



CAHIER DES CLAUSES TECHNIQUES PARTICULIÈRES

MARCHÉ PUBLIC DE FOURNITURES COURANTES ET DE SERVICES

**Acquisition de solutions informatiques de gestion de
base de données**

SOMMAIRE

1. Présentation du cadre	4
1.1 Présentation du programme ANR RCF3	4
1.1.1 <i>Description du corpus :</i>	5
1.1.2 <i>Objectifs et positionnement scientifiques</i>	6
1.2 Structure institutionnelle du consortium	6
2. Objet du marché, dispositions générales	7
2.1 Objet du marché	7
2.1.1 Création d'un formulaire de transcription des données des registres des feux du XIXe siècle et import des données dans le graphe RDF	7
2.1.2 Création d'un outil logiciel d'étiquetage des registres de comité et d'assemblée et des corpus issus des fonds publics canadiens	8
2.1.3 Extension du dépôt RDF	8
2.1.4 Création d'un outil de visualisation des registres de comité et d'assemblée	9
2.1.5 Création d'un outil de visualisation transversal	9
2.2. Lieu de livraison des prestations	9
2.3 Calendrier prévisionnel	9
Article 7. Maintenance	10
7.1. Dispositions Générales	10
7.2. Niveaux d'anomalies	10
7.3. Délai de résolution	10
7.4. Résolution des anomalies des logiciels	11
7.4.1. Résolution des anomalies de niveau 1 (anomalies bloquantes)	11
7.4.2. Résolution des anomalies de niveau 2 (anomalies majeures)	11
7.4.3. Résolution des anomalies de niveau 3 (anomalies mineures)	11

L'ESSENTIEL DU CONTRAT

	Objet	Acquisition de solutions informatiques de gestion de base de données
	Type de contrat	Marché public
	Tranches optionnelles	Sans tranches optionnelles
	Clauses sociales	Sans
	Critère Environnemental	Avec
	Durée / Délai	44 mois
	Prix	Prix global forfaitaire
	Variation des prix	Sans
	Avance	Avec

1. Présentation du cadre

1.1 Présentation du programme ANR RCF3

La Comédie-Française est une des plus anciennes sociétés de comédiens d'Europe. Fondée en 1680, cette institution théâtrale a conservé la quasi-totalité de ses archives. Le programme des Registres de la Comédie-Française (RCF) vise à rendre accessibles les informations contenues dans ces registres¹ par la création de bases de données, par la création d'outils de recherche spécifiques donnant accès aux bases de données, et par l'existence d'un dépôt RDF.

Le programme des Registres de la Comédie-Française repose sur la rencontre entre des technologies récentes appliquées aux sciences humaines et des archives historiques ressortissant au patrimoine théâtral français. Il s'agit d'un **programme international** qui réunit plusieurs institutions et pays : le théâtre de la Comédie-Française, l'Université de Victoria (Canada), New York University (Etats-Unis), et, en France, l'Université Paris Nanterre, Sorbonne Université et l'Université Paul-Valéry de Montpellier.

Ce programme a débuté en 2008 et s'est développé autour de plusieurs phases successives au cours desquelles différents corpus quantitatifs ont été numérisés et constitués en bases de données grâce à différents financeurs. Entre 2008 et 2024, le programme RCF a développé, grâce à ses financements internationaux, plusieurs bases :

- Une base de données pour les recettes du XVIII^e siècle
- Une base de données pour les feux du XVIII^e siècle
- Une base de données pour les dépenses du XVIII^e siècle
- Une base de données pour les recettes du XIX^e siècle
- Un dépôt RDF qui structure la plupart de ces données quantitatives (recettes, dépenses et feux du XVIII^e siècle)

Actuellement, le programme RCF entre dans sa troisième phase (2025-2030). Son objectif sera **d'étendre les bases de données quantitatives** existantes à **de nouveaux jeux de données**, cette fois-ci principalement **qualitatifs**. L'enjeu scientifique est d'éclairer les données chiffrées par des corpus textuels, qui permettent de donner sens et d'interpréter au mieux ces données chiffrées.

