Les registres de la Comédie-Française sur le Web de données liées : de l'hétérogénéité de données vers des données quantitatives en RDF

Charline Granger, Fabien Amarger

Les registres de la Comédie-Française sur le Web de données liées : de l'hétérogénéité de données vers des données quantitatives en RDF

Charline Granger[1], Fabien Amarger[2]
[1]Université Paris Nanterre, 200 avenue de la République, Nanterre, France
charline.granger@yahoo.com
[2]Logilab, 104 avenue Auguste Blanqui, Paris, France
fabien.amarger@logilab.fr

Résumé

La Comédie-Française est une troupe de théâtre née de la volonté centralisatrice de Louis XIV en 1680 : troupe au statut ambigu, elle est pensionnée par le roi, mais fonctionne aussi comme une entreprise privée, dont les sociétaires se partagent les parts. Pour administrer ce partage et permettre le contrôle du pouvoir royal, les comédiens de la compagnie produisent dès sa fondation des registres journaliers, partiellement manuscrits, où ils consignent les recettes et les dépenses. Ces archives comptables sont uniques en leur genre par leur densité et leur continuité, puisque les registres ont été tenus à jour quotidiennement par les membres de la compagnie. Ces données sont particulièrement importantes pour comprendre la vie théâtrale de l'époque. Il n'y a actuellement pas d'autre moyen pour étudier ces registres que de les analyser manuellement, c'est pourquoi le programme de recherche des Registres de la Comédie-Française (RCF) a initié leur transcription et la constitution de bases de données, pour en faciliter l'étude. Ici, nous nous intéressons à la transcription des données des dépenses, qui sont particulièrement hétérogènes. La seconde partie de notre proposition porte sur l'alignement de ces données des dépenses sur les données transcrites antérieurement, qui proviennent des registres des recettes et des feux. L'objectif est de publier toutes ces données dans un même entrepôt SPARQL, afin de permettre des analyses quantitatives transversales.

Mots-clés

archives, history of theatre, ontology, Linked Open Data, Digital Humanities

1 Introduction

La Comédie-Française est une troupe de théâtre née de la volonté centralisatrice de Louis XIV en 1680 : troupe au statut ambigu, elle est pensionnée par le roi, mais fonctionne aussi comme une entreprise privée, dont les sociétaires se partagent les parts. Pour administrer ce partage et permettre le contrôle du pouvoir royal, les comédiens de la compagnie produisent dès sa fondation des registres journaliers, partiellement manuscrits, où ils consignent les recettes et les dépenses. Ces archives comptables sont uniques en leur genre par leur densité et leur continuité, puisque ces registres ont été tenus à jour quotidiennement jusqu'aujourd'hui par les membres de la troupe.

Afin de donner accès à cette archive précieuse, un programme de recherche international bilingue, le programme des Registres de la Comédie-Française [1] [4] a été fondé il y a une dizaine d'années[1]. Il regroupe la Comédie-Française et plusieurs Universités : Paris Nanterre, Sorbonne Université, l'Université de Rouen Normandie, l'Université de Victoria (Canada), New York University et le Massachusetts Institute of Technology. Ce programme a d'abord été consacré à la création d'une base de données concernant les recettes journalières de la troupe entre 1680 et 1793, ce qui représente 113 saisons et plus de 30 000 soirées programmées : les registres de recettes donnent accès à la programmation (habituellement deux pièces par soirée) et au nombre de billets vendus par catégorie de place. Deux bases de données relationnelles ont ainsi été créées : d'une part, la base de données des recettes, qui associe à une date de représentation les titres des pièces jouées ce soir-là, les noms de leurs auteurs et le genre dramatique auquel elles ressortissent ; d'autre part, la base des "feux" (distribution par représentation), qui associe, sur la période 1765-1793, une date de représentation à des noms d'acteurs et d'actrices, eux-mêmes associés aux rôles qu'ils ont tenus ce soir-là, rôles auxquels est rattachée une pièce, un auteur et un genre.

L'enjeu pour nous, aujourd'hui, est de créer une troisième base de données, la base de données des dépenses de la troupe de 1680 à 1776, complémentaire de celle des recettes et des feux. Parallèlement à ce travail, nous créons un dépôt RDF de l'ensemble de nos jeux de données, pour les rendre ouvertes et permettre une liberté de représentation que n'autorisent pas les bases de données relationnelles. [La société Logilab, en tant que prestataire informatique, a en charge la réalisation technique de ces deux chantiers.] Car là est le défi pour nous : rendre compte de manière la plus fidèle possible de données partiellement irrégulières. Pour cela, nous présenterons, dans cet article, les registres eux-mêmes, pour que l'on puisse se rendre compte de ces irrégularités. Nous rendrons compte ensuite de l'application de transcription des dépenses, avant d'aborder, dans un

1. RCF https://cfregisters.org

dernier temps, la création de l'entrepôt SPARQL contenant les données fusionnées des trois bases.

