

Formaliser pour mieux partager : bilan du travail de modélisation pour la BDD du répertoire de la Comédie Italienne de Paris (XVIII siècle)

Soumission anonyme

Résumé

La création d'une base de données relationnelle pour le répertoire de la Comédie Italienne de Paris au XVIII siècle, s'insère dans un projet doctoral, débuté en septembre 2022 pour une durée de trois ans. L'objectif est d'étudier l'évolution du répertoire de manière chronologique - avec la génération de listes, graphes et autres visualisations - à partir des répertoires existants. Après avoir commencé la saisie des données, nous pouvons dresser un bilan du travail effectué lors de la formalisation des données, par rapport aux objectifs du projet ainsi qu'à des préoccupations écologiques, selon la perspective du *minimal computing*. L'article présente les bénéfices et les problèmes que nous avons rencontré dans le processus de définition du modèle entité-association pour la base de données.

1 Introduction

La création d'une base de données (BDD) pour le répertoire de la Comédie Italienne de Paris au XVIII siècle, s'insère dans un projet doctoral, débuté en septembre 2022 pour une durée de trois ans¹. L'objectif est d'étudier l'évolution chronologique du répertoire - avec la génération de listes, graphes et autres visualisations - à partir des répertoires existants (Brenner, 1961 ; Brenner, 1947 ; De Luca, 2011 ; Wild and Charlton, 2005 et Spanu Fremder, 2010).

Après avoir commencé la saisie des données, nous pouvons dresser un bilan du travail effectué lors de la formalisation des informations, par rapport aux objectifs du projet ainsi qu'à nos préoccupations écologiques², selon la perspective du *minimal computing* (Radio, 2020, Gil, 2015, et plus largement sur la décroissance Jarrige, 2022,

1. Le projet, financé par un contrat doctoral de l'Initiative Théâtre, est mené à Sorbonne Université par l'auteur/autrice de cette contribution, sous la direction d'Andrea Fabiano, Emanuele De Luca et Glenn Roe.

2. Sur le numérique et l'écologie : ARCEP (2022), Roussilhe and Berthoud (2021)

Roussilhe and Mateus, 2023). En accord avec cette approche³, notre priorité n'est pas d'utiliser les technologies les plus innovantes, mais de rendre le processus compréhensible (donc appropriable et reproductible), par d'autres chercheurs. Nous utilisons donc une BDD relationnelle, décrite avec un modèle entité-association (EA).

Au prisme de nos objectifs, nous allons présenter les bénéfices et les problèmes que nous avons rencontré dans le processus de définition de modélisation EA, en s'insérant dans deux volets : le fonctionnement du numérique, et son utilisation en histoire du théâtre.

2 Pourquoi cette BDD ?

Parmi les répertoires que nous avons à disposition et qui constituent notre point de départ, les deux documents principaux, dont nous tirons la plupart des données, sont le répertoire de Brenner de 1961 et celui de De Luca (De Luca, 2011). Le répertoire Brenner, 1961 reporte les titres des pièces, associées aux dates auxquelles elles ont été jouées, et aux dépenses et recettes⁴. Celui de De Luca, de son côté, fournit des informations plus détaillées concernant les pièces (auteur, genre, nombre d'actes, forme en prose ou en vers du texte, éventuelles éditions et sources, ainsi qu'une note avec des informations tirées principalement de la presse de l'époque), mais seulement la date de la première représentation.

Le répertoire de De Luca contient les informa-

3. Voir par exemple, Risam and Gil (2022), Chan and Sayers (2022), Wythoff (2022). Le numéro complet dédié au *minimal computing* en HN est consultable à [ce lien](#).