Il s'agira principalement pour le titulaire du marché :

- De concevoir une interface de saisie et une base pour les données issues des registres des feux du XIX^e siècle,
- D'exporter les données des feux du XIX^e sur le graphe RDF,
- De concevoir un outil d'encodage pour les registres de comité et d'assemblée des XVIII^e et XIX^e siècles (corpus manuscrits), mais aussi pour les corpus de presse, mémoires et témoignages relatifs à la programmation de la Comédie-Française (corpus imprimés),
- De concevoir des outils de visualisation pour les feux XIX^e et pour les assemblées, comités et ressources documentaires complémentaires,

¹ Voir 1.2.

- D'exporter les données qualitatives issues des registres de comité et d'assemblée dans le graphe RDF,
- D'adapter le graphe RDF à l'intégration de ces nouvelles données qualitatives,
- De concevoir un outil de visualisation transversale qui traitera l'ensemble des données qualitatives et quantitatives des différents corpus.

Ce programme bénéficie de **deux financements** :

- un financement canadien du Centre de Recherche en Sciences Humaines (CRSH) obtenu par Sara Harvey (Université de Victoria)
- un financement français de l'Agence Nationale de la Recherche (ANR), obtenu par Florence Naugrette (Sorbonne Université) et Charline Granger (Université de Montpellier Paul-Valéry).

Chacun de ces financements couvre **des corpus différents** :

- Au Canada, Sara Harvey et son équipe traitent de corpus de presse, mais aussi des mémoires et témoignages relatifs à la programmation de la Comédie-Française au XVIII^e siècle. Ce corpus est numérisé sur Gallica, en France, pour la majeure partie.
- En France, Florence Naugrette et Charline Granger s'occupent de traiter les registres de « feux » (distribution) du XIX^e siècle et les registres d'assemblée et de comité des XVIII^e et XIX^e siècles. Ce corpus vient des fonds conservés à la Comédie-Française.

Le présent cahier des charges concerne les corpus de la Comédie-Française.

1.1.1 Description du corpus :

- Les registres de « feux » (distribution) du XIX^e siècle

Un « feu » est l'indemnité versée aux artistes les soirs où ils jouent. Les registres des comptes qui les consignent représentent un corpus de 45000 pages pour la période 1765-1900 : ils précisent la distribution pour chaque séance et la répartition des rôles hommes/femmes, sociétaires/pensionnaires. Y figurent aussi des commentaires relatifs aux répétitions, aux « débuts » des artistes, aux reprises de rôle, aux relâches, aux tournées, aux événements politiques. Ces registres (au nombre de 91 pour la période 1799-1899 qui sera l'objet du programme RCF3) sont classés par années théâtrales.

- Les registres d'assemblée et de comité des XVIII^e et XIX^e siècles

Les registres d'assemblée ou de comité constituent un corpus d'environ 40 000 pages pour les XVIII^e, XIX^e et XX^e siècles : ils contiennent les comptes rendus des comités d'administration qui réunissaient le doyen d'âge et les sociétaires élus par leurs pairs ou la tutelle, en présence (selon les périodes et les régimes administratifs de la Maison) du commissaire du Roi auprès du Théâtre-Français, du directeur-gérant de la Société des Comédiens ou du l'administrateur-général. Tenus par les sociétaires et rédigés par leur secrétaire, ils consignent les problèmes et les décisions rendues par le comité de manière hebdomadaire : carrières d'acteurs, dépenses, relations avec la tutelle, règlements intérieurs, autres métiers, etc. Au nombre de 50, ces registres regroupent par volume les

procès-verbaux du comité d'administration de plusieurs années théâtrales. Ils font l'objet d'une transcription au format texte.

Ces deux corpus sont entièrement numérisés et accessibles en *open access* sur le site La Grange (qui inventorie les archives de la Bibliothèque-Musée de la Comédie-Française et héberge les archives qui ont fait l'objet d'une numérisation)² et sur le site du Programme des Registres de la Comédie-Française³. Les registres de comité du XIX^e siècle (14 registres) et du XX^e siècle (3 registres) ont fait l'objet d'une transcription sous format Word. Ceux du XVIII^e siècle (33 registres), quant à eux, restent pour la plupart à transcrire et feront l'objet d'une transcription par *Automatic Text Recognition*. Un modèle d'*Handwritten Text Recognition* spécifique aux registres d'assemblée du XVIII^e siècle a été développé par Marine Tiger (Sorbonne Université) et permettra prochainement de transcrire ce corpus.