2 Les registres de la "Comédie-Française"

L'enjeu de l'établissement de ces bases de données est multiple, pour la communauté des chercheurs. D'abord, il s'agit de donner accès à un large public (étudiants, amateurs, simples curieux et, bien sûr, universitaires et spécialistes) à des données qui ne sont pas aisément consultables par tous, parce qu'elles sont consignées dans les fonds de la Bibliothèque-Musée de la Comédie-Française. Les facsimilés des registres numérisés, auxquels notre site donne accès, est le premier maillon de cette entreprise de diffusion. Mais surtout, les bases de données donnent accès à un surcroît d'information, auquel on ne peut pas avoir accès, ou très difficilement, en se contentant de consulter les registres ou leur version numérisée : grâce à la reconfiguration des données permise par les bases de données, une lecture transversale est possible, qui permet de dégager des constantes et de faire des sondages sur la longue durée. On peut ainsi établir des statistiques sur les pièces qui rapportent le plus en fonction des périodes, sur la variation du nombre de spectateurs en fonction des catégories de places, sur la spécialisation des acteurs par types de rôles, etc. Dans le cas particulier de la base de données des dépenses que nous sommes en train de construire, nous espérons ainsi pouvoir savoir, à terme, quelles sont les pièces dont la représentation a coûté le plus cher en termes de décors, de costumes et de personnel supplémentaire, comme les musiciens et les figurants (les "assistants", comme on disait à l'époque). Grâce à cette base, nous pourrons mieux comprendre l'économie du théâtre, le statut des artistes sous l'Ancien Régime et obtenir de précieuses informations sur les prémisses de l'histoire de la mise en scène.

3 L'application de transcription des registres des dépenses

La politique de RCF en matière de traitement de l'information est de restituer la source de manière la plus précise possible, d'être au plus près de l'information d'origine telle qu'elle apparaît dans les registres. Cette exigence scientifique repose sur la volonté de livrer à l'utilisateur ou l'utilisatrice une information qui ait fait le moins possible l'objet de choix et d'interprétations en amont, ce qui biaiserait la lecture qu'il fait des registres. Une telle exigence est a priori aisément conciliable avec un traitement systématique de l'information, car, de manière générale, les registres journaliers présentent des informations qui sont elles-mêmes systématiques, parce qu'elles sont toujours subordonnées à une date : la "soirée", où a lieu telle représentation, qui a généré telles recettes et telles dépenses. Dans le cadre des registres des dépenses, en plus de cette régularité calendaire, on remarque qu'un certain nombre de catégories restent stables pendant plusieurs années, voire plusieurs décennies, ce qui nous permet de les envisager comme "données massives". Pourtant, l'information est à plusieurs égards hétérogène et non continue. Un premier aspect de cette disparité réside dans le fait que beaucoup de types de frais ne sont pas stables à l'échelle des 96 saisons de la période 1680-1776, mais le sont sur des plus courtes périodes (décennies, années, voire mois). C'est pourquoi les registres peuvent être regroupés par périodes durant lesquelles les catégories pré-imprimées sont globalement stables. La tâche principale, lors du développement de l'application de transcription pour les registres des dépenses, a été de définir les différents formulaires pour chaque période. Chaque dépense est associée à une valeur monétaire représentée en **L** (livres), **S** (sols) and **D** (deniers).

Nous avons défini trois types de champs de dépenses :

- **Champ simple** qui associe une dépense à une seule valeur monétaire : par exemple dans la figure 1 "frais ordinaires".
- **Champ multiple** qui associe une dépense à plusieurs valeurs monétaires, chacune définie par un intitulé manuscrit : par exemple dans la figure 1 "Frais extraordinaires".
- **Champ partitionné** qui associe une dépense à une seule valeur monétaire, qui est la valeur de la part, mais qui l'associe en plus à un nombre de parts et à un total. Ce type de champ est utilisé, par exemple, pour définir le paiement des acteurs, qu'il est possible de retrouver sur la figure 1 avec le libellé "PART".

Un second aspect de la disparité de l'information contenue dans les registres des dépenses tient au fait qu'ils regroupent plusieurs types de données. Nous distinguons quatre catégories différentes de dépenses :

— Les **catégories pré-imprimées** : par exemple, sur la figure 1 "frais ordinaires".
— Les **mentions manuscrites sur les catégories pré-imprimées** (qui permettent de les clarifier ou même de les remplacer) : par exemple sur la figure 1 "et chandelle des religieux".
— Les **informations marginales** (toujours manuscrites, qui ne concernent pas les dépenses, mais donnent des précisions à propos d'une "relâche" [2] ou indiquent la présence d'une personnalité importante dans la salle). Par exemple, sur la figure 1, nous pouvons lire "L'on ne joua point hier à cause du départ de Messieurs Poisson et Raisin pour Fontainebleau".
— Les **dépenses manuscrites** qui sont ajoutées aux catégories pré-imprimées. Par exemple sur la figure 1, "pour l'affiche noire".

L'interface de l'application de transcription a été adaptée pour pouvoir absorber toute cette diversité. Le transcripteur (ou la transcriptrice), peut enregistrer ces différentes annotations : le formulaire contient les dépenses pré-imprimées et certains champs sont laissés libres pour permettre de spécifier la mention manuscrite. Ces champs libres sont sys-

2. "Relâche" désigne les jours de fermeture du théâtre.

