



HAL
open science

Pour une valorisation en ligne des archives audiovisuelles du SPAV du Centre Pompidou : pistes de modélisation

Sarah Ambec

► To cite this version:

Sarah Ambec. Pour une valorisation en ligne des archives audiovisuelles du SPAV du Centre Pompidou : pistes de modélisation. 2024, 10.58079/12hmp . hal-04760225

HAL Id: hal-04760225

<https://hal.science/hal-04760225v1>

Submitted on 30 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Pour une valorisation en ligne des archives audiovisuelles du SPAV du Centre Pompidou : pistes de modélisation

Sarah Ambec (Bibliothèque Kandinsky, MNAM-CCI, Centre Pompidou), Billet du carnet Hypothèses du projet MAVEX (Mémoires audiovisuelles des expositions du Centre Pompidou), 14 octobre 2024, <https://mavex.hypotheses.org/3062> et <https://doi.org/10.58079/12hmp>

Pendant deux mois, du 15 mai au 12 juillet 2024, dans le cadre de son stage d'observation à la bibliothèque Kandinsky, Sarah Ambec (étudiante en M1 du Master Technologies numériques appliquées à l'histoire de l'École des Chartes) a rejoint le projet MAVEX pour travailler sur ses aspects applicatifs. Elle décrit ici les enjeux de sa mission, les étapes et les résultats de son travail de modélisation pour permettre, à terme, de valoriser en ligne les archives audiovisuelles du Service de production audiovisuelle (SPAV) du Centre Pompidou.

Dans la perspective de la préfiguration du développement d'un outil de mise en ligne de l'inventaire du fonds d'archives audiovisuelles du SPAV du Centre Pompidou, j'ai tenté d'identifier les implications techniques du projet. J'ai principalement été amenée à travailler sur l'architecture des données de ce fonds, en cherchant à déterminer quel modèle de données adopter pour documenter les archives audiovisuelles. Cette tâche nécessite une approche technique documentée, afin d'insérer cet outil dans l'écosystème numérique foisonnant du Centre Pompidou, et pour assurer l'interopérabilité des données, tant en interne qu'en externe. Elle est d'autant plus nécessaire que la valorisation de ce fonds implique de se positionner au carrefour de trois logiques documentaires différentes qui animent chacune le Centre Pompidou : la logique documentaire du musée, celle des archives, et enfin celle des bibliothèques. Chacune de ces logiques dispose de son propre modèle de données de référence.

Focus sur les trois modèles de données de référence dans les domaines des musées, des archives et des bibliothèques

- Dans le **domaine des musées**, l'ICOM a développé le [CIDOC CRM](#) (Conceptual Reference Model) dans le but de faciliter l'échange d'informations sur le patrimoine culturel. Cette norme offre une structure riche et détaillée, idéale pour la description des œuvres d'art et objets de patrimoine, et plus largement de tout objet culturel dont des documents audiovisuels. CIDOC CRM permet de modéliser des événements complexes, incluant la production, la conservation, et l'utilisation des documents audiovisuels. Cependant, bien que ce modèle soit très flexible et puisse représenter des relations complexes, il peut s'avérer trop complexe et technique pour les bibliothèques et les archivistes qui ne sont pas spécialisés dans ce domaine. Intégrer les archives audiovisuelles dans CIDOC CRM nécessite une compréhension approfondie du modèle, ce qui peut être une barrière pour une adoption plus large.
- Dans le **domaine des archives**, le Conseil international des archives ([ICA](#)) publie la norme [ISAD\(G\)](#) (International Standard Archival Description) destinée à uniformiser la description des archives. Désormais intégrée au modèle [RIC](#) (Record In Context), elle permet également d'assurer la richesse en métadonnées des instruments de recherches. Ce système de description offre une grande souplesse qui permettrait au mieux d'accommoder chaque fonds audiovisuel du Centre Pompidou. Mais ISAD(G) est avant tout un modèle pensé pour des

archives papiers, et si RIC semble plus prometteur pour embrasser la diversité des objets dans sa description, il n'est pas encore largement appliqué.

- Dans le **domaine des bibliothèques**, la Fédération internationale des associations et institutions de bibliothèque ([IFLA](#)) développe et maintient le modèle [LRM – Library Reference Model](#) (LRM). Celui-ci est une norme bibliographique conçue pour unifier les concepts bibliographiques à travers les différents systèmes de catalogage de bibliothèques. Bien qu'étant une norme générale et inclusive, LRM rencontre des difficultés lorsqu'il s'agit de décrire des archives audiovisuelles. En particulier, LRM n'intègre pas la notion d'événement, essentielle pour décrire le processus de production et les différentes étapes d'un document audiovisuel, y compris les copies et les intermédiaires. Adapter LRM aux archives audiovisuelles nécessite donc des extensions ou des modifications pour inclure des concepts spécifiques à ces types de documents. Héritier du modèle FRBR, le modèle LRM est largement reconnu, utilisé et étudié, ce qui nous permet plus facilement de l'inclure dans un travail de web sémantique. La communauté active autour du LRM apporte également des réponses aux limites du modèle. Des discussions communes avec le CIDOC et IFLA ont donné lieu à la publication d'un modèle commun, FRBROO puis LRMoo (orienté objet). Pour l'audiovisuel, la Fédération internationale des archives du film ([FIAF](#)) s'est appuyée sur LRM pour produire des recommandations de catalogages adaptées aux archives cinématographiques.

