

Valorisation des notices de l'inventaire d'une archive audiovisuelle en Linked Open Data : le cas du Montreux Jazz Digital Project

Travail de Bachelor réalisé

par : **Alain CHARDONNENS**

Mandant :

Alain DUFAUX, Directeur exécutif Metamedia Center (MMC)

Sous la direction de :

Arnaud GAUDINAT, Professeur HES

Basma MAKHLOUF SHABOU, Professeure HES

Carouge, le 12 juillet 2019

Information documentaire

Haute École de Gestion de Genève (HEG-GE)

Déclaration

Ce travail de Bachelor est réalisé dans le cadre de l'examen final de la Haute école de gestion de Genève, en vue de l'obtention du titre Bachelor of Science HES-SO en Information Documentaire.

L'étudiant atteste que son travail a été vérifié par un logiciel de détection de plagiat.

L'étudiant accepte, le cas échéant, la clause de confidentialité. L'utilisation des conclusions et recommandations formulées dans le travail de Bachelor, sans préjuger de leur valeur, n'engage ni la responsabilité de l'auteur, ni celle du conseiller au travail de Bachelor, du juré et de la HEG.

« J'atteste avoir réalisé seul le présent travail, sans avoir utilisé des sources autres que celles citées dans la bibliographie. »

Fait à Carouge, le 12 juillet 2019

Alain Chardonnens

Remerciements

Je remercie mon mandant le Prof. Dr. Alain Dufaux, directeur exécutif du Metamedia Center (MMC) de l'EPFL pour sa disponibilité et sa confiance, au cours de la réalisation de ce mandat. Je tiens tout particulièrement à remercier mon mandant d'avoir répondu positivement, à ma demande d'étude d'un jeu de données de la base de données du Montreux Jazz Digital Project (MJDP) et pour son partage de connaissances concernant l'historique et la technique audiovisuelle du fonds du MJDP.

Je remercie chaleureusement mes deux conseillers pédagogiques, le Prof. Dr. Arnaud Gaudinat et la Prof. Dr. Basma Makhoulf Shabou de m'avoir accordé leur confiance pour le choix de ce sujet. Je les remercie également pour leur bienveillance et pour les nombreux conseils prodigués tout au long de ce travail. Leurs enseignements et expertises respectives en data science et archivistique ont été une source de motivation et d'inspiration importante, au cours de ce travail.

Un grand merci à Olivier Bruchez du MMC pour son expertise informatique et technique de la base de données du Montreux Jazz Digital Project (MJDP).

Merci à l'ensemble de l'équipe actuelle du MMC pour son accueil chaleureux au cours de mon séjour : Alexandre, Candice, Charlotte, Sabrina et Vladimir.

Tout ma gratitude à mes parents et ma sœur pour leur soutien permanent et indéfectible, tout au long de ces quatre années d'études. Merci d'avoir rendu possible la poursuite de ce nouveau projet professionnel dans les meilleures conditions possibles.

Résumé

Le partage et la réutilisation des données liées aux notices d'inventaire, via les technologies du Linked Data (LD) et du Linked Open Data (LOD), émergent peu à peu dans le paysage de la description archivistique.

Ce travail de recherche propose un aperçu des parties prenantes du mouvement GLAM (galleries, libraries, archives, and museums) et de l'Open Data en Suisse. Le résultat est une cartographie des acteurs et des projets. L'étude des projets Wikidata appliqués aux GLAM a mis en perspectives les procédures et outils de transferts disponibles pour l'enrichissement des jeux de données.

Ensuite, l'étude du cas pratique des archives audiovisuelles du Montreux Jazz Digital Project (MJDP), a établi une liste de critères quantifiables et non quantifiables pour un versement vers une plateforme de crowdsourcing. L'évaluation pratique du fonds a utilisé 5 critères différents sur le 27 définis initialement. 5% des tables de la base de données relationnelle du MJDP ont été sélectionnées pour un transfert vers les plateformes Wikidata et opendata.swiss.

Les trois cas d'utilisation de OntoRefine ont été testés pratiquement lors de l'alignement d'un échantillon du MJDP vers Wikidata et d'autres bases de données musicales. Le cas *Choix entité et propriété* a permis de diminuer significativement le taux d'erreur au cours de la phase d'alignement des jeux de données.

La preuve de concept a mis en lumière l'adaptabilité du standard Record In Contexte (RiC), à partir d'une version simplifiée du schéma de données du MJDP.

