

Métadonnées, RiC et TheaterFalle

Cédric Degrange

Le modèle de description archivistique RiC-CM permet la saisie d'un fonds d'archives sous la forme de graphes de connaissances - ou ontologies. Cette solution signifie un enrichissement indéniable de la compréhension d'un fonds et sa mise en œuvre requiert des solutions techniques. Cependant, la description archivistique dans RiC relève avant tout de questions épistémiques sur la nature des documents et leur production, dont les réponses relèvent de l'archivistique. A travers l'exemple de la saisie d'un fonds, l'article montre l'importance pour l'archiviste de s'approprier le vocabulaire de RiC-CM afin de répondre à ces questions, mais surtout afin de conceptualiser de manière cohérente la description d'une collection documentaire.

*Der archivische Verzeichnungsstandard RiC-CM ermöglicht die Erfassung eines Archivbestands mittels Wissensgraphen – oder Ontologien. Diese Lösung bedeutet eine deutliche Bereicherung des Verständnisses eines Bestands, und ihre Umsetzung erfordert technische Lösungen. In erster Linie wirft die archivische Verzeichnung in RiC jedoch epistemische Fragen zur Natur der Dokumente und ihrer Produktion auf, die archivwissenschaftlich beantwortet werden müssen. Am Beispiel der Erfassung eines Bestands zeigt der Artikel, wie wichtig es ist, dass sich Archivar*innen das Vokabular von RiC-CM aneignen, um auf diese Fragen antworten zu können, aber vor allem, um die Beschreibung einer Dokumentensammlung in kohärenter Weise konzeptualisieren zu können.*

The RiC-CM archival description standard enables the description of archival holdings using knowledge graphs or ontologies. This solution significantly enriches the understanding of a collection and requires technical solutions for its implementation. First and foremost, however, archival description in RiC raises epistemic questions about the nature of documents and their creation, which need to be answered from an archival science perspective. Using the example of the descriptive processing of an archival fonds, the article shows how important it is for archivists to acquire the vocabulary of RiC-CM to be able to answer these questions,

but above all to be able to conceptualise the description of a document collection in a coherent way.

1 Introduction

L'intention de ce travail est double. D'une part, il y a la volonté d'analyser le cas pratique de la saisie du fonds d'archives du TheaterFalle dans la base de données de la fondation SAPA¹. D'autre part, cette base de données étant construite sur le modèle de description archivistique RiC-CM²³, il s'agit de présenter les notions qui fondent cette norme dans la perspective du cas pratique TheaterFalle.

La première partie théorique aborde les notions de métadonnées, langage RDF et ontologie, qui permettent une représentation des connaissances comme un réseau de données liées. Cette partie se termine par la présentation de la norme de description archivistique RiC. Nous examinerons brièvement les différents éléments de ce modèle qui permet de décrire un document à travers les contextes dans lesquels il s'insère.

La seconde partie, après une présentation du fonds TheaterFalle, se concentre sur les formulaires de saisie et décrit la solution mise en place par SAPA. Nous en faisons une analyse critique et proposons une autre approche qui nous semble plus en phase avec le modèle RiC et plus cohérente épistémologiquement.⁴

-
- 1 Fondée en 2017, la fondation SAPA, archives suisses des arts de la scène, ci-après SAPA, est issue de la réunion des collections du théâtre à Berne et de la danse à Lausanne, Zürich. L'acronyme SAPA vient de la désignation anglaise *Swiss Archives for Performing Arts*. La saisie du Fonds TheaterFalle a été effectuée dans le cadre d'un stage de six mois.
 - 2 International Council on Archives, « Records in Contexts Conceptual Models 0.2 (RiC-CM 0.2) », 2021.
 - 3 Cet article provient d'un travail déposé en octobre 2023 et se fonde sur la version 0.2 du modèle RiC-CM publiée en 2021. Il ne prend pas en compte la version 1.0 publiée en novembre 2023. Cependant, les principes théoriques restent inchangés d'une version à l'autre.
 - 4 Cette solution de SAPA fait l'objet d'un article de Baptiste de CoulonDe Coulon Baptiste, « Déploiement de la norme Records in Contexts pour la gestion des collections de la Fondation SAPA », *Revue électronique suisse de science de l'information (RESSI)* (24), 29.02.2024. En ligne: <<https://doi.org/10.55790/journals/ressi.2024.e1511>>, consulté le 28.03.2025., data archiviste de la fondation, article paru après la remise de ce travail. Lors de mon stage chez SAPA en 2023, nous avons eu l'occasion de discuter de l'utilisation de RiC pour la plateforme de SAPA. Si nos articles traitent de l'application de la norme archivistique, nos perspectives, complémentaires, sont différentes: plus technique et pragmatique dans l'article de de Coulon, plus épistémique et conceptuelle dans ce travail.

Notre perspective porte sur les aspects conceptuels et épistémiques du travail archivistique de saisie et de description. À travers notre examen de la littérature, il nous a semblé que ces aspects étaient peu abordés et que les discussions se focalisaient rapidement sur les enjeux techniques. L'examen de notre cas pratique soutient l'idée qu'une réflexion sur les notions épistémiques utilisées pour structurer une représentation des connaissances est nécessaire afin d'obtenir un modèle cohérent, épistémiquement plus riche, et probablement aussi plus facile en termes de recherche d'informations.