Les archives audiovisuelles du Centre Pompidou sont aujourd'hui conservées dans une base de données relationnelle interne et il nous faut, pour les valoriser, normaliser leur description par la modélisation. Elle doit nous permettre de conserver un maximum d'informations sur ces archives ainsi que les rendre plus accessibles.

La réalisation d'une modélisation conceptuelle des données

Pour représenter au mieux les besoins de valorisation des archives audiovisuelles, j'ai conçu un modèle de données qui offre une représentation graphique simple des données et de leurs relations. Ce modèle, bien que non conforme aux normes UML, s'en inspire pour rester lisible et servir d'outil de dialogue entre acteurs techniques et non techniques. Il permettrait de créer une base de données SQL respectant le schéma et les liens entre les tables établies par le modèle. Ce modèle « idéal » est le fruit d'une réflexion qui a impliqué plusieurs étapes, dont je tiens ici à expliquer les détails. Le travail de modélisation implique une fine analyse du corpus concerné, une recherche parmi les normes les plus adaptées, et enfin de la créativité pour réussir à dépasser un cadre qui pourrait être trop rigide pour ces archives.

Réflexion sur le corpus à documenter

Pour initier la modélisation, j'ai d'abord réalisé un schéma sans contraintes, basé uniquement sur le contenu du corpus. Cette étape permet d'inventorier les différents types d'objets présents dans les archives audiovisuelles, tels que films, captations d'expositions et d'événements, vidéos, etc. L'objectif est de bien représenter la diversité du fonds et d'identifier tous les éléments à intégrer dans la base de données.

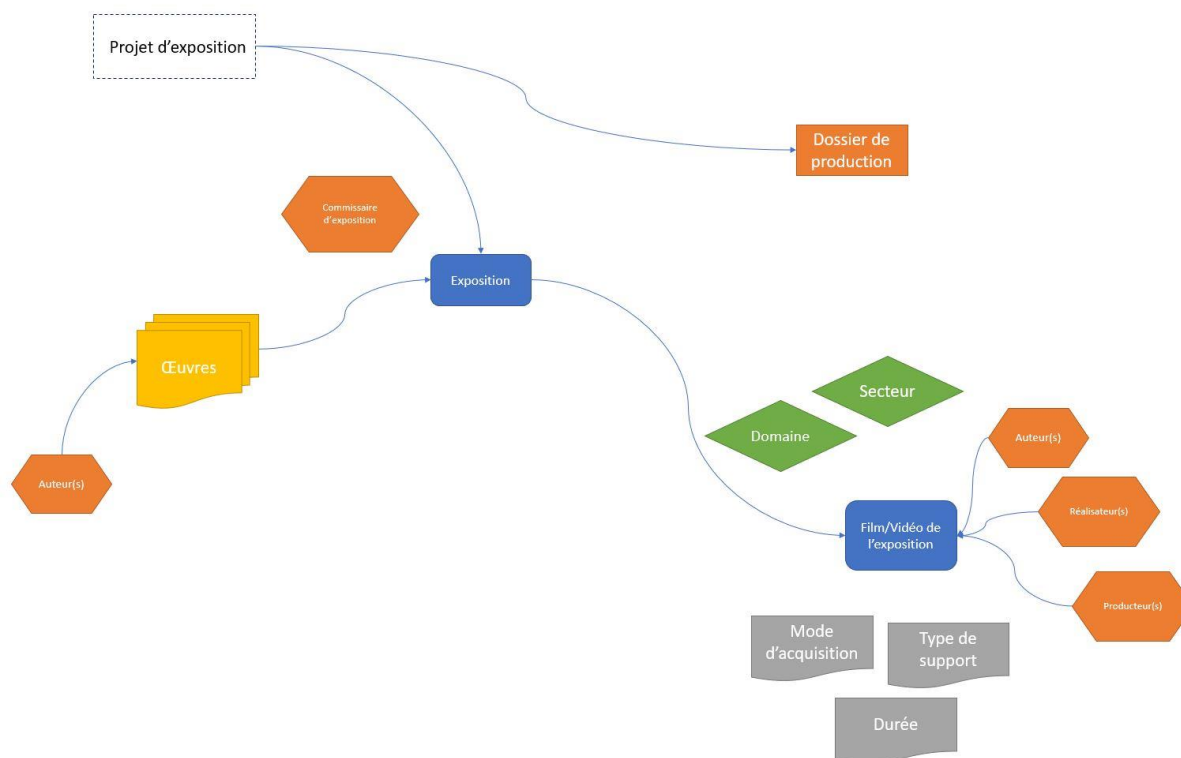


Figure 1 - Processus de création d'une archive audiovisuelle d'exposition (schéma de Sarah Ambec, MAVEX, bibliothèque Kandinsky]

Reprendre le modèle LRM

Étant donné la diversité des archives audiovisuelles du Centre Pompidou, un cadre normé est nécessaire. Nous nous sommes tournés vers le modèle LRM (Library Reference Model) qui est donc une norme de bibliothèque. Elle est notamment adoptée pour le catalogage des collections de la Bibliothèque Kandinsky mises en ligne sur son portail. La [Transition Bibliographique](#) a fourni des ressources utiles pour documenter ce cadre et son application dans les bibliothèques françaises. Le modèle LRM présente l'avantage d'être simple à appliquer par sa rigidité. C'est ce cadre immuable qui permet de décrire de la même manière une livre que l'on se trouve à Paris ou à Montréal. Toutefois, ce modèle est pensé pour des objets produits de manière identiques en plusieurs exemplaires, des livres édités, non pour des objets uniques comme nos archives. La norme des musées, le CIDOC CRM, permettrait de prendre en compte la dimension unique de ce fonds. Mais le modèle conceptuel du CIDOC est très complexe et donc couteux en termes de formation requise pour indexer les objets, et en termes d'infrastructures numériques pour maintenir toutes ces relations. De ce fait, nous avons fait le choix du modèle LRM afin de garantir une pérennité à notre indexation des archives audiovisuelles du Centre Pompidou, en l'inscrivant dans l'écosystème de la Bibliothèque Kandinsky et en permettant plus aisément de cataloguer les nouvelles entrées.

