies lors de la mission 2, à l'exception des interfaces utilisateur (qui feront l'objet de la mission 10). Le développement devra s'appuyer sur le socle technique mis en place lors de la mission 5 et suivre les principes et consignes définis au cours de cette même mission.

##### *4.11.2 Livrables*

- ensemble des programmes et codes source, scripts d'installation ou de modification de la base de données, scripts d'initialisation ;
- cahier de recettes contenant la liste des tests effectués, les résultats attendus et les résultats observés ;



- documentation technique.

#### **4.12 Mission 9 : Mise en œuvre du référentiel des supports de publication**

##### *4.12.1 Objectifs et attendus de la mission*

Cette mission consiste à implémenter le référentiel des supports de publication et les processus associés sur la base des spécifications définies lors de la mission 3, à l'exception des interfaces utilisateur (qui feront l'objet de la mission 10). Le développement devra s'appuyer sur le socle technique mis en place lors de la mission 5 et suivre les principes et consignes définis au cours de cette même mission.

##### *4.12.2 Livrables*

- ensemble des programmes et codes source, scripts d'installation ou de modification de la base de données, scripts d'initialisation ;
- cahier de recettes contenant la liste des tests effectués, les résultats attendus et les résultats observés ;
- documentation technique.

#### **4.13 Mission 10 : Mise en œuvre des interfaces utilisateur**

##### *4.13.1 Objectifs et attendus de la mission*

Cette mission a pour objet d'implémenter les interfaces utilisateur définies lors des missions précédentes. Le développement devra s'appuyer sur le socle technique mis en place lors de la mission 5 et suivre les principes et consignes définis au cours de cette même mission.

##### *4.13.2 Livrables*

- ensemble des programmes et codes source, scripts d'installation ou de modification de la base de données, scripts d'initialisation ;
- cahier de recettes contenant la liste des tests effectués, les résultats attendus et les résultats observés ;
- documentation utilisateur ;
- documentation technique.

## **ARTICLE 5 - EXIGENCES TECHNIQUES**

---

### **5.01 Exigences techniques générales**

La base de données utilisée est Oracle 11g (le Hcéres dispose d'une licence Oracle Enterprise).

Pour des raisons de performance, les traitements de données seront réalisés dans la mesure du possible directement en base de données.

Le chargement des données et l'orchestration des traitements seront réalisés via l'outil Talend Integration Suite.

Pour les raisons d'évolutivité et de maintenabilité du système évoquées dans la partie 4.01, les modèles et le code produits par le titulaire devront :

- être accompagnés de documentation technique détaillée ;
- permettre une maintenance et une reprise aisées, notamment en utilisant :
  - des noms d'objets et de variables pertinents et cohérents dans l'ensemble du système ;
  - des commentaires dans le code pour indiquer la signification des variables, fonctions, constantes (y compris dans les fichiers de paramétrage) ;
  - des commentaires en base de données (tables et colonnes) pour indiquer la signification et / ou l'utilisation des tables et colonnes ;
- être conçus de façon à permettre une évolution, que ce soit au niveau des modèles de données ou des processus.

## 5.02 Exigences techniques relatives aux interfaces utilisateur

Les interfaces utilisateur seront accessibles par tous les utilisateurs depuis un client léger (navigateur internet) et compatible avec les navigateurs du marché (Firefox, Chrome, Internet Explorer, Safari) dans leurs versions les plus utilisées et supérieures.

La solution devra intégrer une fonctionnalité d'authentification unique SSO qui sera rendue opérationnelle après la mise en œuvre de l'outil. Le Hcéres menant une réflexion de mise en œuvre d'une solution d'authentification SSO qui sera connectée à toutes les applications gérées par le Hcéres, l'outil devra intégrer cette fonctionnalité sans être activée lors de la mise en production de l'outil dans un premier temps.

Elle devra intégrer un mécanisme de gestion des droits d'accès.

La solution appliquera l'identité visuelle du Hcéres dont les éléments sont :

- logo :



- police utilisé pour les supports digitaux du Hcéres : Muli

-

### COULEURS INSTITUTIONNELLES

#### RÉFÉRENCES ET ÉQUIVALENCES

PANTONE  
Quadri CMJN  
Réfs RVB  
Réfs HEXA



Pantone 267 C  
C 80 - M 95  
R 92 - V 45 - B 145  
# 5C2D91



Pantone 1925 C  
M 100 - J 50  
R 237 - V 20 - B 91  
# ED145B

Le titulaire devra mettre en œuvre les bonnes pratiques de développement de façon à respecter les règles de sécurités telles que décrites par l'Open Web Application Security Project, pour assurer la sécurité et éviter la vulnérabilité de l'outil lorsqu'il sera mis en production.