2 Discussion

2.1 Notions théoriques

Ce travail porte sur une description numérique d'un fonds d'archives. Cette description s'appuie sur un vocabulaire constitué de métadonnées, sur une syntaxe – dite RDF – les mettant en relation, et d'un standard de description archivistique RiC-CM. Ce sont ces trois notions qui font l'objet de cette section.

2.1.1 Métadonnées

Une description s'appuie sur un vocabulaire, qui dans le cas d'une description numérique, est constitué de métadonnées. Au sens étymologique, les métadonnées sont des données qui portent sur des données. Par conséquent le vocabulaire pour décrire le fonds concerné est constitué de données et de métadonnées.

Dans la littérature, il semble parfois difficile de savoir clairement ce qui distingue données et métadonnées. Une approche est celle de Pringuet « Une donnée est une information ; elle peut être par exemple une date, un lieu, une mesure, un événement, ou encore la description de quelque chose. »⁵ Nous notons qu'une donnée est une information qui porte sur « quelque chose ». Rapporté à des documents à décrire, nous dirons qu'une donnée porte alors sur une réalité du monde. Les métadonnées, elles, portent sur une donnée et « [font] ré-

5 Pringuet Virginie, « Du Thésaurus à l'ontologie : de l'usage des mots à la visualisation des données », in: *Vers un nouvel archiviste numérique*, Éditions de la Maison des sciences de l'homme, Paris, l'Harmattan INA éd, 2022, p. 153.

férence à des détails spécifiques de celle-ci. »⁶ Ainsi, une donnée permet un ancrage épistémique dans une réalité.

Cette première approche doit être précisée, l'usage montrant que le statut de donnée ou métadonnée peut changer. Une métadonnée peut être utilisée comme la donnée sur laquelle porte une métadonnée d'un niveau supérieur, de sorte que les niveaux en viennent à se superposer : « Just as you can always add an extra lego piece on top of another, you can always add another layer of meta-data to describe metadata. »⁷ Les niveaux supérieurs correspondent alors à des informations plus abstraites que ceux inférieurs suivant le point de vue descriptif adopté : « Was als Metaebene oder Datenebene gelten soll, ist abhängig von der Betrachtungsperspektive [...]. »⁸ Cette idée de point de vue peut s'appliquer aux données, puisque celles-ci reflètent le niveau de détail désiré sur la chose. Ce faisant, nous retrouvons l'idée de granularité de la description : la donnée étant alors l'information la plus proche de la réalité et permet un ancrage, bien que relatif, dans celle-ci.

Alemu souligne encore une autre caractéristique de la notion de métadonnée : « Metadata provides the context for information. Metadata makes context explicit. »⁹ Tandis que la donnée caractérise une réalité, une métadonnée, en étant d'un niveau d'abstraction plus abstrait, explicite le contexte de l'information sur laquelle elle porte et en permet une compréhension. Par exemple, les métadonnées « production » ou « institution » participent chacune à expliciter un contexte et une compréhension différente de l'information « théâtre ».

En résumé, nous dirons qu'une donnée est une information qui caractérise une chose, tandis qu'une métadonnée donne le contexte d'une information (qu'elle soit du niveau d'une donnée ou d'une métadonnée) ; où le rapport entre réalité, donnée et métadonnée dépend du point de vue descriptif choisi par un agent en accord avec ses intentions.

6 Ibid.

7 Hooland Seth van et Verborgh Ruben, *Linked data for libraries, archives and museums: how to clean, link and publish your metadata*, London, Facet Publishing, 2014, p. 3.

8 Kuhlen Rainer, Lewandowski Dirk, Semar Wolfgang et al. (éds.), *Grundlagen der Informationswissenschaft*, De Gruyter, 2022, p. 245. En ligne: <<https://doi.org/10.1515/9783110769043>>, consulté le 05.12.2023.

9 Alemu Getaneh, *The future of enriched, linked, open and filtered metadata: making sense of ifla, lrm, rda, linked data and bibframe*, London, Facet Publishing, 2022, p. 9.

2.1.2 RDF, RDFS et ontologie

Données et métadonnées font partie des éléments constitutifs du langage RDF – l’abréviation pour *Ressource Description Framework*. Plutôt que la notion de modèle de graphe¹⁰, nous utiliserons ici la notion de langage suivant Allemange et Hendler : « RDF, RDFS, and OWL are the basic representation languages of the Semantic Web, with RDF serving as the foundation. »¹¹ Pour notre propos, seuls les langages RDF et RDFS nous importent.¹² L’idée fondamentale de ces langages est le modèle de proposition Sujet-Prédicat-Objet, nommé triplet, qui permet d’associer deux informations à l’aide d’une relation.¹³ Le langage RDF correspond à la syntaxe de base indiquant comment agencer les mots d’un vocabulaire.

Dans le modèle RDF, le monde est décrit en premier lieu en termes de choses et de noms de choses : « In essence RDF reduces everything in the world that is knowable, or has existed, into either a thing or a thing’s label ; this separates the identity of a thing from strings of text [...]. »¹⁴ On peut aussi dire qu’un schéma RDF relie une ressource à une autre ressource ou à un littéral ; où une ressource est une chose décrite par des données et métadonnées, et un littéral est le nom d’une chose, à savoir une information qui caractérise directement une entité, c’est-à-dire une donnée. Par exemple, « Shakespeare a écrit Othello en 1603 » est représenté sous la forme de deux triplets: « Shakespeare » – « a écrit » – « Othello » et « Othello – date de – 1603 », où « 1603 » est considéré comme un littéral, « Shakespeare » et « Othello » des ressources¹⁵, « a écrit » et

10 « [RDF] est un modèle de graphe destiné à décrire formellement les ressources web et leurs données [...] », Wikipedia.fr, « Ressource Description Framework » (consultation : 05.09.2023)

11 Allemang Dean et Hendler Jim, *Semantic Web for the working Ontologist: effective modeling in RDFS and OWL*, Second Edition, Waltham, Mass, Elsevier, 2012, p. 27.