4. Nous ne traitons pas du tout les dépenses et recettes dans notre BDD pour plusieurs raisons : cette information n'est pas homogène tout au long du siècle (parfois on a dépenses et recettes, mais parfois seulement une des deux, voire aucune) ; cette information n'est que partiellement représentative de l'affluence du public au théâtre, car on ne connaît jamais le nombre de spectateurs des loges, qui étaient en abonnement ; de plus, cette information est déjà traitée dans la future base de données du projet [Recital](#).

tions sur environ 650 pièces et couvre la période 1716-1762 (dates de l'arrivée à Paris de, respectivement, Luigi Riccoboni et Carlo Goldoni). Les informations sur les pièces qui ne sont pas traitées, seront tirées des autres répertoires, en particulier de [Brenner, 1947](#). [Brenner, 1961](#) couvre en revanche le siècle en entier, et contient moins que 41000 représentations selon une estimation approximative⁵.

La future BDD est conçue pour fusionner les informations contenues dans ces deux répertoires, avec le but de contribuer à l'étude de l'articulation entre tradition italienne et française au sein du théâtre de la Comédie-Italienne dans la période étudiée⁶. Pour cela, nous voulons explorer les évolutions du répertoire du théâtre au cours du siècle : premières et reprises, distribution des genres des pièces, présence et absence d'auteurs, par exemple. Par *répertoire* on entend ici l'ensemble, cohérent par rapport au genre, des pièces qui étaient jouées dans un certain théâtre - ensemble qui définissait à son tour le théâtre ([de Rougemont, 2001](#)).

Cette volonté d'étudier le répertoire du théâtre Italien, au moyen des répertoires⁷ existants, a des conséquences sur la modélisation des données. Par exemple, nous ne distinguons pas d'éventuelles versions différentes d'une même pièce, et nous ne distinguons pas non plus une pièce comme texte mis en scène de son édition (lorsqu'elle existe) : la pièce nous intéresse comme l'objet abstrait qui sous-tend à toutes ses matérialisations, car nous étudions le théâtre du point de vue des spectateurs. De plus, cela permet de retracer de manière simple les variations au sein du répertoire.

En revanche, suivant le modèle utilisé pour les [Registres de la Comédie-Française](#)⁸, nous avons distingué la soirée comme moment spatio-temporel où le théâtre est ouvert au public, de la représentation comme mise en scène d'une certaine pièce lors d'une certaine soirée. Nous avons donc une configuration où il peut y avoir plusieurs représentations lors d'une même soirée, et où à chaque représen-

5. Cette estimation a été faite en calculant la moyenne du nombre de pièces mentionnées dans onze pages, prises aléatoirement, et en multipliant ce nombre par le nombre de pages du répertoire.

6. Concernant la Comédie-Italienne à cette époque voir par exemple [Fabiano, 2018](#), [De Luca, 2015](#), [Fabiano and De Luca, 2023](#)

7. On utilise le mot répertoire avec deux significations distinctes : comme l'ensemble des pièces jouées par une compagnie à un moment donné, et comme un recueil d'informations synthétiques.

8. Ce modèle m'a été fourni par Motasem Alrahabi au début de ma thèse.

Une entité dans un diagramme entité-relation

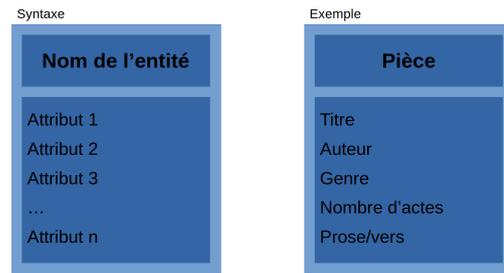


FIGURE 1 – Exemple d'entité

Entités et relations dans un diagramme entité-relation

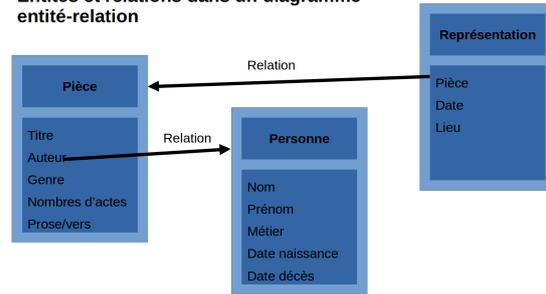


FIGURE 2 – Exemple de relation.

tation n'est associée qu'une seule pièce. Chaque pièce peut à son tour être représentée plusieurs fois.