Le titulaire devra mettre en œuvre les bonnes pratiques de développement de façon à respecter l'application du Règlement général sur la protection des données.

### 5.03 Éléments de volumétrie

Le titulaire devra prendre en compte les éléments de volumétrie suivants :

Élément	Ordre de grandeur	Commentaire
Nombre de documents par an	De 1 à 2,5 millions	a tendance à augmenter
Nombre de documents au total actuellement (1999-2018)	37 millions	
Poids de la base de données du système existant	3 To	ne sont pas inclus les schémas projets, utilisateurs, et de développement
Poids des fichiers XML du stock par an	De 10 à 50 Go	
Poids des fichiers XML du stock 1999-2018	570 Go	
Poids des fichiers XML hebdomadaires pour un an	400 Go	Estimation à partir des 20 premières semaines de 2019

## ARTICLE 6 - PRESTATIONS ET LIVRABLES ATTENDUS

Tous les livrables seront soumis à validation du Hcéres.

Le tableau suivant récapitule les livrables attendus :

N° mission	Intitulé de la mission	Sous-partie	Livrables
1	Conception du référentiel géographique	Conception fonctionnelle	<ul style="list-style-type: none"> <li>- modélisation conceptuelle des données du référentiel géographique</li> <li>- modélisation et cartographie des processus (mise à jour du référentiel et des rapprochements)</li> <li>- règles de gestion applicables lors des processus (mise à jour du référentiel et des rapprochements)</li> <li>- spécifications fonctionnelles détaillées</li> <li>- maquette des interfaces utilisateur (rendu IHM, ...)</li> </ul>
		Conception technique	<ul style="list-style-type: none"> <li>- spécifications techniques détaillées</li> <li>- plan de test</li> <li>- documentations, comptes rendus d'ateliers</li> </ul>
2	Conception du référentiel institutionnel	Conception fonctionnelle	<ul style="list-style-type: none"> <li>- modélisation conceptuelle des données du référentiel institutionnel</li> <li>- modélisation et cartographie des processus (mise à jour du référentiel et des rapprochements)</li> <li>- règles de gestion applicables lors des processus (mise à jour du référentiel et des rapprochements)</li> <li>- spécifications fonctionnelles détaillées</li> <li>- maquette des interfaces utilisateur (rendu IHM, ...)</li> </ul>
		Conception technique	<ul style="list-style-type: none"> <li>- spécifications techniques détaillées</li> <li>- plan de test</li> <li>- documentations, comptes rendus d'ateliers</li> </ul>
3	Conception du référentiel des supports de publication	Conception fonctionnelle	<ul style="list-style-type: none"> <li>- modélisation conceptuelle des données du référentiel des supports de publications</li> <li>- modélisation et cartographie des processus (mise à jour du référentiel et des rapprochements)</li> <li>- règles de gestion applicables lors des processus</li> </ul>

			(mise à jour du référentiel et des rapprochements) - spécifications fonctionnelles détaillées - maquette des interfaces utilisateur (rendu IHM, ...)
		Conception technique	- spécifications techniques détaillées - plan de test - documentations, comptes rendus d'ateliers
4	Conception du modèle relationnel métier	Conception fonctionnelle	- modélisation conceptuelle des données du modèle métier - modélisation et cartographie des processus - règles de gestion applicables lors des processus - spécifications fonctionnelles détaillées
		Conception technique	- spécifications techniques détaillées - plan de test - documentations, comptes rendus d'ateliers
5	Socle technique de développement, intégration et validation	Définition du socle technique	- architecture technique et logicielle du système ; - environnement de développement ; - consignes de développement ; - méthodologie de suivi de version ; - fichiers de configuration ; - code source qui servira de squelette pour les interfaces utilisateur (par exemple, affichage minimaliste de l'écran d'accueil) ; - procédures de construction des binaires ; - plan d'intégration ; - plan de tests d'intégration ;
		Intégration	- ensemble des programmes et codes source, scripts d'installation ou de modification de la base de données, scripts d'initialisation ; - cahier de recettes contenant la liste des tests d'intégration effectués, les résultats attendus et les résultats observés ; - documentation technique finalisée ; - procédure détaillée d'installation ; - système en production, avec logs d'installation.
6	Mise en œuvre du modèle métier		- ensemble des programmes et codes source, scripts d'installation ou de modification de la base de données, scripts d'initialisation ; - cahier de recettes contenant la liste des tests effectués, les résultats attendus et les résultats observés ; - documentation technique.
7	Mise en œuvre du référentiel géographique		- ensemble des programmes et codes source, scripts d'installation ou de modification de la base de données, scripts d'initialisation ; - cahier de recettes contenant la liste des tests effectués, les résultats attendus et les résultats observés ; - documentation technique.
8	Mise en œuvre du référentiel institutionnel		- ensemble des programmes et codes source, scripts d'installation ou de modification de la base de données, scripts d'initialisation ; - cahier de recettes contenant la liste des tests effectués, les résultats attendus et les résultats observés ; - documentation technique.
9	Mise en œuvre du référentiel des		- ensemble des programmes et codes source, scripts d'installation ou de modification de la