12 Les langages RDF et RDFS sont utilisés pour rendre compte d’un univers fermé, tel un fonds d’archives, tandis que OWL est plus riche en terme d’expressivité et permet de rendre compte d’univers ouverts, tel que le web.

13 « Proposition » est utilisé au sens de la logique des propositions : une affirmation vraie ou fausse. Le modèle de triplet Sujet-Prédicat-Objet est hérité de la logique des prédicats, une extension de la logique des propositions. Dans le triplet, l’objet est alors compris comme l’objet du prédicat.

14 Carlson Scott, Lampert Cory, Melvin Darnelle et al., *Linked data for the perplexed librarian*, Chicago, ALA Editions, 2020 (An ALCTS monograph), pp. 33-34.

15 Il s’agit ici alors de la personne et de la pièce, qui peuvent être décrites, non pas du nom qui désignent celles-ci.

« date de » des relations. Les deux triplets pouvant être liés par la ressource commune « Othello » (cf. schéma ci-dessous)

On peut alors prédiquer d'autres objets écrits par la ressource « Shakespeare » : « Hamlet » ou « La Tempête ». On peut former le triplet indiquant qu'il « a vécu » « en Angleterre », une caractéristique partagée avec la ressource « Virginia Woolf » dont on pourrait affirmer qu'elle « a lu » « Shakespeare ». À l'aide de triplets RDF, nous pouvons décrire les différentes caractéristiques d'un sujet identifié et lier des entités ayant des ressources communes, formant ainsi un réseau de triplets.

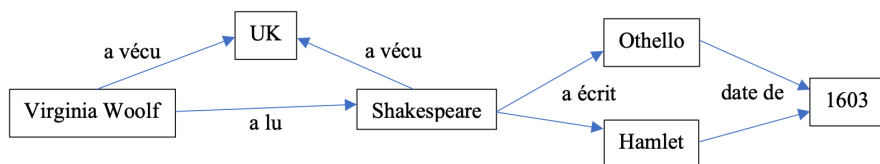


Figure 1. Exemple de réseau de triplets

Un intérêt de ce genre de réseau est la possibilité pour la machine d'établir des inférences. Par inférence, nous entendons l'émergence de nouveaux triplets à partir du réseau existant.¹⁶ Par exemple, de nos quelques triplets, nous pouvons inferer le triplet « Virginia Woolf » – « a lu » – « Hamlet » qui n'existe pas dans le réseau.¹⁷ Une nouvelle information est affirmée, qui enrichit ce réseau. Ce principe appliqué à un réseau plus grand permettra des connexions moins attendues et plus intéressantes.

Dans l'analyse du cas pratique, nous aurons aussi besoin de la notion de classe qui permet de réunir les éléments « Virginia Woolf » et « Shakespeare » sous la dénomination « auteur-ice ». Le langage RDF *Schema* (RDFS) – une extension de RDF – permet à une métadonnée de désigner une classe ou un ensemble¹⁸.

Une classe peut être définie par son extension,¹⁹ c'est-à-dire par la collection d'éléments partageant au moins une propriété – une classe peut aussi ne contenir qu'un seul élément. La classe « Auteur-ice » contient, dans notre exemple, la collection des deux éléments « Virginia Woolf » et « Shakespeare »

16 Allemang Dean et Hendler Jim, *Semantic Web for the working Ontologist*, op. cit., 2012, p. 123.

17 Où le triplet « Virginia Woolf – a lu – Shakespeare » signifie bien qu'elle a lu tous ses textes.

18 Nous utilisons indistinctement « classe » et « ensemble ».

définit une classe. Cette classe peut aussi être définie par son intention qui, ici, est « Auteur-ice ». L'intention correspond alors à une métadonnée. Une classe peut aussi contenir une sous-classe, qui en est une spécification.²⁰ Par exemple, « Auteu-ric(e) anglais-e » spécifie la classe « Auteur-ice ». Nous retrouvons le rapport d'abstraction entre les différents niveaux de métadonnées.

Les langages RDF et RDFS permettent de modéliser un réseau d'informations ou de connaissances nommés ontologies : « [...] ontologies are specification (rules) that deal with concepts (information) in a formal (standard) and shared (community-accepted) way. »²¹ Ramené à ce qui précède, nous dirons qu'une ontologie est une description dans un langage RDF ou RDFS (*rules*) d'une composition de classes, de sous-classes et d'instances (*standard*) caractérisées par des données, métadonnées et de relations (*information*) s'appuyant sur la représentation (*community-accepted*) d'une réalité.²²

Notons que si une ontologie se fonde sur des éléments formels, celle-ci demeure une représentation subjective d'un ensemble de connaissances - un objet épistémique. La notion d'ontologie autorise des perspectives dissemblables, de sorte que plusieurs représentation d'une même réalité peuvent se compléter, voire s'opposer.

19 CIDOC-CRM International Committee for Documentation, « Definition of the CIDOC Conceptual Reference Model », 2022, pp. 12-13. Ce modèle, qui a pour objectif « the reconstruction and interpretation of the past at a human scale », offre une présentation claire des notions de la théorie des ensembles.