3 Bilan technique de la formalisation

Le modèle EA prévoit la définition d'objets (entités) et des relations (associations) qui les caractérisent. De nombreuses notations existent pour dessiner un diagramme EA, la plus commune étant probablement celle de Chen. Afin de nous aligner avec des projets proches dont nous avons la documentation à disposition (les [Registres de la Comédie-Française](#), et surtout la future BDD Opéra-Comique Database, grâce à Motasem Alrahabi) nous avons utilisé une notation mixte, où les entités sont des tableaux avec la liste détaillée des attributs (Figure 1), et où les relations sont des simples flèches sans label - comme dans une table de relation (Figure 2). Il s'agit d'une notation proche du modèle logique pour une BDD relationnelle.

Le schéma est relativement simple à concevoir dans ses grandes lignes, mais nécessite la manipulation de certains concepts mathématiques pour être bien défini : cela concerne particulièrement la notion de relation, de cardinalité, et la cohérence du modèle. Concernant les relations, l'utilisation de simples flèches a vite trouvé ses limites, car nous ne pouvons pas dessiner toutes les relations néces-

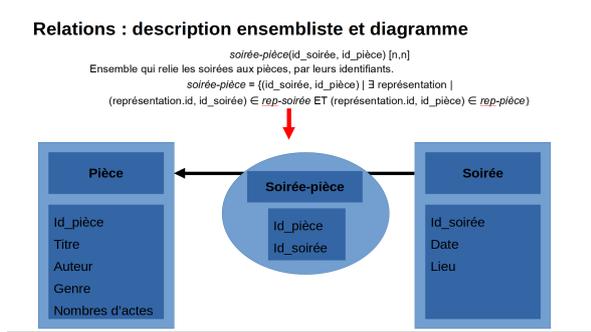


FIGURE 3 – Différentes manières d’écrire une relation

saïres pour caractériser notre répertoire ⁹. Pour en-
 diguer le problème, nous avons décrit les relations
 manquantes au moyen du formalisme de la théorie
 des ensembles (Figure 3), car les relations, sont
 en fait des types particuliers d’entités, que nous
 pouvons définir par des ensembles. Cela permet
 d’intégrer rapidement les relations à un nouveau
 diagramme, mais ne facilite pas la communication
 avec les spécialistes en humanités.

La cardinalité est une notion indiquant le nombre
 de fois où les données peuvent être mises en rela-
 tion. Cette information doit être définie pour décrire
 la nature de la relation, mais aussi pour guider la
 recherche des informations (il est important de sa-
 voir, par exemple, qu’à partir d’une représentation
 nous pouvons trouver une unique soirée, alors qu’à
 partir d’une soirée, nous obtenons plusieurs repré-
 sentations ¹⁰). La complexité de l’attribution d’une
 cardinalité à une relation, réside dans la transfor-
 mation d’une phrase assez complexe en langage
 naturel, en un couple de symboles ¹¹.

Vérifier la cohérence du modèle, consiste à éli-
 miner les informations rédundantes et à optimi-
 ser la structure pour la recherche d’informations.
 Pour cela, on utilise des processus de transfor-
 mations successives de la structure, afin de la rendre
 plus efficace (processus de normalisation). Les mé-
 thodes de normalisation sont peu intuitives pour
 des spécialistes en humanités, et une intervention
 informatique, attentive à l’interprétation des don-
 nées, est nécessaire. Pour notre BDD, la normali-
 sation a mené à des changements importants dans
 la structure des données ¹² (Figures 4 et 5). L’enjeu

9. Avec des flèches il n’est possible de dessiner que les relations créées par la présence d’un identifiant dans l’attribut d’un entité (clé étrangère).