	supports de publication		base de données, scripts d'initialisation ; - cahier de recettes contenant la liste des tests effectués, les résultats attendus et les résultats observés ; - documentation technique.
10	Mise en œuvre des interfaces utilisateur		- ensemble des programmes et codes source, scripts d'installation ou de modification de la base de données, scripts d'initialisation ; - cahier de recettes contenant la liste des tests effectués, les résultats attendus et les résultats observés ; - documentation utilisateur ; - documentation technique.

## ARTICLE 7 - CALENDRIER PREVISIONNEL DU PROJET

Le planning prévisionnel souhaité du projet est le suivant :

- ➔ Achèvement de toutes les missions pour fin mars 2020, toutes validations effectuées (VA et VSR) ;
- ➔ L'achèvement de la mission M4 présente un intérêt particulier pour le Hcéres afin de pouvoir concevoir et réaliser les développements nécessaires au maintien des outils existants qui ne font pas l'objet de ce marché.

Durant le déroulement de la prestation, les missions de mise en œuvre peuvent démarrer avant l'achèvement complet de toutes les missions de conception. Dans le même esprit, les travaux de mise en œuvre peuvent permettre d'affiner les spécifications dans un processus itératif.

Les experts fonctionnels sollicités dans le cadre des ateliers de spécifications fonctionnelles sont les mêmes pour les trois référentiels. Il est donc nécessaire d'étaler dans le temps les missions de conception des référentiels.

Le soumissionnaire devra proposer un planning de déroulement du projet. Le planning proposé devra être motivé afin que le Hcéres puisse en apprécier sa pertinence.

## ARTICLE 8 - GARANTIE

Le titulaire s'engagera pour une période de 12 mois, à compter de la recette définitive des outils marquée par la recette en vérification de service régulier (VSR), sur une garantie de bon fonctionnement des outils livrés et s'oblige à remédier à tout dysfonctionnement en apportant les corrections nécessaires.

Le titulaire remet au Hcéres :

- les composants logiciels de la nouvelle version corrigée (les programmes sources et exécutables, la liste des fichiers, des tables) ;
- la mise à jour des outils connexes si nécessaire ;
- la documentation technique mise à jour ;
- les protocoles de tests d'intégration, de non-régression et de performance, les résultats attendus et les résultats constatés.

## ARTICLE 9 - CADRE DE REPONSE

Afin de permettre une lecture et une analyse comparative des offres, il est demandé à ce que la réponse respecte la structure suivante :

- Présentation générale
  - o présentation générale de la société ;

- références significatives ;
  - références et contacts de clients ayant déjà bénéficié des prestations similaires à celle de la proposition commerciale ;
  - organisation et gouvernance projet prévues ;
  - moyens humains affectés au projet et CV des différents profils ;
  - moyens techniques dédiés pour la réalisation des prestations ;
  - méthodologie de suivi du projet et de suivi de la qualité et outils utilisés ;
  - partenariats technologiques.
- Présentation détaillée de l'offre
    - Note de compréhension des problématiques exposées
    - architecture envisagée de la solution, socle technique, composition des différentes briques et type de licence associée ;
    - présentation des modalités de réalisation et des briques techniques sous-jacentes envisagées ;
    - présentation de l'ergonomie de la solution : maquette d'écrans pour les interfaces utilisateur ;
    - démarche projet, description détaillée des étapes ;
    - présentation de la méthode d'animation des ateliers de travail ;
    - planning, charges prévisionnelles, hypothèses de chiffrage.