1.1.2 Objectifs et positionnement scientifiques

Ces registres renseignent sur la conception du répertoire, la proportion des reprises et des créations, les logiques de programmation et de distribution, les politiques culturelles, la jurisprudence, le rôle des comédiens dans l'émergence des dramaturgies nouvelles, les liens qu'ils entretiennent avec les auteurs et entre eux (composition de la troupe, rapports de force selon l'âge et le genre, querelles, vedettariat, etc.). C'est donc un pan peu connu de l'histoire du théâtre (centrée sur la troupe) qui sera frayé et théorisé. En concevant à l'aide du Titulaire du présent marché un outil d'exploration et de visualisation ergonomique, RCF3 permettra l'interrogation rapide des registres de feux et d'assemblée et l'interopérabilité de ces deux bases avec les bases RCF existantes.

1.2 Structure institutionnelle du consortium

Piloté par Charline Granger (CNRS – IRCL (Université Paul-Valéry de Montpellier)) et Florence Naugrette (Sorbonne Université – CELLF), le programme RCF 3 est en partenariat avec 13 institutions, dont 7 universités françaises, 2 états-uniennes, 1 canadienne, 1 écossaise, 1 entreprise publique et une entreprise privée. Il rassemble ainsi 20 chercheurs et EC titulaires en Études théâtrales, en Littérature française, en Histoire, en Humanités numériques et en informatique.

Établissement	Lieux	Laboratoire	Participants
Sorbonne Université	Paris	CELLF	F. Naugrette (PR co-directrice de RCF3, porteuse de projet), S. Marchand (MCF), V. Yvernault (MCF), M. Alrahabi (IR), R. Bret-Vitoz (PR), M. Bouchardon (PR)
CNRS – Université Paul Valéry	Montpellier	IRCL	C. Granger (CR, co-directrice de RCF3)

² <https://comedie-francaise.bibli.fr>

³ <https://www.cfregisters.org/#/>

Paris Nanterre Université	Nanterre	HAR	T. Karsenti (PR, directrice du programme ANR RCF2), C. Bouteille (MCF)
Comédie-Française	Paris	Bibliothèque -Musée	L.-G. Pairault (conservateur-archiviste)
Université de Victoria	Canada	Fonds publics canadiens du CRSH	Sara Harvey (PR)

2. Objet du marché, dispositions générales

2.1 Objet du marché

Le présent marché a pour objet l'acquisition d'une solution informatique complète de la conception au déploiement pour :

- La création d'un formulaire de saisie des registres des feux du XIX^e siècle,
- L'import des données saisie des feux du XIX^e siècle dans le graphe RDF,
- La création d'un outil de visualisation des registres des feux des XVIII^e et XIX^e siècles,
- La création d'un outil logiciel d'étiquetage automatique des textes (format texte, XML et XML-TEI)
- L'import des données étiquetées dans le graphe RDF,
- La création d'un outil de visualisation des registres d'assemblée, de comité et des corpus complémentaires traités au Canada (presse, mémoires, etc. : voir plus loin),
- La création d'un outil transversal de visualisation de l'ensemble des données (aussi bien qualitatives que quantitatives) du graphe RDF (recettes, dépenses et feux du XVIII^e siècle, recette et feux du XIX^e siècle, comité et assemblée des XVIII^e et XIX^e siècles, corpus de presse, mémoires et témoignages traités au Canada).

2.1.1 Création d'un formulaire de transcription des données des registres des feux du XIX^e siècle et import des données dans le graphe RDF

Le formulaire de saisie permettra d'enrichir la base des feux du XVIII^e siècle. On **réutilisera**

- les identifiants uniques des pièces et des rôles de la base des feux du XVIII^e siècle
- les identifiants uniques des pièces représentées pour la première fois au XIX^e siècle fournis par la base des recettes du XIX^e siècle.

Le formulaire de saisie devra également permettre de **créer de nouveaux identifiants** (acteurs, pièces, rôles), en particulier pour les rôles des pièces nouvelles du XIX^e siècle (qui n'ont pas encore d'identifiant unique).