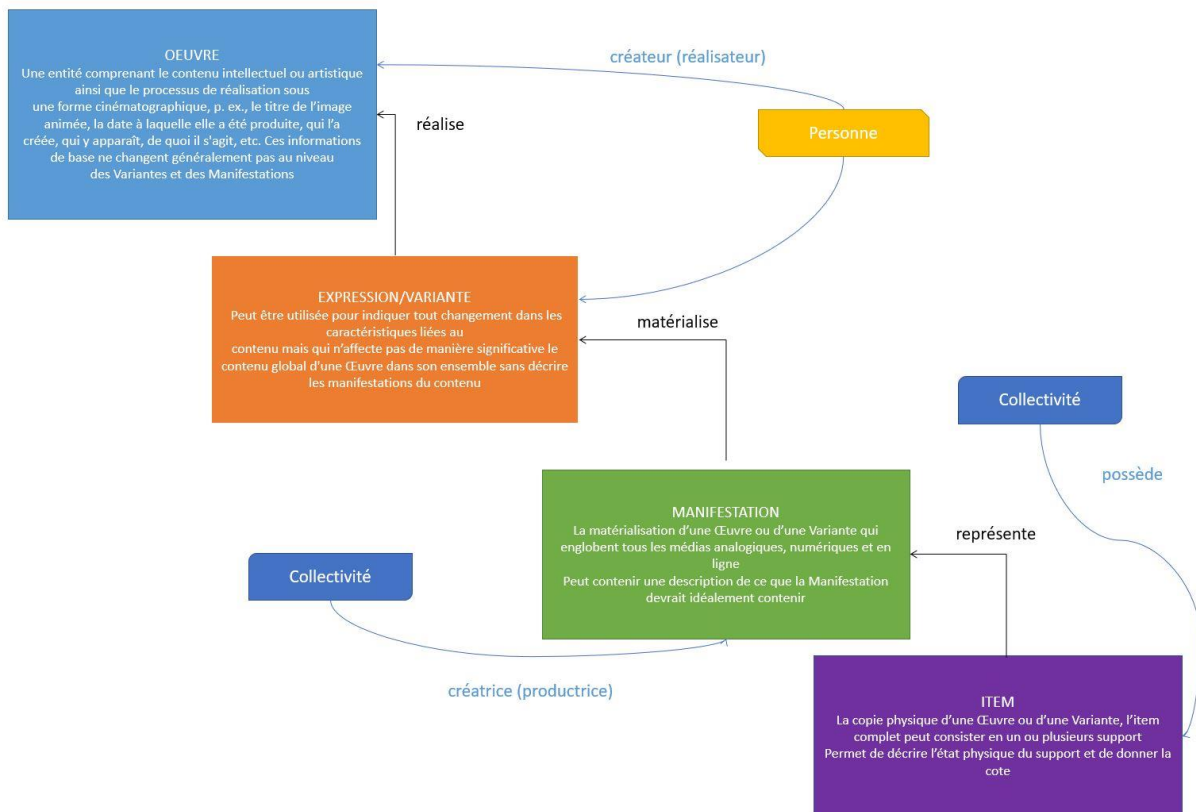


Figure 2 - Le modèle LRM selon les définitions de la Transition Bibliographique (schéma de Sarah Ambec, MAVEX, bibliothèque Kandinsky)

Les défis de l'application des normes

LRM, bien qu'étant une norme bibliographique large et générale, ne s'applique pas parfaitement à tous les types d'objets comme nous l'avons évoqué. C'est le cas pour les archives audiovisuelles. Le modèle LRM ne prend pas en compte la notion d'événement – cruciale pour décrire le processus de production des documents audiovisuels –, la diversité de ses composantes – versions intermédiaires, alternatives et copies. Cette notion d'évènement, présente dans le CIDOC, nous permettrait de retracer le processus de production d'un document audiovisuel. Cette lacune complique l'adaptation de LRM aux films et aux captations d'expositions. Pour pallier ce problème, nous avons étudié les solutions proposées par les ontologies de la FIAF (Fédération internationale des archives du film) et de l'IFLA (Fédération internationale des associations et institutions de bibliothèques). La FIAF a publié en 2022 un [Manuel de catalogage des images animées de la FIAF](#), offrant une structure de métadonnées adaptée aux films. L'IFLA a publié en 2020 une nouvelle version du modèle LRMoo. LRMoo (pour object oriented) est un nouveau modèle construit en collaboration entre la logique de catalogage des bibliothèques et celle des musées. C'est une extension de LRM qui cherche une interopérabilité avec CIDOC CRM, avec notamment l'introduction dans la [version du 0.6](#) de l'élément « Recording Event » qui nous permet de qualifier les captations d'expositions.