20 Delestre Nicolas et Malandain Nicolas, *Du Web des documents au Web sémantique*, Bois-Guillaume, Klog, 2017, p. 97.

21 Carlson et al., *Linked data for the perplexed librarian*, op. cit., 2020, p. 62.

22 RDF, RDFS, RiC, ou encore CIDOC-CRM, sont des langages - avec des grammaires et des sémantiques plus ou moins riches - permettant de décrire une ontologie. Ils opèrent de manière similaire au français, à l'anglais, à l'arabe ou au pidgin, qui sont autant de langages pouvant servir à rendre compte d'une collection d'objets. Ces langages permettent ainsi la description d'une représentation de la réalité, mais description et représentation, bien qu'étroitement liées, sont de natures distinctes. La description est un objet linguistique, la représentation un objet épistémique. Le travail de saisie d'un fonds d'archives consiste alors d'abord dans la représentation de ce qui est - la compréhension du quoi -, puis dans la recherche des moyens de décrire cette représentation - le comment - en usant au besoin de différents langages. Si une fois achevée, une ontologie est à la fois une description et une représentation, il nous paraît important de souligner qu'elle est, en tant que compréhension de la réalité, d'abord un objet épistémique.

2.1.3 RiC-CM

Le modèle de description archivistique RiC – *Records in Contexts* – est fondé sur les langages RDF et se présente sous deux formes : l’une conceptuelle, le RiC-CM²³, l’autre technique, le RiC-O²⁴. Ce dernier correspond à l’implémentation informatique du RiC-CM selon le standard de langage *Web Ontology Language* (OWL). Ici, il ne sera question que du modèle conceptuel, le RiC-CM.

Dans RiC, nous retrouvons l’idée de triplets conçus alors à partir des notions d’entité, de relation et d’attribut. Une entité est liée à des entités ou des attributs, de manière similaire que dans le modèle RDF.

Les entités sont des objets conçus au sens large des choses (*things*). Quatre types d’entités constituent le cœur du modèle RiC²⁵ : *record Ressources*, *agent*, *activity* et *instantiation*. Les trois premières font l’objet de sous-catégories ou de catégories rattachées – reprenant la hiérarchie classe, sous-classe. Cas particulier, l’*instantiation* constitue une catégorie à elle seule. Au total, il y a 22 entités.

Pour caractériser ces entités, le modèle contient 41 attributs - plus quelques-uns spécifiques aux relations – et 78 relations pour désigner les liens entre ces entités. Les attributs servent ainsi principalement à décrire les entités, tandis que les relations servent à décrire les contextes dans lesquels s’insèrent les entités.

Si nous regardons plus en détail les entités, nous notons qu’une instantiation porte sur la matérialité – physique ou digitale – du document, renvoyant ainsi au support de l’information. Il peut s’agir d’un texte papier ou d’un fichier numérique. Les *record ressources* renvoient eux à la conception intellectuelle des documents. Il y a trois types de *record ressources* : le *record set*, le *record* et le *record part*. Ces deux derniers sont nécessairement liés à une instantiation au moment de leur création – celle-ci peut néanmoins ensuite disparaître²⁶. Par exemple, une édition de la pièce *Othello* est un *record* dont un exemplaire papier sera une instantiation, de sorte qu’il peut y avoir plusieurs instantiation

23 RiC-CM : *Records in Contexts - Conceptual Model*.

24 RiC-O : *Records in Contexts - Ontology*

25 Cette description s’appuie sur Clavaud et WildiClavaud Florence et Wildi Tobias, « ICA Records in Contexts-Ontology (RiC-O) : a Semantic Framework for Describing Archival Ressources », *Proceedings of Linked Archives International Workshop 2021*, 2021..

26 Ibid., p. 71.

d'un *record* ou *record part*.²⁷ Si ces instanciations disparaissaient, le *record* en tant qu'entité intellectuelle perdure. Enfin, un *record set* est un ensemble qui possède pour élément un ou plusieurs *records* ayant un ou des attributs communs, de sorte qu'il ne dépend pas directement d'une ou de plusieurs instanciations : « A *record set* [...] is an intellectual and not a physical aggregation. »²⁸ *Record set*, *record* et *record part* correspondent à des entités intellectuelles dépendant indirectement ou directement d'une instanciation, l'instanciation correspondant à une entité matérielle. L'ancrage dans la réalité se fait alors à travers l'instanciation.

La notion d'*agent* est définie en relation avec celles d'activité et de *records* : « *Agents* are entities that act or perform *activities* in the world and in the course of performing the *activities* that are the products or by-product of the *activity* performance »²⁹. Un agent peut être une personne ou un groupe de personnes. Dans ce dernier cas, il est possible de distinguer entre famille et corporation. Quant à l'entité *activity*, elle est définie comme : « The doing of something for some human purpose »³⁰. Une activité peut être comprise par rapport à sa finalité. Par exemple, la mise en scène d'une pièce de théâtre est le résultat d'une activité. Une activité peut aussi être comprise comme le processus lui-même. La focalisation se fera alors par exemple sur le travail de mise en scène d'une pièce de théâtre. Ainsi, suivant le point de vue adopté et les documents (les *records*) à disposition, une activité pourra être conçue selon un résultat ou selon un processus.