Le formulaire de saisie devra permettre aux utilisateurs de renseigner les informations suivantes présentées dans le registre :

- Date de la séance (représentation théâtrale)
- L'horaire de la représentation (matinée, soirée)
- Pièce(s) représentée(s) (entre 1 et 4 par jour, par ordre de représentation)
- Les acteurs distribués et le rôle qu'ils ont tenu

- Le(s) lieu(x) de représentation (parfois extérieur(s) à la Comédie-Française pour une ou plusieurs pièces par soirée)

Un outil permettra à un administrateur de créer des instances utilisateurs pour donner accès aux pages de formulaire, classées par registre ou par année civile. Il ne sera pas nécessaire d'intégrer une image de la page du registre sur l'interface de saisie — les registres numérisés étant disponibles en *Open Access* sur La Grange et le site cfregiters.org.

Le prestataire **importera les données saisies dans le graphe RDF**.

2.1.2 Création d'un outil logiciel d'étiquetage des registres de comité et d'assemblée et des corpus issus des fonds publics canadiens

Un outil logiciel d'étiquetage des textes **permettra d'encoder** :

- Les corpus de presse, mémoires et témoignages relatifs à la programmation de la Comédie-Française (corpus de Sara Harvey, Université de Victoria)
- Les registres de comité et d'assemblée des XVIII^e, XIX^e et XX^e siècles (corpus de Charline Granger et Florence Naugrette, Université Paul-Valéry de Montpellier et Sorbonne Université).

Le développement de cet outil logiciel comporte **deux phases**, qui feront l'objet de **deux financements** :

- **PHASE 1** : création d'une interface web permettant de sauvegarder les textes (en format texte et XML-TEI), d'ajouter ou supprimer des entités nommées et de rechercher une entité. **Cette phase 1 sera financée par le CRSH au Canada.**
- **PHASE 2** : développement d'un outil de reconnaissance automatique d'entités nommées. **Cette phase 2 sera financée par l'ANR en France.**

Le titulaire du présent marché sera en charge du développement de la PHASE 2.

L'outil de reconnaissance d'entités nommées **permettra d'encoder** :

- Les noms de personne
- Les noms de lieux
- Les dates
- Les titres des pièces

L'outil d'étiquetage sera à la fois :

- **Automatique** : en s'appuyant sur le travail d'étiquetage déjà réalisé sur le corpus de presse, mémoires et témoignage par le CRSH (ce qui constitue un corpus étiqueté de 30 000 pages), l'outil logiciel étiquètera automatiquement les entités connues. Puis il étiquètera automatiquement les nouvelles entités que l'utilisateur pourra créer.
- **Semi-automatique** : il permettra à l'utilisateur de vérifier et éventuellement de corriger l'entité nommée reconnue automatiquement par la machine (notamment en cas d'homonymie).
- **Manuel** : il permettra à l'utilisateur de créer de nouvelles entités nommées manuellement, qui enrichiront l'outil de reconnaissance automatique.

Les corpus à étiqueter seront au format texte (Word), XML et XML-TEI.

2.1.3 Extension du dépôt RDF

Toutes les nouvelles données quantitatives et qualitatives devront être **importées sur le graphe RDF**.

Le graphe RDF devra être adapté à l'extension de notre écosystème de données, issu des corpus des différents siècles (XVIII^e, XIX^e, et XX^e siècles) et en particulier à l'intégration des nouvelles données qualitatives.

La restructuration de nos jeux de données devra permettre la mise en relation des données qualitatives et quantitatives issues de l'ensemble de nos corpus (anciens et nouveaux).

2.1.4 Création d'un outil de visualisation des registres de comité et d'assemblée

L'outil de visualisation des assemblées et des comités sera développé. Il permettra :

- d'afficher le texte du procès-verbal d'une assemblée ou d'un comité
- de faire des recherches plein texte dans les procès-verbaux
- de visualiser ergonomiquement les entités nommées et de faire des recherches ciblées grâce à ces entités nommées

2.1.5 Création d'un outil de visualisation transversal

Un **outil de visualisation transversal** permettra de mettre en relation l'ensemble de nos jeux de données et de les visualiser de manière originale.

2.2. Lieu de livraison des prestations

<https://www.cfregisters.org/fr/>

2.3 Calendrier prévisionnel

Phases	Livraison
Création de l'interface de saisie des feux 18e-19e siècles	M+4
Création de l'outil de reconnaissance automatique d'entités nommées pour les assemblées et corpus complémentaires	M+6
Création d'un outil de visualisation des feux 18e-19e siècles	M+12
Création de l'outil de visualisation des registres de comité et d'assemblée et les corpus complémentaires (à partir des entités nommées)	M+24

Création de l'outil de visualisation transversal permettant d'interroger l'ensemble des jeux de données du programme RCF	M+28
Imports des données dans le graphe RDF	M+30
Maintenance et vérification de service régulier	M+44

La mention M0 correspondant à la date de notification du marché.