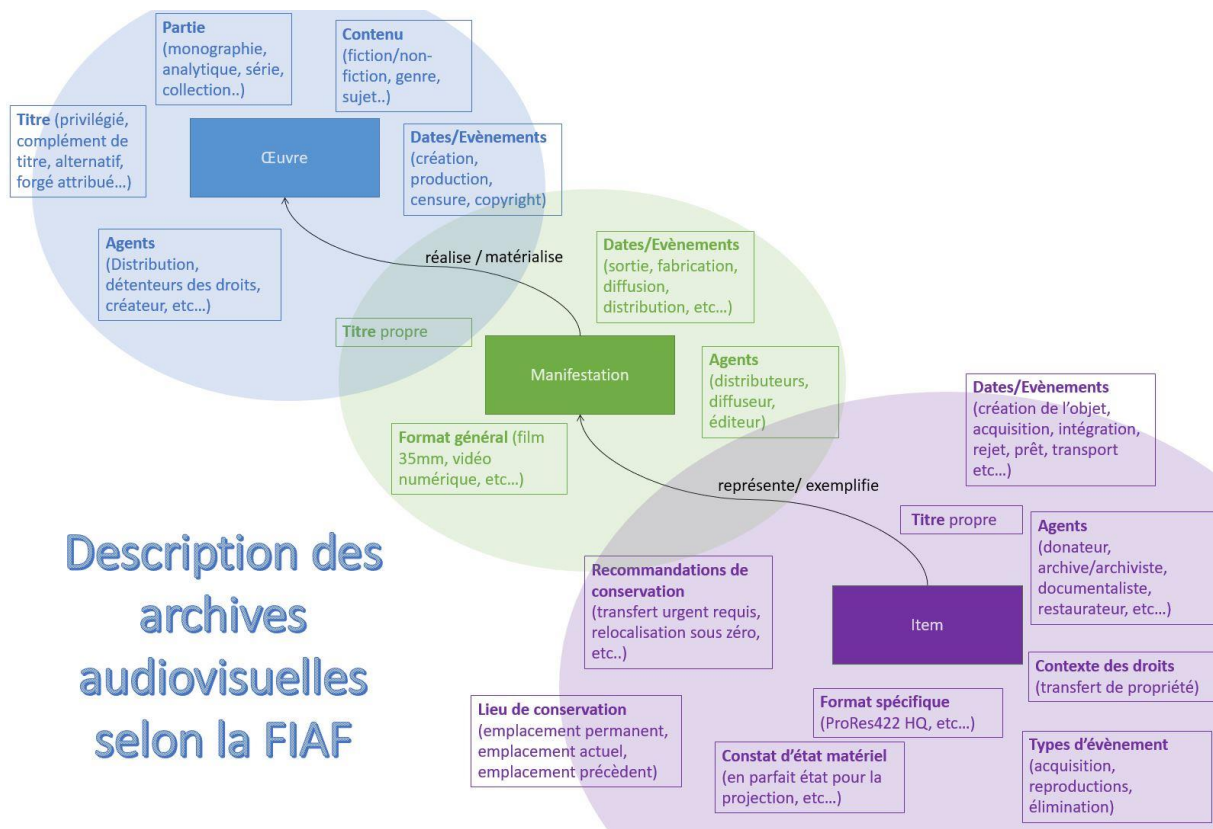


Figure 3 - Description des archives audiovisuelles selon la FIAF (schéma de Sarah Ambec, MAVEX, bibliothèque Kandinsky)