10. Dans ce cas, la cardinalité de la relation (représentation, soirée) est [n,1].

11. Les seules valeurs admises sont : [1,1], [1,n], [n,1], [n,n].

12. Ce travail a été mené grâce à l’intervention fondamen-

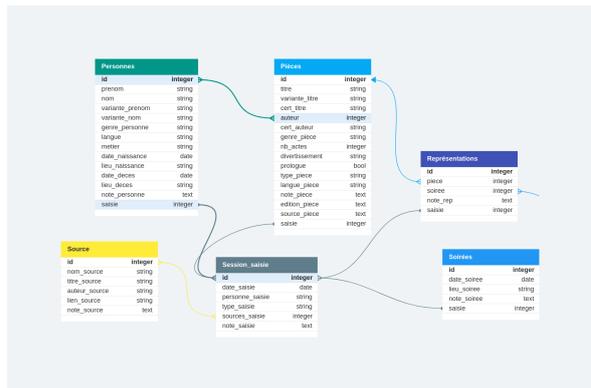


FIGURE 4 – Diagramme EA avant la normalisation.

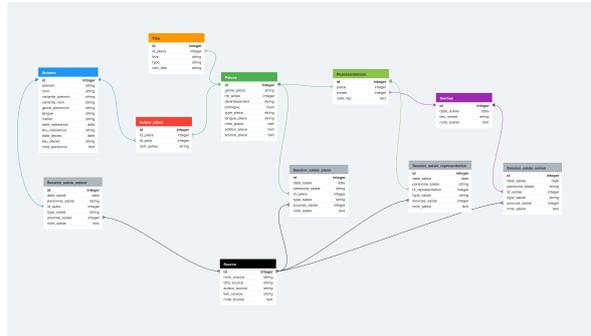


FIGURE 5 – Diagramme EA après la normalisation (image créée par Camilla Cantiani).

principal a été de modifier le diagramme sans gé-
 nérer des incohérences d’interprétation, mais grâce
 à une analyse fine des besoins et des contraintes, le
 résultat final les respecte d’avantage.

4 Bilan réflexif de la formalisation

La formalisation est le moment de transforma-
 tion d’une information discursive et multidimen-
 sionnelle, vers une information strictement codi-
 fiée et non ambiguë. Ce moment est conceptualisé
 dans les humanités numériques (HN) par la dis-
 tinction entre *information* et *donnée*. Il s’agit d’un
 travail interprétatif, qu’il est nécessaire d’expli-
 citer (Mouza et al., 2023, Moretti and Sobchuk, 2022,
 Bowker and Leigh Star, 2023, Bollen, 2016). Cela
 représente une étape intermédiaire vers la numé-
 risation, et constitue un moyen pour décrire les
 informations de manière relativement courte, et
 très précise, car chaque symbole a une signification
 particulière, le privant de toute ambiguïté. La dé-
 finition de concepts précis, a l’avantage de rendre
 visibles les biais méthodologiques et disciplinaires,
 et d’enrichir ainsi une communication transdisci-
 plinaire (Mouza et al., 2023, Bollen, 2016). Cela

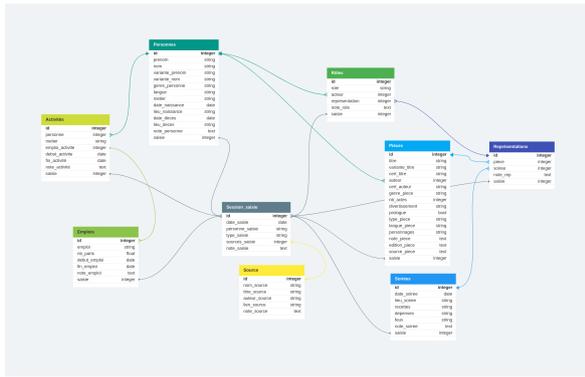


FIGURE 6 – Diagramme EA complet de la BDD (non normalisé).

réduit le caractère multidimensionnel des objets historiques, mais donne une profondeur particulière à l'interprétation que l'on choisit de développer.