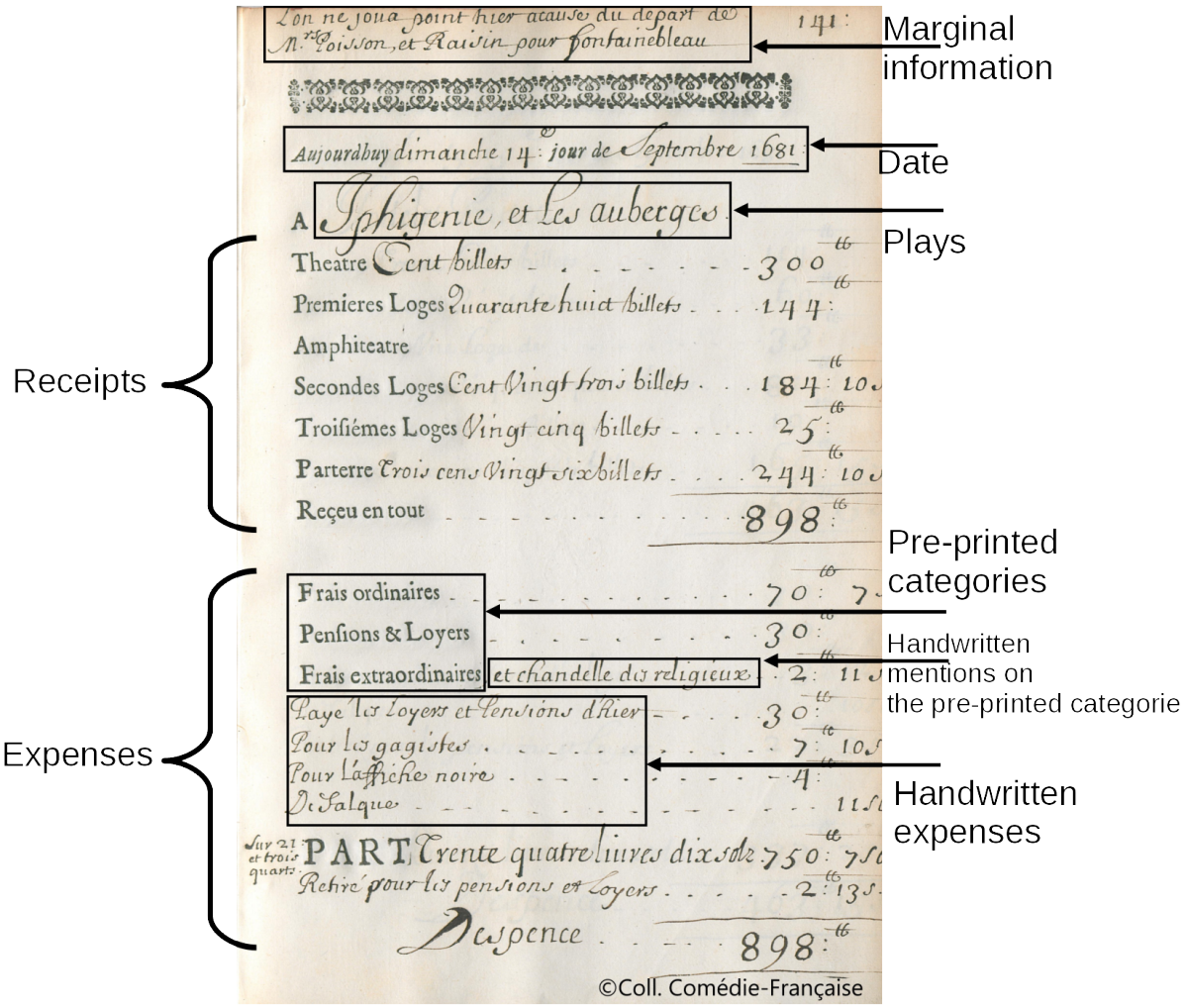

FIGURE 1 – Exemple d'une page d'un registre des dépenses

tématiquement rassemblés en "super-catégories" pour regrouper des frais analogues quand cela est possible. Par exemple, la catégorie "Frais extraordinaires" est utilisée pour réunir les dépenses exceptionnelles qui concernent ce jour spécifiquement. Par exemple dans la figure 1, "pour l'affiche noire" ou dans d'autres pages "vin" ou encore "une robe de chambre".

Néanmoins, ces frais extraordinaires posent problème car cette catégorie de dépenses définit des dépenses irrégulières, contrairement aux dépenses ordinaires : aussi n'ont-ils pas de représentation fixe durant toute la période. Pour certaine saison, il existe la catégorie pré-imprimée "Frais extraordinaires", comme nous pouvons le voir sur la figure 1. Mais dans d'autres registres, on note la présence de plusieurs mentions manuscrites qui ne sont pas catégorisées. Dans cette situation, il est périlleux de définir une catégorie pour ces dépenses : sont-elles réellement des "frais extraordinaires" ou sont-elles des dépenses non catégorisables ? Le problème ici est double : comment transcrire ces mentions manuscrites non catégorisées et peut-on trouver un moyen efficace et pertinent, durant un traitement post-transcription, de catégoriser quand même ces dépenses ? Cette étape de catégorisation post-transcription est importante puisqu'elle oriente les possibilités d'analyses quantitatives des données et leur interprétation. En fonction des catégories que nous définirons dans cette étape, les résultats peuvent varier significativement, ce qui n'est pas sans poser un problème scientifique majeur : en ajoutant une information qui ne figure pas dans l'archive, nous risquons d'introduire un biais interprétatif dans les données transcrites, qui ne rendent pas compte fidèlement du document source. Pour l'instant, nous ajoutons les mentions manuscrites non catégorisées comme des "Frais extraordinaires", sans catégorisation supplémentaire. C'est une des limites de notre approche : nous y reviendrons dans les perspectives présentées à la fin de cet article.