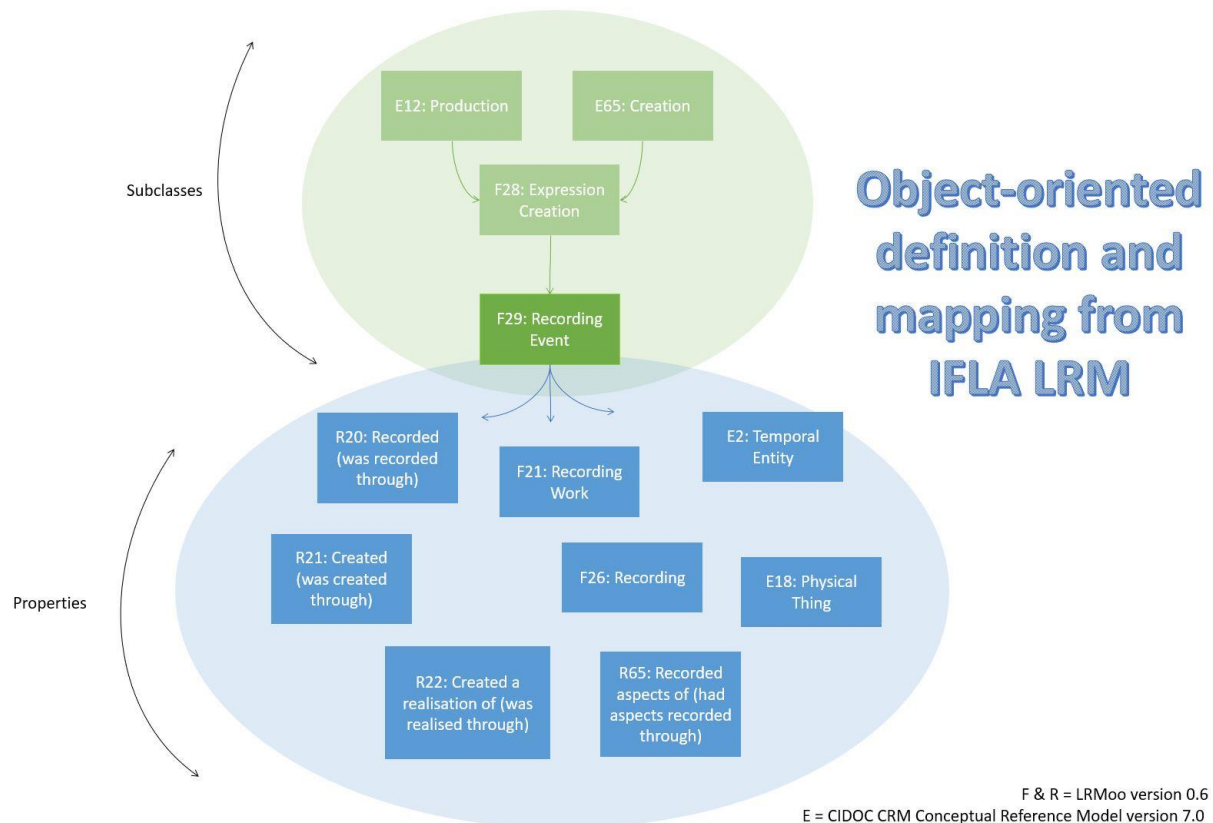


Figure 4 - Modélisation des attributs de l'évènement selon LRMoo (schéma de Sarah Ambec, MAVEX, bibliothèque Kandinsky)

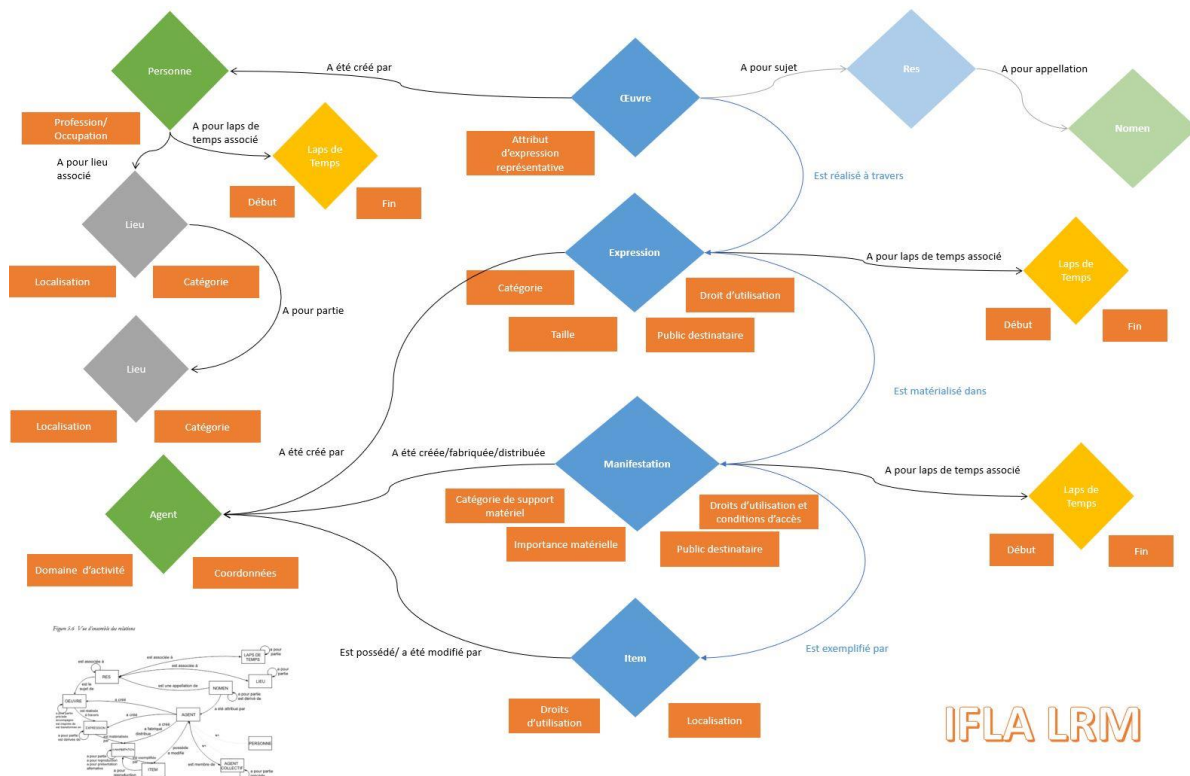


Figure 5 - Modélisation de l'ontologie LRM (schéma de Sarah Ambec, MAVEX, bibliothèque Kandinsky)

Déterminer un « modèle idéal »

En combinant les caractéristiques données aux entités selon les manuels de catalogages d'IFLA LRM et selon la FIAF, nous pouvons dresser un nouveau schéma qui permet de décrire au mieux les archives audiovisuelles de notre corpus. J'ai dans les schémas suivants mis face à face les deux descriptions selon les recommandations différentes afin de pouvoir les comparer.