Dans RiC-CM, les entités occupent une place centrale dans la structure et le développement d'une ontologie. Elles sont les nœuds du réseau et sont les éléments que relient les relations. Le développement d'une ontologie fondée sur RiC-CM commencera par définir les entités nécessaires, puis les décrire au moyen d'attributs : « Once the metadata are reorganized per entity according to RiC, each entity described can be assigned a unique identifier and linked to other entities by using one of RiC relations. »³¹ Leur mise en réseau, et ainsi la description de leurs contextes se font dans un second temps. Par conséquent,

27 Le *record part* pourra correspondre à un acte de la pièce, une instanciation du *record part* sera l'acte imprimé dans l'édition papier, une autre instanciation en sera la photocopie.

28 International Council on Archives, « RiC-CM », *art. cit.*, 2021, p. 21.

29 *Ibid.*, p. 26.

30 *Ibid.*, p. 32.

31 Clavaud Florence et Wildi Tobias, « ICA Records in Contexts-Ontology (RiC-O): a Semantic Framework for Describing Archival Ressources », *art. cit.*, 2021, p. 4.

la démarche impose un important travail pour concevoir les entités et déterminer les attributs nécessaires. Elle impose par là une réflexion sur la perspective adoptée et les parties de la réalité à décrire que l'on souhaite modéliser.

2.2 Description du cas pratique

2.2.1 SAPA et www.performing-arts.ch

Les collections de SAPA comportent une partie documentaire et une partie archivistique. La partie documentaire (nommée « Special Collections ») possède une section « Personalia », composée des coupures de presse sur les personnalités des arts de la scène suisse, et une section « Panorama », rassemblant les documentations publiques produites par les institutions suisses, ainsi que les articles de presse. La partie archivistique rassemble des fonds de personnes ou d'institutions pouvant comporter un mélange de documents papier, documents audiovisuels ou artefacts. Cette politique d'acquisition permet aux collections de se compléter, de sorte qu'un fonds d'archives pourra être relié avec un ou des collections documentaires – et réciproquement.

Étant donnée la structure de ses collections et constatant l'augmentation à venir des documents digitaux, SAPA a choisi de développer une base de données sur le modèle archivistique RiC et les langages associés RDF et RDFS. Cette plateforme, toujours en développement, a été mise en ligne en 2020 (www.performing-arts.ch) et permet à la fois la consultation des documents et leur saisie. Les données de « 30'000 acteur·ices du théâtre suisse »³² ont ainsi été mises à disposition, l'accent étant mis sur la saisie et la préservation des documents numériques.

2.2.2 Le Fonds TheaterFalle

Le TheaterFalle est une compagnie de théâtre bâloise fondée par Ruth Widmer en 1986, active jusqu'en 2016. Ce fonds est une donation faite en 2023 et correspond à environ 1,50 m linéaire organisé autour des productions théâtrales, chacune faisant l'objet d'un dossier physique.

32 SAPA, « SAPA : Rapport annuel 2020 », Lausanne, Berne, Fondation SAPA, Archives suisses des arts de la scène, 2020, p. 19. En ligne: <https://sapa.swiss/wp-content/uploads/2021/06/SAPA_FR_2020_CMKY_RZ.pdf>.

L'inventaire du fonds a montré que les dossiers contenaient des documents textuels (tapuscrits des pièces, matériel de production ou matériel pédagogique) ou audiovisuels (vidéos, photographies ou cassettes audio). La priorité étant mise sur les documents vidéo, une douzaine de productions ont été sélectionnées pour numérisation. Les documents textes ont fait l'objet d'un tri en vue de compléter le dossier documentaire du « Panorama » consacré au TheaterFalle. Des documents ont ainsi été retirés du fonds au profit au dossier du « Panorama ».³³ Enfin, les doublons ont été éliminés. Au final, la saisie a porté sur des documents textes – tapuscrits, matériels de production ou matériel pédagogique –, et sur des documents iconographiques – photographies, diapositives des productions.

Pour cette saisie, l'unité de description correspond au dossier d'une production théâtrale, dossier conçu par la fondatrice du TheaterFalle Ruth Widmer.

2.2.3 Le Formulaire de saisie

Nous présentons tout d'abord la saisie telle qu'elle a été décidée par l'archiviste responsable, puis, dans la section suivante, nous en faisons l'analyse critique. Pour cette saisie, deux formulaires ont été utilisés : le formulaire correspondant au *record set* et celui correspondant au *record*. Le formulaire lié à l'entité *instantiation* étant en cours de validation, il n'a pas été utilisé, de sorte que la saisie reste au niveau intellectuel de la description des documents.

Le formulaire *record set* a été utilisé pour décrire un dossier de production lorsque celui-ci contenait une partie texte et une partie iconographique. Celles-ci étaient alors saisies avec la sous-catégorie *record*. Si un dossier contenait uniquement des documents texte, il était traité comme un *record*. Le plan de classement se présente alors de la manière suivante avec à droite les dossiers constitutifs du fonds:

33 Cette décision de SAPA vise un objectif d'exhaustivité de la collection documentaire et un gain de place.

<u>Plan de classement</u>	<u>Documents du fonds</u>
<i>Record set</i> : Leo und Ich	Dossier
<i>Record</i> : Leo und Ich : Textdokumente	
<i>Record</i> : Leo und Ich : Fotomaterial	
<i>Record</i> : Tango in Tanger	Dossier

Figure 2. Détail du classement du fonds TheaterFalle dans la plateforme de SAPA

Un dossier du fonds est ainsi traité comme un *record set* ou un *record* suivant qu'il possède un contenu photographique ou non. Singulièrement, l'entité correspondant à un dossier peut alors correspondre à des niveaux conceptuels différents. La distinction conceptuelle entre *record set* et *record* paraît alors pour le moins ambiguë et on se demande ce qui distingue la saisie du *record set* « Leo und Ich » de celle du *record* « Tango in Tanger ». Matériellement, les deux sont des dossiers de production.