Un problème supplémentaire se pose pour certaines catégories de dépenses qui sont assimilées les unes aux autres sur une même ligne de frais. Dans la figure 1, les "frais extraordinaires" et les "chandelles des religieux" sont représentées comme une même dépense, puisqu'il y a un seul nombre pour ces deux postes. Il n'y a donc aucun moyen de déterminer si la dépense "2 livres et 11 sols" correspond aux "frais extraordinaires" dans leur intégralité ou juste aux "chandelles des religieux".

La figure 2 présente l'exemple d'un formulaire de transcription d'une page d'un registre. La partie gauche de l'application est dédiée au fac-similé de la page de registre qui doit être transcrite. La partie droite contient les champs des dépenses à renseigner, en fonction de la période à laquelle la page fait référence. En bas de l'application, apparaissent en commentaires des méta-données concernant le processus de transcription. Est renseigné aussi dans ces méta-données l'état de la transcription pour cette page.

Le processus de transcription d'une page d'un registre suit différentes étapes. Tout d'abord, la page se trouve dans l'état *to do* (à faire). Puis, lorsque le transcripteur ou la transcriptrice a fini de transcrire la page, elle passe dans l'état *to review* (à relire) pour informer un expert qu'il faut vérifier la page, pour la valider ou la corriger. Une fois que les données de transcription sont validées, la page passe dans l'état *validated* (validée). À ce moment, les données peuvent être analysées puisqu'elles ont été transcrites et validées. Une page peut être validée même si des erreurs apparaissent. Par exemple, l'erreur la plus courante réside dans une différence entre le total transcrit des dépenses sur la page d'archive et le total calculé par la machine, à partir des données rentrées dans le formulaire de saisie. La plupart du temps, cette erreur est présente dans la page elle-même et n'est pas liée à une mauvaise transcription. Nous avons décidé de ne pas corriger l'erreur dans les données, en notifiant que la page contient une erreur. Cela permet de laisser aux chercheurs la possibilité d'étudier ces erreurs et de les interpréter, par exemple pour déterminer à partir des récurrences si elles sont délibérées ou non, et pour analyser s'il y a une forme de détournement d'argent de la part du comédien chargé de remplir quotidiennement le registre.

4 Des données RDF pour fusionner les trois bases de données

Une fois que les trois bases de données (*dépenses*, *recette* et *feux* (rôles et acteurs/actrices)) sont transcrites et validées, il faut ensuite pouvoir les fusionner, afin de pouvoir interroger l'intégralité des données. Le processus d'alignement a été relativement simple puisque chaque page transcrite, dans les trois bases, est associée à une date de représentation, qui constitue l'élément pivot de notre alignement. L'utilisation des formalismes RDF et des technologies du Web Sémantique nous a semblé le plus appropriée, dans la mesure où l'objectif principal est de donner accès à ces données au plus grand nombre de personnes : les technologies du Web Sémantique et plus spécifiquement le Web de données liées sont destinés précisément à cet usage [5]. Mais le formalisme RDF permet aussi de résoudre un des problèmes présentés dans le chapitre 2. La diversité des données est complexe à manipuler dans une base de données relationnelle, alors que le RDF permet de créer facilement des liens lorsque cela est nécessaire. Nous n'avons pas besoin de connaitre au préalable l'intégralité des attributs et des champs.

Afin d'exporter ces données au format RDF, il nous faut définir une ontologie. Nous pouvons observer une partie de cette ontologie, unifiant les concepts des trois bases, dans la figure 3 : pour faciliter la lisibilité, nous n'en affichons qu'une partie (par exemple, la notion de "période" n'apparaît pas).

Nous l'avons dit : le concept pivot pour créer les liens entre les différentes données des trois bases est la **Journée**. Toute la partie supérieure sur la figure 3 regroupe les concepts de la base des *Feux* (acteurs/actrices et rôles), en bas à gauche pour les *Dépenses* (Expenses) et en bas à droite pour les *Recettes*(Receipts). Tous les détails (catégories de dépenses, catégories de sièges, etc.) ne sont pas présents dans l'ontologie initiale puisque ces détails sont générés dynamiquement lors de l'export des données. L'ontologie évolue donc