Article 7. Maintenance

7.1. Dispositions Générales

La maintenance corrective s'entend comme les mesures consistant à corriger les anomalies.

Cette prise en charge pourra se faire soit par livraison d'une nouvelle version du logiciel soit par la mise à disposition d'une solution de contournement. La maintenance corrective des logiciels comprend les interventions demandées par la personne publique en cas d'anomalies, c'est-à-dire de difficultés de fonctionnement du logiciel répétitives et reproductibles, au regard des spécifications fonctionnelles.

Par ailleurs, le titulaire doit fournir un support, c'est-à-dire une assistance interactive (assistance téléphonique, messages électroniques ...) pour la résolution des problèmes techniques rencontrés par la personne publique dans l'utilisation du ou des logiciels fournis.

7.2. Niveaux d'anomalies

La maintenance corrective des logiciels comprend les interventions demandées par l'Université en cas d'anomalies, c'est-à-dire de difficultés de fonctionnement du logiciel (répétitives et reproductibles).

Ces anomalies sont classées en trois niveaux :

- Niveau 1 : Bloquant. Anomalie rendant une ou plusieurs fonctions totalement indisponibles.
- Niveau 2 : Majeur. Anomalie rendant une fonction fortement dégradée (temps de réponse supérieur au double des temps de réponse définis dans les exigences du CCTP) ou impactant plusieurs fonctionnalités.
- Niveau 3 : Mineur. Anomalie n'impliquant pas de forts ralentissements.

Les demandes d'information ou d'évolution des logiciels sont classées dans le dernier niveau 4.

- Niveau 4 : Information. Le logiciel fonctionne correctement. La personne publique attend des informations quant à l'utilisation ou à l'exploitation du logiciel. Ces informations doivent être fournies sous 48 heures.

Ce dernier niveau prévoit une pénalité de 50 € HT par jour de retard pour le Titulaire en cas de non-respect des engagements des délais de prise en compte et des délais de résolution.

7.3. Délai de résolution

Le décompte du délai de résolution commence dès sollicitation du Titulaire par l'Université par l'un des moyens mis à sa disposition et clairement définis dans la proposition du Titulaire (Site internet, Numéro de téléphone de l'assistance en ligne ou de la messagerie électronique).

En cas d'anomalie de niveau 1 à 3 pour un logiciel, le Titulaire doit respecter le délai de résolution pour le niveau de gravité correspondant.

En cas de non-respect de ce délai de résolution des pénalités pourront être appliquées.

Le décompte du délai de résolution imparti au titulaire pour répondre à une demande de sollicitation ne court que pendant la période de sollicitation. La période de suivi court du lundi au vendredi de 9h00 à 17h30, hors samedi, dimanche et jours fériés.

7.4. Résolution des anomalies des logiciels

Une anomalie est résolue lorsqu'il est contradictoirement constaté que le titulaire y a mis fin.

7.4.1. Résolution des anomalies de niveau 1 (anomalies bloquantes)

Le Titulaire dispose d'un délai de résolution maximum de 4h pour corriger l'anomalie ou fournir une solution de contournement, si elle existe, à compter de la demande de résolution.

Cette obligation est une obligation de résultat.

Le Titulaire doit informer la personne publique de la durée des anomalies, s'il estime que celle-ci doit dépasser la durée maximum fixée par le marché.

7.4.2. Résolution des anomalies de niveau 2 (anomalies majeures)

Le titulaire dispose d'un délai de résolution maximum de deux jours ouvrés pour corriger l'anomalie ou fournir une solution de contournement, si elle existe, à compter du lendemain de la demande de résolution

Le respect de délai est une obligation de résultat.

7.4.3. Résolution des anomalies de niveau 3 (anomalies mineures)

Le titulaire dispose d'un délai de résolution maximum de trois jours ouvrés pour corriger l'anomalie ou fournir une solution de contournement, si elle existe, à compter du lendemain de la demande de résolution

Le respect des délais est une obligation de résultat.