Un document élaboré par SAPA explique la procédure de saisie. Il indique les champs à remplir obligatoirement et les métadonnées utilisées pour décrire une entité. La saisie est ainsi facilitée et peut se faire relativement rapidement.

Le formulaire pour un *record set* contient 20 métadonnées obligatoires, 1 recommandée, 15 facultatives – et 11 à ne pas utiliser. Seules les métadonnées obligatoires ont été utilisées.

Record set		
Nb métadonnées	Visées	Désignations
3	Identification	No de versement (RiC-A16), identifications individuelle (RiC-A16) et géographique (RiC-A16).
2	Unité de description	Niveau (RiC-A36), entrée parent (RiC-A36).
8	Entité	Titre (RiC-A28), dates extrêmes (2x) (RiC-A19 et A29 ?), auteur·rices (RiC-A28), importance matérielle (RiC-A04), forme documentaire (RiC-A05), langue (RiC-A25), localisation (RiC-A27 ?).
5	Aspects légaux	Mode d'acquisition (?), accès (3) (RiC-A01), utilisation (RiC-A09).
2	Saisie	Date (RiC-A19 et A29 ?), auteur·rices (RiC-A28).

Figure 3. Tableau des métadonnées utilisées pour le *record set*

Le formulaire pour un *record* contient 22 métadonnées obligatoires, 2 recommandées, 13 optionnelles et 12 à ne pas utiliser.

Record		
Nb métadonnées	Visées	Désignations
3	Identification	No de versement (RiC-A16), identifications individuelle (RiC-A16) et géographique (RiC-A16).
1	Unité de description	Niveau (RiC-A36), entrée parent (RiC-A36).
9	Entité	Titre (RiC-A28), dates extrêmes (2x) (RiC-A19 et A29 ?), auteur·rices (RiC-A28), importance matérielle (RiC-A04), forme documentaire (RiC-A05), langue (RiC-A25), localisation (RiC-A27 ?), Sujet (RiC-A28).
5	Aspects légaux	Mode d’acquisition (events ?), accès (3) (RiC-A01), utilisation (RiC-A09).
3 ajoutées non utilisées	Droits	Déclarations sur le droit d’auteur, droits d’utilisation capture vidéo, droits d’utilisation streaming
2	Saisie	Date (RiC-A19 et A29 ?), auteur·rices (RiC-A28).
2 recommandées et utilisées	Droit d’auteur et contenu	Mention du droit d’auteur (RiC-A09), description du contenu (RiC-A16).

Figure 4. Tableau des métadonnées utilisées pour le record

Étonnamment, il y a peu de différences entre les métadonnées obligatoires d’un *record set* et celles d’un *record*. Les changements – en couleurs dans le tableau –, sont la disparition de la métadonnée « niveau » RiC-A36, et l’ajout d’une visée « droit » avec trois métadonnées spécifiques aux documents vidéos – soulignant cette priorité. À cela s’ajoute une métadonnée permettant d’introduire des mots-clefs, mais étant obligatoire seulement pour les documents audiovisuels, elle n’a pas été utilisée. Par contre, deux métadonnées facultatives – en jaune dans le tableau – ont été employées : l’une permet la description du contenu, l’autre la mention du droit d’auteur pour les documents photographiques.

Relevons que dans le plan de classement, le fonds « TheaterFalle » s’insère sous « Personnes morales », mais que le « TheaterFalle » en tant qu’agent produisant des pièces de théâtre n’apparaît pas. L’entité RiC-E11 « corporate body » n’est ainsi pas utilisée. Le « TheaterFalle » n’apparaît que dans la partie documentaire « Panorama » renvoyant aux dossiers contenant articles de presse et documentation. Par contre, la fondatrice Ruth Widmer est décrite en tant

qu'agent auquel les dossiers du fonds TheaterFalle sont liés (grâce à la métadonnée RiC-A28 « Auteur-ices »).

2.2.4 Analyse critique et propositions

La comparaison de l'ontologie développée pour décrire les collections de la fondation avec les prescriptions de RiC-CM suscite plusieurs interrogations quant aux choix des entités et de leur description.

Un premier élément à souligner est que le document de saisie ne porte que sur la description des entités appartenant à la catégorie *record ressources*, ainsi qu'à la catégorie *instantiation*. Les catégories d'entités *agent* et *activity* ne sont pas décrites et ne sont pas utilisées dans une saisie standard. L'ontologie développée est constituée principalement de *records ressource*. Les contextes liés aux producteurs et à leurs actions ne sont donc pas décrits. Cette absence nous paraît entraîner des confusions et des complications inutiles.

Une série d'interrogations est causée par la similarité entre les entités *record set* et *record*. Cette similarité apparaît notamment dans la redondance des métadonnées de sorte que l'on ne voit pas le gain épistémique d'un niveau *record set*.

Si un *record set* rassemble des *records* ayant un ou des attributs en commun, cela ne signifie pas que ces attributs doivent être repris dans la description de l'ensemble. Au contraire, le principe de la hiérarchie des *record ressources* est que les métadonnées portant sur les attributs des entités, elles ne *sont* pas ces attributs. Autrement dit, le passage au *record set* devrait induire un changement de niveau d'abstraction et ainsi permettre un gain épistémique. Ce principe du gain épistémique, nous semble aussi aller dans le sens d'une hiérarchie d'informations en cascade, qui permettrait de clarifier l'attribution des métadonnées suivant les niveaux et éviter tant que faire se peut les redondances, mais aussi le choix des entités à utiliser.