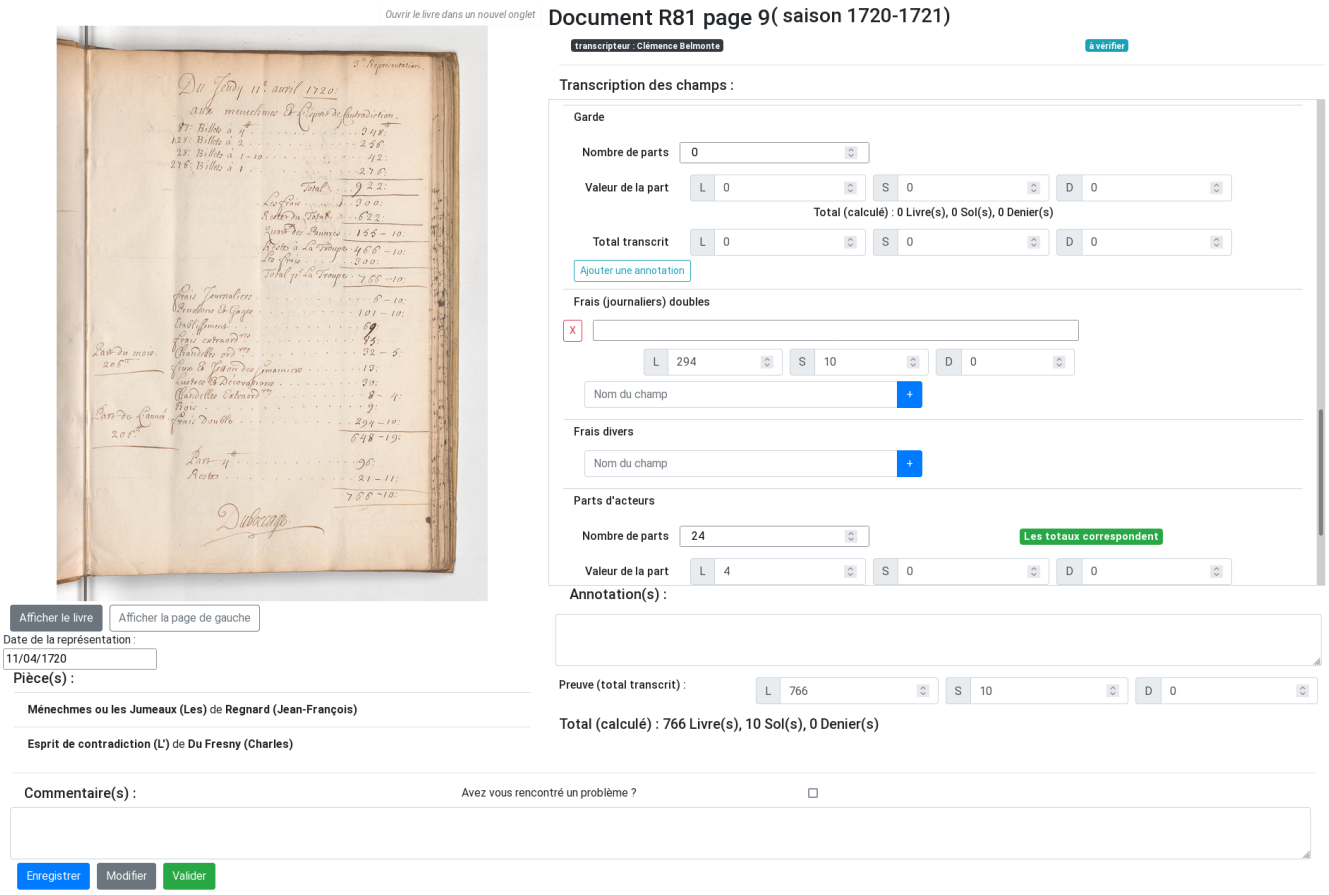

FIGURE 2 – Application de transcription des dépenses

en fonction des éléments découverts lors de l'export.

Une fois l'ontologie définie, nous alignons les concepts clés de cette ontologie avec des ontologies de références. Cet alignement est crucial, notamment parce que notre ontologie utilise, par fidélité à la source archivistique, des termes français. Mais il n'y a aucune raison de ne pas permettre à des personnes non francophones d'accéder à ces données : favoriser la réutilisation et l'appropriation des données par le plus grand nombre est un des objectifs du programme RCF, et cela est facilité lorsque les ontologies sont alignées sur une même ontologie de référence. Ces alignements peuvent ensuite être utilisés pour enrichir les données, par exemple en y adjoignant la date de naissance d'un acteur ou d'une actrice.

La table 1 présente les alignements entre l'ontologie RCF et FRBRoo, schema.org, DubinCore et FoaF. Nous n'avons toutefois pas aligné l'intégralité de FRBRoo, bien que cette dernière constitue une des ontologies les plus reconnues dans ce domaine. Parce qu'elle hérite de Cidoc CRM, la notion d'évènement est centrale dans sa modélisation. Or, dans nos données, il y a bien des évènements, mais tout n'est pas évènement : par exemple, nous n'avons pas "l'évènement" de la naissance d'une personne, ce qui est nécessaire dans l'ontologie FRBRoo. De plus, nous voulons éviter des alignements complexes qui ne rendraient pas la réutilisation du modèle aisée. Une fois que nous avons défini l'ontologie, exporté toutes les données à partir des trois bases de données en RDF, nous déployons l'intégralité des triplets RDF (ontologie et données) dans un entrepôt SPARQL public [3]. Nous utilisons Virtuoso Open-Source Edition [4]. Il est alors possible d'interroger cet entrepôt pour obtenir des réponses aux questions telles que "quel est le prix des représentations dans lesquelles un acteur donné a joué et quels sont les frais extraordinaires pour ces représentations ?" Une autre requête possible consiste à comparer les recettes de ces représentations et à les mettre en lien avec la distribution des rôles : ce peut être un moyen d'interroger le processus de vedettarisation des acteurs et actrices.