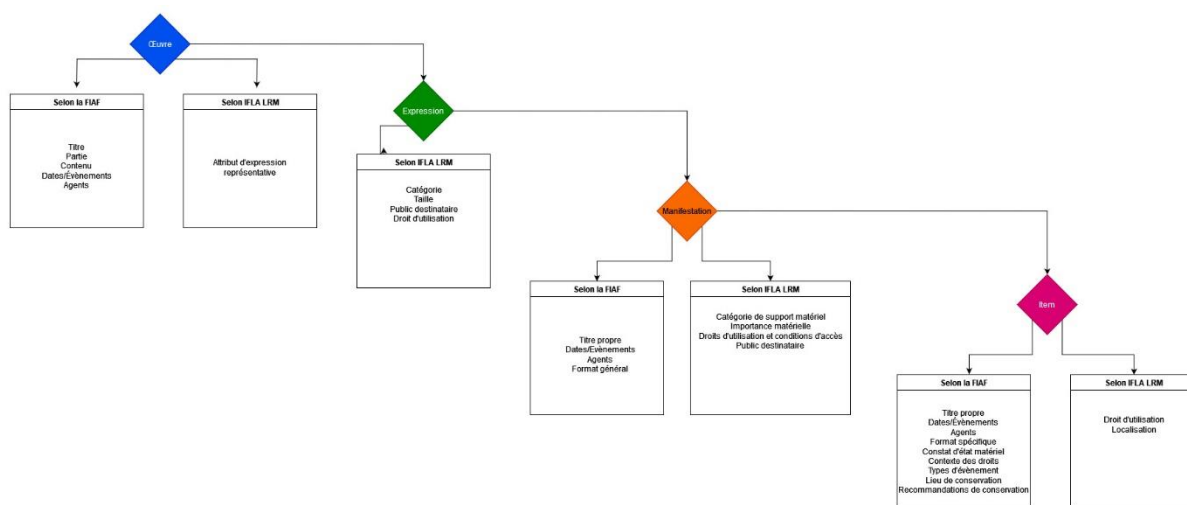


Figure 6 - Comparaison des attributs proposés aux entités LRM par IFLA LRM et par la FIAF (schéma de Sarah Ambec, MAVEX, bibliothèque Kandinsky)

Appliquer à un exemple

Pour mieux se représenter l'intérêt de chacune de ses propositions (IFLA LRM ou FIAF), un exemple détaillé permet d'appréhender la difficulté que représente ce fonds. En effet, l'état de la description actuelle des archives audiovisuelles est parcellaire. D'une part, tous les attributs du modèle de la FIAF ou d'IFLA LRM ne trouvent pas de correspondance dans la base de données actuelle. D'autre part, tous les champs de cette base ne sont pas uniformément remplis. Le schéma ci-dessous traduit ces lacunes.

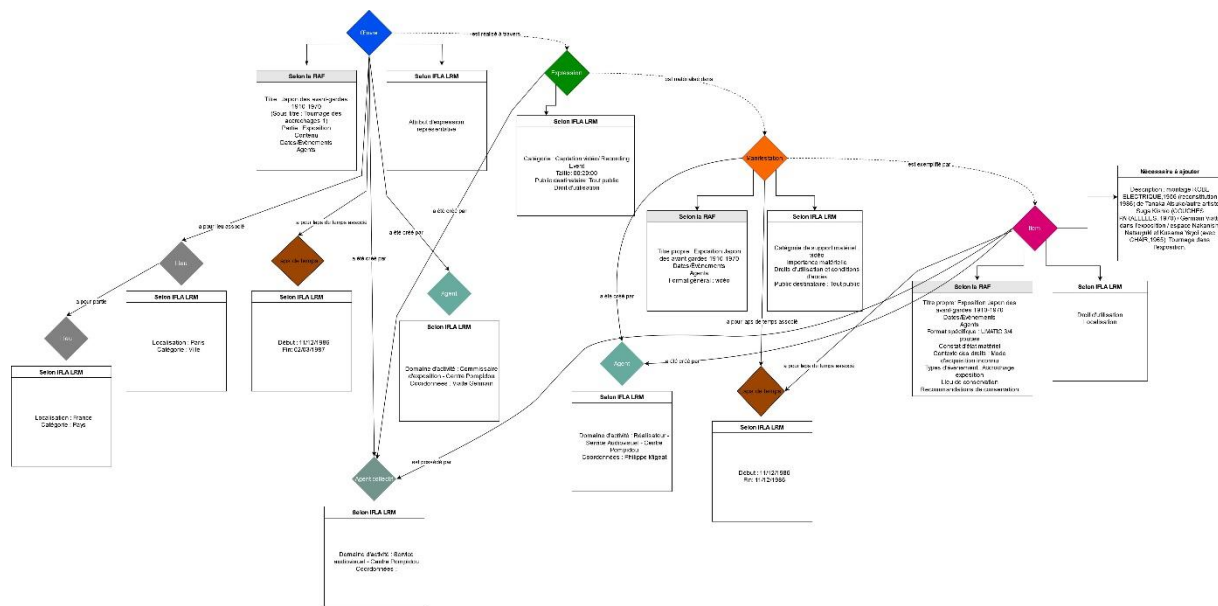


Figure 7 - Application du LRM sur une captation de l'exposition [Japon des avant-gardes 1910-1970 \(1986-1987\)](#) (schéma de Sarah Ambec, MAVEX, bibliothèque Kandinsky)

En deuxième figure, l'exemple appliqué est plus développé, indiquant cette fois le nombre réel de manifestations et d'items existants pour cette même œuvre. Chaque bobine est un item, le montage est une manifestation, la demande de production est une expression et l'œuvre est l'idée de cette captation, liée à l'œuvre exposition.