Concernant les questions écologiques, le processus de formalisation a représenté un outil d'analyse des besoins et des contraintes du projet, facilitant une implémentation appropriée au contexte (étant indépendant des technologies). Cela constitue le point de départ pour une approche *minimal computing* aux HN, telle qu'elle a été définie par Risam et Gil dans l'introduction au numéro de Digital Humanities Quarterly dédié au *minimal computing*¹³. La modélisation a facilité aussi la simplification de la structure des données, non seulement lors de la normalisation, mais aussi pour sélectionner les informations essentielles, et que nous sommes en mesure de traiter avec nos moyens. Concrètement, nous avons défini une structure qui représente toute l'activité de la compagnie (Figure 6)¹⁴, mais nous n'implémentons que la partie essentielle pour étudier les pièces et la chronologie de leurs représentations (Figure 4)¹⁵. Cette double structure facilite, en outre, une éventuelle prise en main successive de la BDD, pour développer l'interconnexion et l'interopérabilité, ou pour enrichir les données.

5 Conclusions

La formalisation mathématique a été pour nous un outil d'analyse des besoins, et de communication interdisciplinaire, répondant à nos préoccupations écologiques. Néanmoins, afin de profiter pleinement de son potentiel communicatif, une in-

13. Puisque « not every cultural heritage or scholarly outfit around the world has access to the same infrastructure or resources [...] minimal computing encourages solutions that can be implemented universally. » (Risam and Gil, 2022)

14. Version interactive à ce lien [ce lien](#).

15. Version interactive à ce lien [ce lien](#).

tervention attentive d'informaticiens ou mathématiciens est nécessaire, car la prise en main d'un outil comme le modèle EA demande une certaine familiarité avec les langages formels, pour les utiliser de manière créative.

La BDD est actuellement dans la phase de saisie des données¹⁶, effectuée à travers une interface créée avec Django par Camilla Cantiani. Son implémentation finale pourra donc être générée dans plusieurs formats. Les *Registres de la Comédie-Française* utilisent PostgreSQL, mais, dans une perspective *minimal computing*, il serait plus pertinent dans notre cas d'utiliser SQLite. En effet, c'est un logiciel plus léger, et notre BDD est assez simple pour ne pas avoir besoin des fonctionnalités complexes que propose PostgreSQL. Mais surtout, SQLite ne doit pas nécessairement s'appuyer sur un serveur pour fonctionner et, à ce stade du projet, cela nous semble l'aspect le plus intéressant pour la perspective *minimal computing* : une BDD légère et sans serveur peut tourner sans accès à internet, elle est peu dépendante d'autres technologies, et peut être stockée voire installée sur une simple clé USB - donc fonctionner sur la plupart des ordinateurs.

Le *minimal computing* nous a ici servi comme un outil méthodologique pour guider et évaluer nos choix en termes d'écologie. Néanmoins les pratiques de recherche littéraire ne sont pas totalement adaptées à cette démarche universaliste, qui ne fournit que peu de pistes pour intégrer dans la modélisation l'acte interprétatif des chercheurs. L'article d'Erik Radio sur les ontologies (Radio, 2020), propose deux principes pour aborder cette question : « maximum materiality » et « maximum transparency ». Le premier sert à ancrer le modèle dans les aspects matériels et physiques des objets étudiés, le deuxième à expliciter l'écart entre « those elements or properties that are based in concrete reality and those that are interpretative »¹⁷.

Or l'interprétation apparaît ainsi comme un élément étranger à la modélisation, qui doit être le plus possible éliminé (« maximum materiality »), ou explicité lorsque cela n'est pas possible (« maximum transparency »). Mais est-ce que l'acte interprétatif est véritablement étranger à la modélisation ? Comment pourrait-on le valoriser à l'intérieur de nos modèles ? Ces questions, que ce court article ne permet pas d'aborder, ont un intérêt au delà des HN, pour comprendre le numérique et sa capacité

16. Cette tâche sera bientôt accompagnée par le travail de deux stagiaires.

17. Radio, 2020

271

272

273
274

275
276
277

278
279
280
281

282
283
284

285
286
287

288
289
290

291
292
293
294

295
296
297
298
299
300

301
302

303
304
305
306

307
308
309

310
311

312
313
314

315
316
317

318
319
320
321

322
323

324
325
326

à représenter le monde.