Pour aider les utilisateurs et utilisatrices ne maitrisant pas le SPARQL à accéder à ces données, nous avons développé une application web`https://upnd.pages.logilab.fr/rcf-ui-expense/` qui utilise les données de l'entrepôt SPARQL. Cette application est pour l'instant dédiée aux dépenses, mais il serait intéressant de l'étendre pour prendre en compte les données des deux autres bases.

La figure 4 présente une capture d'écran de l'application. Nous pouvons observer qu'il est possible de comparer plusieurs types de dépenses dans une période donnée. C'est particulièrement utile pour déterminer quelle dépense représente une grande part des dépenses totales en fonction des autres types de dépenses et d'une période donnée. Cette application permet aux utilisateurs ou utilisatrices de com-

3. `https://rcf-sparql.demo.logilab.fr/`
4. `https://vos.openlinksw.com`

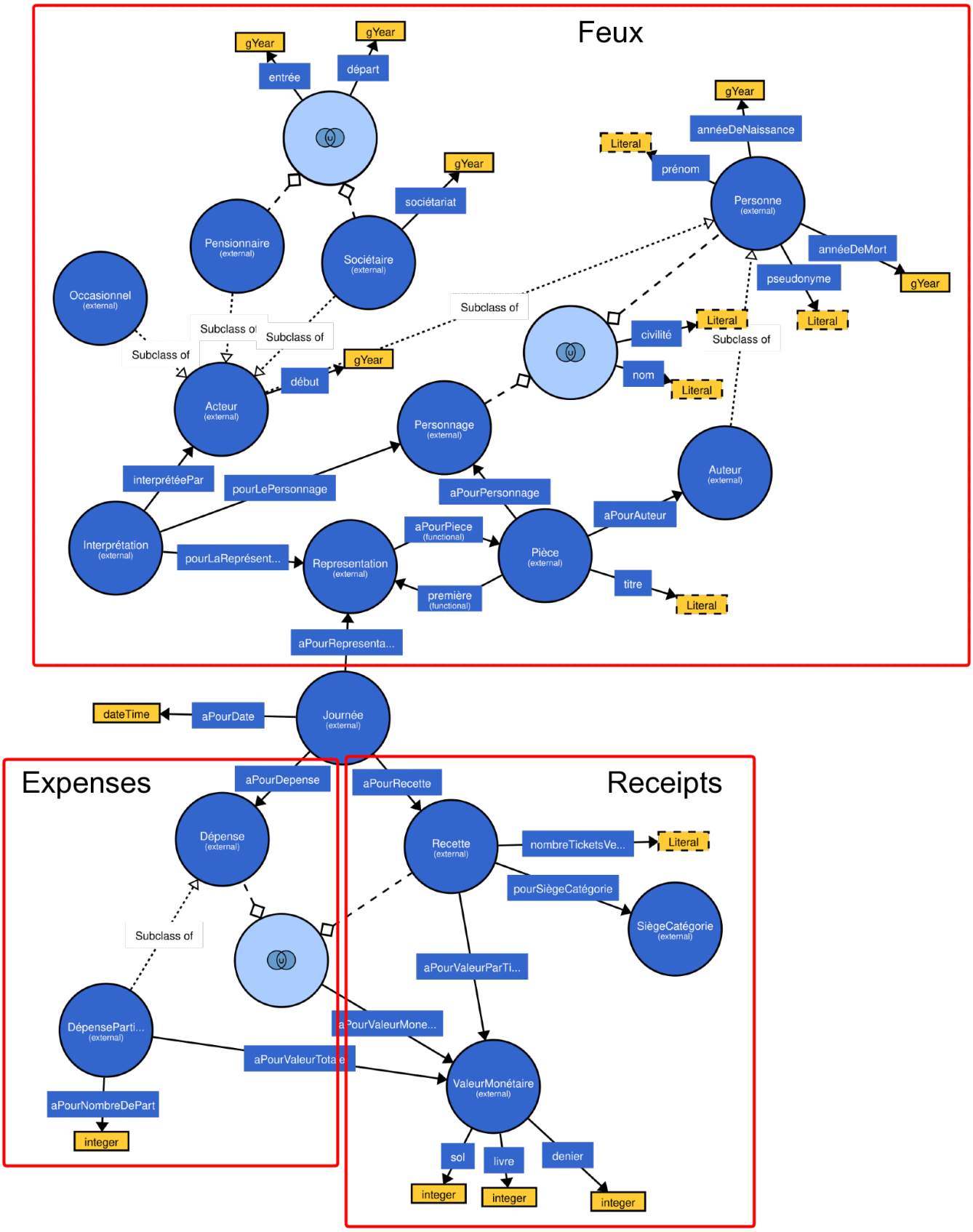

FIGURE 3 – Ontologie unifiée RCF

RCF	**FRBRoo**[2]	**schema.org**	**DublinCore**[6]	**FoaF**[3]
		Class		
rcf :Personne	frbroo :E21_Person	schema :Person		foaf :Person
rcf :Journée	frbroo :E5_Event	schema :Event	dctype :Event	
rcf :Pièce	frbroo :F1_Work	schema :Play		
rcf :Inteprétation	frbroo :F31_Performance			
rcf :Représentation	frbroo :F2_Expression	schema :TheaterEvent		
rcf :ValeurMonétaire		schema :MonetaryAmount		
		Object Properties		
rcf :aPourAuteur		schema :author	dce :creator	
rcf :aPourPersonnage		schema :character		
rcf :aPourPiece	frbroo :R40i_is_representative_expression_for.	schema :workPerformed		
rcf :aPourReprésentation		schema :subEvent		
rcf :première		schema :firstPerformance		
		Data Properties		
rcf :annéeDeMort		schema :deathDate		
rcf :annéeDeNaissance		schema :birthDate		
rcf :civilité		schema :gender		
rcf :nom		schema :familyName		foaf :surname
rcf :prénom		schema :givenName		foaf :givenname
rcf :pseydonyme				foaf :nick