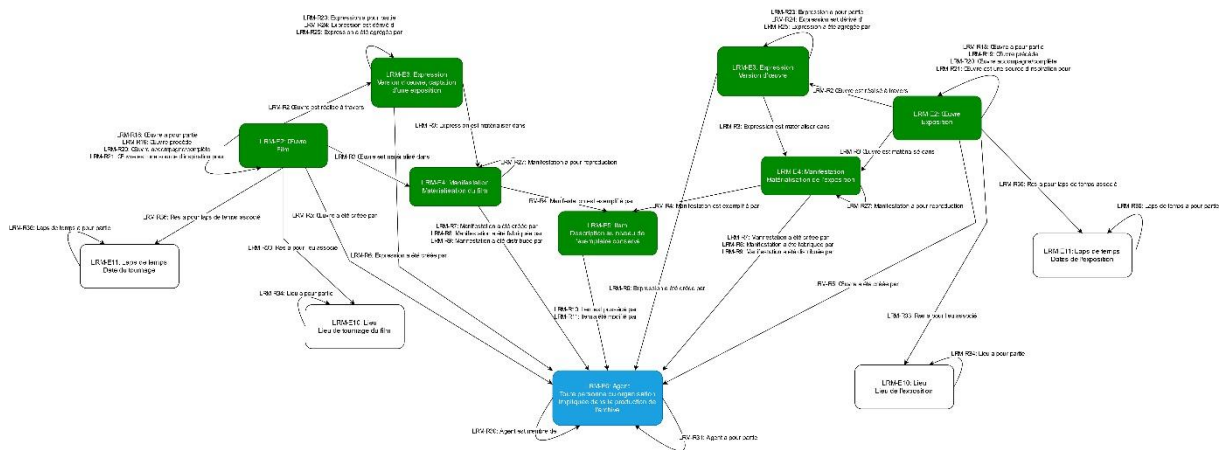


Figure 9 - Modèle LRM adapté aux archives audiovisuelles du Centre Pompidou (schéma de Sarah Ambec, MAVEX, bibliothèque Kandinsky)

Construire une nouvelle base de données relationnelle ?

Dans le cadre du projet MAVEX, nous nous posons la question de la préfiguration d'un outil de valorisation de ces archives, alors pourquoi pas modéliser ce à quoi pourrait ressembler cette base de données ? Et comment passer de la modélisation conceptuelle précédemment décrite à une structure de base de données relationnelle ? Inspirée par la structure de base de données proposée par la [FANA Danse et Arts Vivants](#) (en figure 10 ci-dessous)¹, je propose dans ce schéma de différencier le film et l'exposition, la captation et l'évènement, et de rassembler tous les agents dans une même table. Ce dernier modèle présenté en Figure 11, plus simple, nous permet de mettre en évidence les éléments principaux qui sont essentiels à la description de nos fonds.

¹ Voir Aurore Després, Sébastien Jacquot, « FANA - Danse & Arts vivants. Une fabrique de l'archive : conception et structuration de la plateforme » in *FANA Danse et Arts vivants*, Laboratoire ELLIADD EA 4661, Université de Franche-Comté, publication en ligne, 2018, <https://fanum.univ-fcomte.fr/fana/?p=1>, consulté le 21 janvier 2018 ; ainsi que Aurore Després, « Penser l'archive audiovisuelle pour la recherche en danse. Le Fonds d'Archives Numériques Audiovisuelles FANA Danse contemporaine », *Recherches en danse* [En ligne], no 5, mis en ligne le 15 décembre 2016, consulté le 08 octobre 2024. URL : <http://journals.openedition.org/danse/1307> ; DOI : <https://doi.org/10.4000/danse.1307>