References

ARCEP. 2022. [Pour un numérique soutenable](#). Technical report, ARCEP.

Jonathan Bollen. 2016. [Data models for theatre research: People, places, and performance](#). *Theatre Journal*, 68(4) :615–632.

Geoffrey C. Bowker and Susan Leigh Star. 2023. *Arranger les choses : des conséquences de la classification*. Number 9 in EHESS translations. Éditions de l'EHESS.

Clarence Brenner. 1947. *A bibliographical list of plays in the French language 1700-1789*. Ann Arbor, Mich. : printed by Edwards brothers, Berkley.

Clarence Brenner. 1961. *The Theatre italien, its repertory, 1716-1793*. University of California press, Berkley-Los Angeles.

Tiffany Chan and Jentery Sayers. 2022. Minimal Computing from the Labor Perspective. *Digital Humanities Quarterly*, 016(2).

Emanuele De Luca. 2011. *Il repertorio della Comédie-Italienne di Parigi (1716-1762) / Le répertoire de la Comédie-Italienne de Paris (1716-1762)*. Les savoirs des acteurs italiens. IRPMF, Paris.

Emanuele De Luca. 2015. *"Un uomo di qualche talento" : François Antoine Valentin Riccoboni (1707-1772) vita, attività teatrale, poetica di un attore-autore nell'Europa dei Lumi*. Ph.D. thesis, Fabrizio Serra editore, Pisa. ISBN : 9788862278041 Series Number : 9 Series : Biblioteca di drammaturgia.

Martine de Rougemont. 2001. *La vie théâtrale en France au XVIIIe siècle*. H. Champion, Paris.

Andrea Fabiano. 2018. *La Comédie-Italienne de Paris et Carlo Goldoni : De la commedia dell'arte à l'opéra comique, une dramaturgie de l'hybridation au XVIIIe siècle*. PU Paris-Sorbonne, Paris.

Andrea Fabiano and Emanule De Luca, editors. 2023. *L'apothéose d'Arlequin*. Sorbonne Université Presses, Paris.

Alex Gil. 2015. [The User, the Learner and the Machines We Make](#). *Minimal Computing*.

François Jarrige. 2022. *On arrête parfois le progrès : histoire et décroissance*. Le pas de côté. l'Échappée, Paris.

Franco Moretti and Oleg Sobchuk. 2022. *Falso movimento : la svolta quantitativa nello studio della letteratura*. Extrema ratio. Nottetempo, Milano.

Cédric du Mouza, Stéphane Lamassé, and Philippe Rygiel. 2023. [De l'histoire numérique à l'histoire données ? Les Cahiers de Framespa. e-STORIA](#), (42). Number : 42 Publisher : UMR 5136 – FRAMESPA.

Erik Radio. 2020. [Minimal Computing and Ontologies](#). *Minimal Computing*.

Roopika Risam and Alex Gil. 2022. Introduction : The Questions of Minimal Computing. *Digital Humanities Quarterly*, 016(2).

Gauthier Roussilhe and Françoise Berthoud. 2021. [Le numérique face à la crise environnementale](#). *Alternatives Non-Violentes*, 199(2) :11–13. Publisher : Alternatives Non-Violentes. 327
328
329
330

Gauthier Roussilhe and Quentin Mateus. 2023. *Perspectives low-tech*. Divergences, Paris. 331
332

Silvia Spanu Fremder. 2010. *Le répertoire et la dramaturgie de la Comédie-Italienne de Paris durant la seconde moitié du XVIIIe siècle*. Thèse de doctorat en études italiennes (sous la direction d'Andrea Fabiano), Université Paris-Sorbonne, Paris. 333
334
335
336
337

Nicole Wild and David Charlton. 2005. *Théâtre de l'Opéra-Comique Paris : répertoire 1762-1972*. Collection Musique, musicologie. Mardaga, Sprimont. OCLC : ocm61753992. 338
339
340
341

Grant Wythoff. 2022. Ensuring Minimal Computing Serves Maximal Connection. *Digital Humanities Quarterly*, 016(2). 342
343
344