TABLE 1 – Alignement de l'ontologie RCF avec FRBRoo, schema.org, DublinCore et FoaF

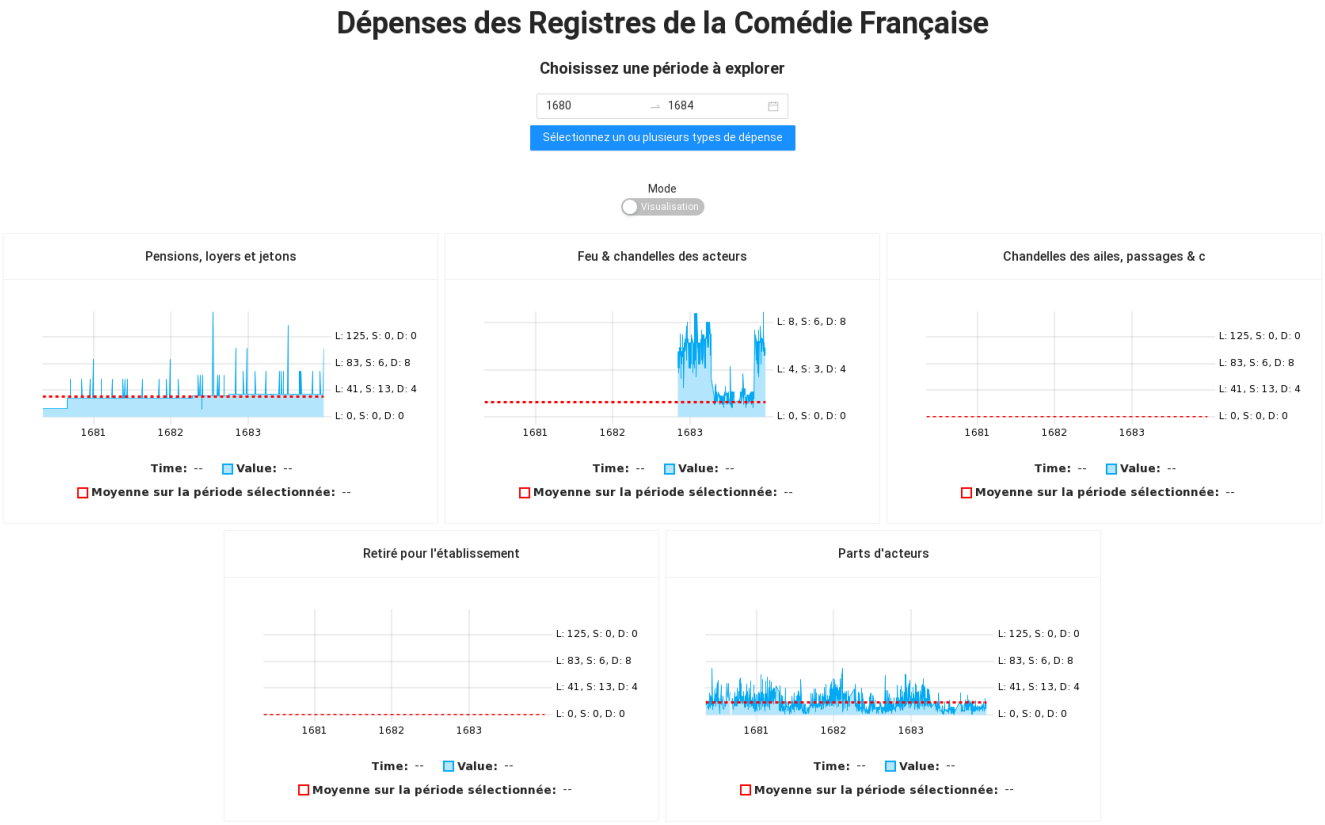

FIGURE 4 – Application Web pour visualiser les données des dépenses à partir de l'entrepôt SPARQL

parer la somme ou la moyenne des dépenses sélectionnées pour certaines pièces ou auteurs ou autrices (toujours selon une période donnée).
Le développement de cette application était nécessaire pour permettre à tout le monde d'accéder à ces données. Le problème, néanmoins, est que lors des développements, il a été nécessaire de faire des choix. Nous avons dû arbitrer par exemple sur la façon dont les dépenses sont affichées et dont elles sont filtrées ou agrégées. Dans le cas des dépenses partitionnées, par exemple, nous les considérons comme des champs simples et nous comparons uniquement le total des parts. Si des utilisateurs ou utilisatrices souhaitent explorer ce type de dépenses et étudier finement leur évolution, l'application sera limitée. C'est pourquoi l'entrepôt SPARQL est si important, d'un point de vue technique et scientifique : il permet d'accéder aux données RDF directement, telles qu'elles ont été saisies. Mais il est nécessaire de savoir écrire des requêtes SPARQL, c'est pourquoi nous prenons en considération l'utilisation d'outils comme Sparnatural [5].