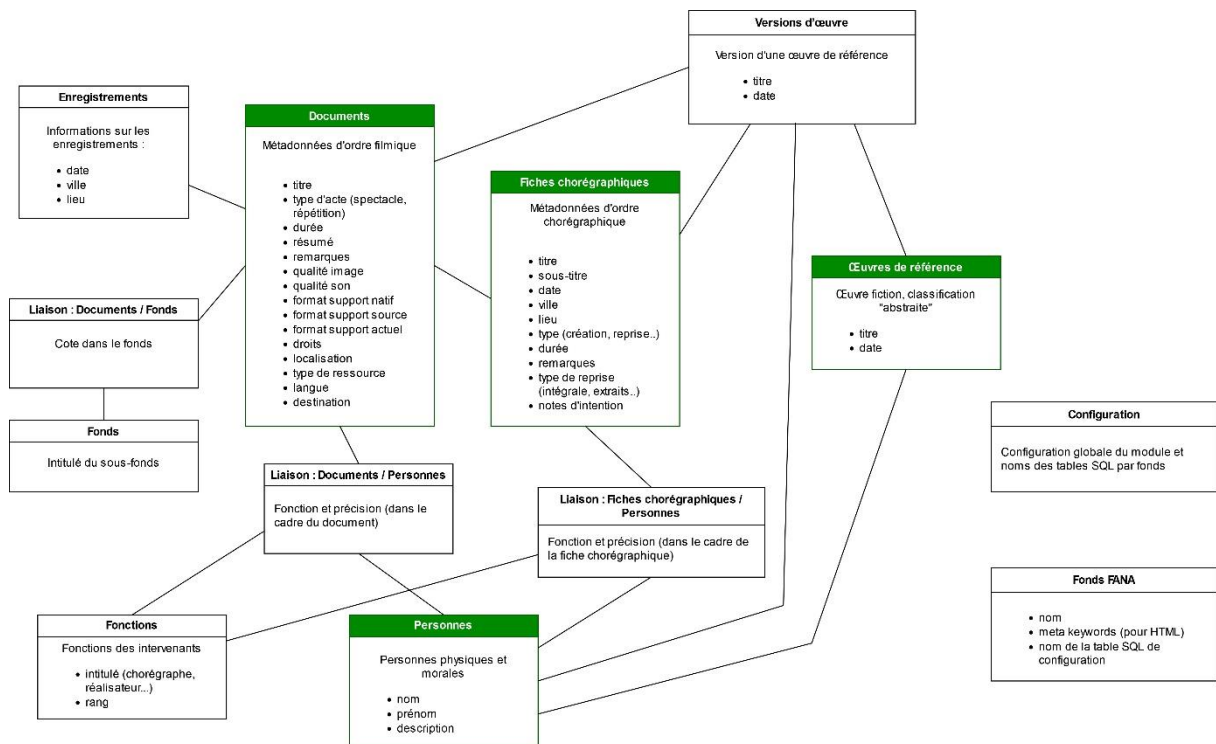


Figure 10 - Modélisation de la base de données SQL permettant la valorisation des archives audiovisuelles de la [FANA Danse et Arts vivants](#)

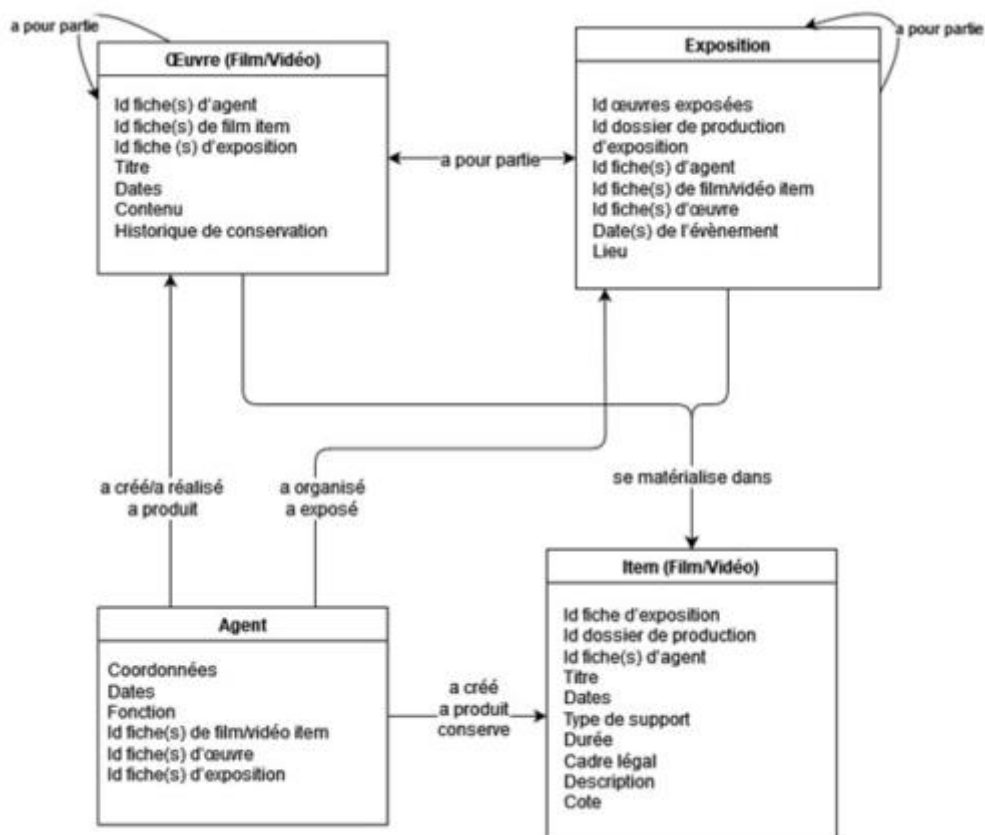


Figure 11 - Modélisation conceptuelle de la structure d'un outil de valorisation des archives audiovisuelles du Centre Pompidou (schéma de Sarah Ambec, MAVEX, bibliothèque Kandinsky)

Le *data mapping* pour envisager une migration

Le *data mapping* ou cartographie des données est le processus de mise en correspondance des champs de données d'une source avec les champs de données d'une autre source. Il permet de garantir que les données d'une source peuvent être transformées ou transférées avec précision et efficacité vers une autre destination tout en conservant leur intégrité, leur cohérence et leur signification. Nous sommes encore dans une étape de préfiguration de l'outil dans le cadre du projet MAVEX, ainsi est-il trop tôt pour réaliser un *data mapping*. Toutefois, une fois la documentation nécessaire à ce travail rassemblée, il s'agit d'une étape cruciale pour vérifier l'interopérabilité du modèle proposé et identifier les contraintes potentielles.

Conclusion

J'ai pu au cours de ce stage établir un premier constat des défis techniques que représente la préfiguration d'un outil en vue de la valorisation des archives audiovisuelles du Centre Pompidou. Après avoir passé en revue les différents modèles qui s'offrent à nous, nous n'avons pas proposé de choix entre la logique du musée, celle des archives et celle de la bibliothèque. Nous proposons un compromis, un consensus tel que doit l'être un outil qui est amené à être utilisé par des commissaires d'exposition, des archivistes, des historien·nes ou des artistes.

Il s'agit donc d'une première réponse, d'une ouverture, qui doit servir à la construction d'une solution plus pérenne, une fois le retour des usagers collecté par [le biais du questionnaire](#), une fois la documentation nécessaire au *data mapping* réunie, et une fois la décision d'un outil de valorisation appropriée actée. Cette phase exploratoire nous aura permis d'appréhender les défis qui se poseront alors et je l'espère, donnera de précieux outils pour y remédier.