Sur la question des entités, il nous semble plus cohérent d'utiliser les entités *record* et *activity*. Les documents textes et photographiques sont alors traités comme des *records* et font alors l'objet d'une description à l'aide des métadonnées du document de saisie avec des cotes d'identification et de localisation uniques. Ces *records* sont ensuite considérés comme les produits d'une activité, à savoir la *création* d'une pièce de théâtre décrite comme une entité *activity*. Poussant plus loin l'usage de RiC-CM, cette entité *activity* peut être liée à l'entité *Person* « Ruth Widmer » et à l'entité à créer *corporate body* « TheaterFalle ». Dès lors, il y a une différence évidente entre la description de la création d'une pièce

de théâtre liée à des agents (institutions, metteur·ses en scène ou acteur·ices) et les produits de cette activité qu'ils soient sous forme papier, photographiques ou vidéo.

Nous obtiendrions ainsi un plan de classement structuré comme suit :

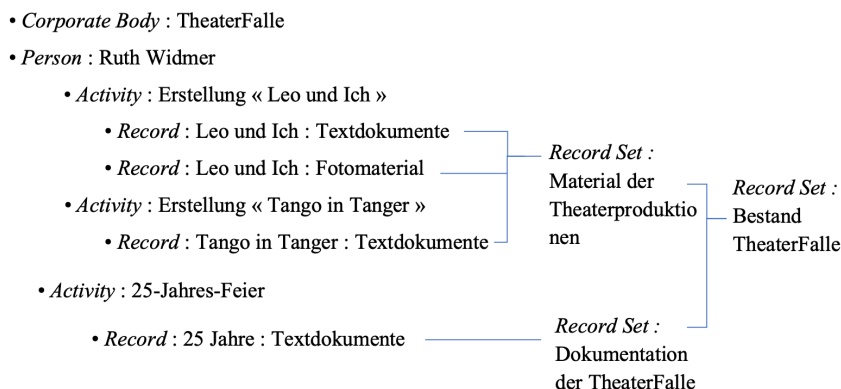


Figure 5. Exemple d'utilisation des entités *agency* et *activity* pour le fonds TheaterFalle

L'ensemble des *records* liés au TheaterFalle serait alors un *record set* intitulé « Material der Theaterproduktionen » et constitue un classement parallèle ou en sus à celui de gauche. Il y a ainsi un début de réseau qui offre à la fois une description du contexte de production des documents, ainsi qu'une description de leur contenu matériel et des ensembles qu'ils constituent. Ce faisant, on génère un gain informationnel qui pourrait augmenter en spécifiant d'autres ensembles ou en élargissant le réseau.

Ces gains épistémiques se font aussi au niveau des descriptions des entités. Par exemple, le nombre de productions que comporte le fonds n'apparaît pas en l'état, mais peut être attribué de manière cohérente au *record set* « Material der Theaterproduktionen ». Cette approche évite aussi d'avoir une multiplication des cotes uniques (dans la saisie actuelle, *record* et *record set* possèdent chacun une cote d'identification unique). Dans notre description, seuls les *records* possèdent une cote d'identification unique, le *record set* possédant une cote générique.³⁴ Cette solution fait des *records* les entités centrales de cette ontologie autour desquelles se développent les contextes décrits avec les entités à disposition dans RiC-CM.

L'usage des entités *agent* et *activity* permettrait aussi de palier un manque lié à la spécificité de SAPA, qui est les arts de la scène. RiC-CM offre peu d'attributs pour caractériser le sujet ou le domaine des *record ressources*. En n'utilisant que celles-ci, l'ajout devrait passer par l'attribut *descriptive note* (RiC-A16) et la rédaction d'un texte. Par contre, l'emploi des entités *agent* et *activity* donne l'accès à des attributs spécifiques au type d'institution ou à l'activité de cette entité. Ainsi, l'attribut *occupation type* (RiC-A30), répétable, permet d'indiquer la ou les différentes occupations d'une entité *person* comme acteur·ice, metteur·se en scène ou danseur·se. De même, l'attribut *corporate body type* (RiC-A12) permet de décrire une entité tels que Centre de création théâtrale, Compagnie de danse ou de théâtre. Enfin, le type d'une entité *activity* peut être caractérisé par l'attribut *activity type* (RiC-A02) et désigner une production, une création ou une représentation. Ce faisant, il est possible d'enrichir le contenu sémantique d'une ontologie.

L'usage de multiples entités et ce faisant l'ajout d'attributs peuvent paraître une complication exigeant temps et énergie, ce qui est vrai. Cependant cela paraît vrai surtout au début du processus de conceptualisation de l'ontologie. La difficulté semble principalement de choisir les différentes entités nécessaires structurant l'ontologie à décrire puis de spécifier et répartir les métadonnées entre elles. La saisie paraît ensuite devoir gagner en clarté et efficacité, et donc en temps et énergie. Par ailleurs, l'usage des différentes entités, et c'est l'un des intérêts de RiC-CM, permet d'enrichir la représentation d'un fonds et d'en souligner pour partie sa valeur épistémique.