5 Conclusion et perspectives

Les Registres des dépenses de la Comédie-Française sont une source d'informations particulièrement riche pour les recherches en humanités. Leur transcription apporte une vraie plus-value, puisqu'elle permet une analyse quantitative des données. Mais l'information y est disparate, à cause des irrégularités dans les types de frais. Nous avons montré que notre application de transcription des dépenses est conçue pour prendre en compte une partie de ces irrégularités, puis nous avons expliqué comment nous avons fusionné les trois bases de données dans un entrepôt SPARQL dédié, pour permettre au plus grand nombre de personnes d'accéder à ces données.

La disparité des données observée dans les registres RCF nous oblige à avoir une attitude critique à l'égard de l'approche quantitative. Il nous a semblé important de préciser qu'il existe un biais d'interprétation dans l'application de visualisation. Observer un graphique sur une page web peut donner une impression d'objectivité quant aux données présentées. Mais c'est un leurre : l'aspect quantitatif, qui est seulement une possibilité de lire et de comprendre les informations, résulte d'un processus fait de choix successifs, qui filtrent nécessairement l'information, qui l'orientent et qui lissent les disparités. Mais ces choix sont aussi nécessaires pour rendre l'information lisible et intelligible. C'est la raison pour laquelle les informations quantitatives, telles qu'elles sont présentées dans l'application de visualisation des dépenses, nécessitent d'être interprétées par les chercheurs ou chercheuses. Contextualiser des données devient alors essentiel. Nous avons commencé la création d'une encyclopédie RCF, qui définit les termes utilisés dans les données, telles que "Frais extraordinaires", "Frais ordinaires", "saisons", "loges", "preciput", etc. Le qualitatif vient en quelque sorte au secours du quantitatif, en apportant un

5. https://sparnatural.eu/

éclairage contextuel indispensable [6].
Comme discuté dans le chapitre 3, certaines dépenses manuscrites ne sont pas catégorisées. C'est un problème, car de ce fait, nous créons une nouvelle catégorie de dépenses qui vient spécialiser "Frais extraordinaires" pour chaque mention manuscrite non catégorisée. À cause de la nature libre de la mention manuscrite, chaque label identifiant ces catégories sera donc différent, sauf pour quelques exceptions. Le résultat est un grand nombre de catégories ne contenant qu'une seule dépense. Si les données ne sont pas, ou peu, catégorisés, alors l'analyse quantitative perd de son intérêt. C'est pour cela que nous souhaitons explorer les possibilités d'utiliser des algorithmes de catégorisation sur les sous-catégories de "Frais extraordinaires". Il s'agirait d'utiliser l'encyclopédie RCF pour détecter des entités nommées dans les labels des catégories manuscrites et de les regrouper, si elles partagent ces mêmes entités.
Mais, au-delà de la nécessité de contextualiser les données, le problème reste l'accès aux données brutes, afin qu'aux chercheurs et chercheuses puissent être délivrées des informations qui soient le moins biaisées possible. Puisque SPARQL n'est pas si simple à appréhender, il sera nécessaire de former des non spécialistes à ce langage de requête en proposant un moyen pédagogique de l'apprendre : la connaissance, même rudimentaire, de SPARQL peut permettre d'éviter les biais, inévitables, des interfaces de visualisations. Comment rapprocher les chercheurs et chercheuses en humanités de l'accès aux données brutes des registres, qui constituent un si riche fonds ? Comment mettre à profit l'ontologie OWL pour proposer un outil de visualisation qui soit le moins partial et le moins biaisé possible ? Ces questions seront au centre des travaux de ces prochaines années.

Références

[1] Burrows, S., Roe, G. : Digitizing Enlightenment : Digital Humanities and the Transformation of Eighteenth-century Studies (2020)

[2] Doerr, M., Bekiari, C., LeBoeuf, P., nationale de France, B. : Frbroo, a conceptual model for performing arts. In : 2008 Annual Conference of CIDOC, Athens. pp. 15–18 (2008)

[3] Graves, M., Constabaris, A., Brickley, D. : Foaf : Connecting people on the semantic web. Cataloging & classification quarterly **43**(3-4), 191–202 (2007)

[4] Harvey, S., Sanjuan, A. : Le projet des registres journaliers de la comédie-française : Les humanités numériques, dialogue entre les mondes de la recherche et de la documentation. Bulletin des bibliothèques de France (9), 102–109 (2016)

[5] Meroño-Peñuela, A., Ashkpour, A., Van Erp, M., Mandemakers, K., Breure, L., Scharnhorst, A., Schlobach, S., Van Harmelen, F. : Semantic technologies for historical research : A survey. Semantic Web **6**(6), 539–564 (2015)

[6] Weibel, S.L., Koch, T. : The dublin core metadata initiative. D-lib magazine **6**(12), 1082–9873 (2000)

6. https://cfregisters.org/#!/encyclopedie/mots