Un autre argument en faveur de l'emploi, ici, de nouvelles entités est lié à la recherche d'information dans une base contenant des données au format RDF. De telles bases de données exigent pour faire des requêtes de passer par l'emploi du langage SPARQL conçu pour les machines et les humains. Si ce langage permet d'obtenir des réponses précises à des questions complexes, il implique une difficulté d'emploi qu'il s'agit de surmonter. Cette difficulté m'a paru d'autant plus grande dans le cas de la plateforme de SAPA que quasi toutes les entités qui structurent son ontologie sont des *record ressources*, de sorte que la distinction entre *records set* et *record* devient ambiguë. Par conséquent, une requête portant uniquement sur les *records set* n'inclut pas le *record* « Tango in

34 Notre réflexion met de côté l'entité *Instantiation*, celle-ci n'étant pas disponible lors de cette saisie. En l'introduisant, une cote dérivée de celle de son *record* devrait lui être attribuée. Cependant, le *record* en tant qu'entité intellectuelle perdure même si toutes ses *instantiations* venaient à disparaître, de sorte que c'est lui qui fait foi en dernier lieu.

Tanger » qui correspond pourtant aussi à un dossier de production (cf. figure 2). La requête devra alors porter sur les *records set* et les *records*, et signifie une connaissance préalable de l'ontologie³⁵. En utilisant, différentes entités et des attributs spécifiques à celles-ci, les requêtes sur les *records* liées aux activités de tel agent nous paraissent facilitées. Néanmoins, cette remarque est, en l'état, une hypothèse, car il ne nous a malheureusement pas été possible d'analyser plus en détail la base de données de SAPA, ou d'en discuter avec la personne responsable. De plus, si nous avons pu avoir la présentation d'une autre base de données sur ce modèle et découvrir une manière de l'interroger bien plus simple, nous n'avons pu véritablement tester celle-ci. Néanmoins, les requêtes SPARQL paraissent moins complexes à rédiger et les résultats obtenus plus précis. Il semble bien y avoir un gain manifeste en termes de valorisation du fonds et de recherche d'informations à structurer clairement et distinctement une ontologie.

3 Conclusion

Avec ce travail, nous avons voulu décrire les différentes notions qui fondent une ontologie de type RDF, cela afin d'examiner leur portée épistémique. Il nous semble au final que nous n'avons qu'effleuré l'articulation de ces notions et leur portée. L'analyse et la critique d'un cas pratique ont cependant permis d'aborder une partie des enjeux en suggérant la nécessité d'une réflexion sur la partie de réalité à décrire, sur le choix des entités pour ancrer et développer une ontologie qui permet une compréhension plus précise des informations. L'examen des enjeux épistémiques nous paraît ainsi important pour aller au-delà d'une archive constituée uniquement de *dossiers*, et capter des connaissances qui permettent une meilleure compréhension et un usage facilité du fonds décrit.

Bibliographie

Alemu Getaneh, *The future of enriched, linked, open and filtered metadata: making sense of ifla, lrm, rda, linked data and bibframe*, London, Facet Publishing, 2022.

35 Il semble que le principe de *monotonicity* n'est alors pas respecté : « Monotonicity requires that adding new classes and properties to the model or adding new statements to a knowledge base does not invalidate already modelled structures and existing statements. » International Committee for Documentation, « CIDOC CRM », *art. cit.*, 2022, p. 29.. Dans la description de SAPA, les classes *Records set* et *Records* sont dans les faits fusionnées formant une nouvelle classe.

- Allemang Dean et Hendler Jim, *Semantic Web for the working Ontologist: effective modeling in RDFS and OWL*, Second Edition, Waltham, Mass, Elsevier, 2012.
- Carlson Scott, Lampert Cory, Melvin Darnelle et al., *Linked data for the perplexed librarian*, Chicago, ALA Editions, 2020 (An ALCTS monograph).
- Clavaud Florence et Wildi Tobias, « ICA Records in Contexts-Ontology (RiC-O) : a Semantic Framework for Describing Archival Ressources », *Proceedings of Linked Archives International Workshop 2021*, 2021.
- De Coulon Baptiste, « Déploiement de la norme Records in Contexts pour la gestion des collections de la Fondation SAPA », *Revue électronique suisse de science de l'information (RESSI)* (24), 29.02.2024. En ligne: <<https://doi.org/10.55790/journals/ressi.2024.e1511>>, consulté le 28.03.2025.
- Delestre Nicolas et Malandain Nicolas, *Du Web des documents au Web sémantique*, Bois-Guillaume, Klog, 2017.
- Hooland Seth van et Verborgh Ruben, *Linked data for libraries, archives and museums : how to clean, link and publish your metadata*, London, Facet Publishing, 2014.
- International Committee for Documentation, « Definition of the CIDOC Conceptual Reference Model », 2022.
- International Council on Archives, « Records in Contexts Conceptual Models 0.2 (RiC-CM 0.2) », 2021.
- Kuhlen Rainer, Lewandowski Dirk, Semar Wolfgang et al. (éds.), *Grundlagen der Informationswissenschaft*, De Gruyter, 2022. En ligne: <<https://doi.org/10.1515/9783110769043>>, consulté le 05.12.2023.
- Pringuet Virginie, « Du Thésaurus à l'ontologie: de l'usage des mots à la visualisation des données », in: *Vers un nouvel archiviste numérique*, Éditions de la Maison des sciences de l'homme, Paris, l'Harmattan INA éd, 2022.
- SAPA, « SAPA : Rapport annuel 2020 », Lausanne, Berne, Fondation SAPA, Archives suisses des arts de la scène, 2020. En ligne: <https://sapa.swiss/wp-content/uploads/2021/06/SAPA_FR_2020_CMKY_RZ.pdf>.