Mots clés : Open Archives; Linked Open Data (LOD); Linked Data (LD); Crowdsourcing; GLAM; Archives audiovisuelles; Wikidata; RiC

Table des matières

Déclaration.....	i
Remerciements.....	ii
Résumé	iii
Liste des tableaux	vi
Liste des figures.....	vi
Liste des abréviations.....	vii
1. Introduction.....	1
2. États des lieux.....	3
2.1 Le Metamedia Center.....	3
2.2 La plateforme de crowdsourcing Wikidata.....	3
2.2.1 Procédure et outils pour accomplir un versement.....	5
2.2.2 Exemples de valorisation GLAM & Wikidata.....	7
2.3 Alignement d'une base de données relationnelle	8
3. Méthodologie	9
4. Open data et GLAM en Suisse.....	16
5. Ontologies applicables à une archive audiovisuelle.....	19
6. Bases de données musicales et vocabulaires contrôlés	21
7. Évaluation du fonds MJDP	22
7.1 Inventaire typologique des supports physiques	22
7.2 Inventaire typologique des enregistrements numériques	24
7.3 Inventaire typologique des données.....	26
7.4 Liste des critères	27
8. Adapter le modèle de données du MJDP aux linked data	29
8.1 Choix des entités, propriétés et relations	29
8.2 Validation du nouveau modèle conceptuel	30
9. Réalisation du nouveau modèle conceptuel.....	32
9.1 Diagramme de classe	32
9.2 Matrice d'adjacence.....	34
9.3 Fichier de triplet.....	34
10. Partager les données du MJDP en linked open data	37
10.1 Exploration et sélection des données externes sur Wikidata	37
10.1.1 Statistique sur les vocabulaires d'ontologies GLAM.....	37
10.1.2 Choix et sélection manuelle de propriétés.....	38
10.1.3 Validation du choix d'un identifiant externe.....	39
10.2 Préparation d'un versement vers Wikidata	40

10.2.1	Choix et mise en pratique des critères d'évaluation.....	40
10.2.2	Préparation du fichier d'export de données.....	43
10.3	Préparation d'un versement vers opendata.swiss	44
10.3.1	Choix et mise en pratique des critères d'évaluation.....	44
10.3.2	Préparation du fichier de description de données partagées.....	44
10.4	Alignement avec Wikidata et d'autres bases de données	45
11.	Conclusion	50
	Bibliographie	52
Annexe 1 :	« GLAM & Wikimedia » 2019 à Berne	61
Annexe 2 :	Wikidata:Events/Atelier 2019 à Lausanne	68
Annexe 3 :	Plan de recherches	71
Annexe 4 :	Workflow GLAM pour un versement	72
Annexe 5 :	La recherche d'information sur Wikidata	73
Annexe 6 :	Table de conversion de la largeur des coffrets.....	76
Annexe 7 :	Matrice d'adjacence des entités RiC	77
Annexe 8 :	Requête CONSTRUCT en SPARQL	78
Annexe 9 :	Sélection manuelle des propriétés Wikidata	80
Annexe 10 :	Requêtes SPARQL avec identifiants de type external-id..	82
Annexe 11 :	Distribution des propriétés MusicBrainz.....	84
Annexe 12 :	Modèle de versement opendata.swiss.....	85
Annexe 13 :	Exemple de fichier R2RML.....	87
Annexe 14 :	Diagramme du nouveau modèle conceptuel.....	89
Annexe 15 :	Entités de l'exemple liées à la plateforme Wikidata	90

Liste des tableaux

Tableau 1 : Liste des bases de données musicales	21
Tableau 2 : Inventaire typologique des supports physiques	23
Tableau 3 : Format des enregistrements numériques du fonds.....	26
Tableau 4 : Liste des critères quantifiables du fonds	27
Tableau 5 : Liste des critères non quantifiables du fonds	28
Tableau 6 : Structure de la matrice d'adjacence du graphe.....	34
Tableau 7 : Liste des identifiants de bases de données musicales	39
Tableau 8 : Extrait du résultat des valeurs des propriétés de type <i>external-id</i>	40
Tableau 9 : Intitulé des tables et colonnes du MJDP sélectionnées pour un transfert.....	42
Tableau 10 : Choix des propriétés sélectionnées sur Wikidata pour un transfert	43
Tableau 11 : Résultats de l'alignement des données du premier échantillon	46
Tableau 12: Résultats du cas pratique <i>Ajouter une propriété</i> pour la colonne artist_computed_public_name	47
Tableau 13 : Valeurs initiales du cas pratique <i>Choix entité et propriété</i>	47
Tableau 14 : Résultats du cas pratique <i>Ajouter une propriété</i> pour la colonne artist_computed_public_name	48

Liste des figures

Figure 1 : Conceptualisation du nouveau modèle	11
Figure 2 : Réalisation de la preuve de concept.....	13
Figure 3 : Requête SPARQL pour rechercher des ontologies sur Wikidata	14
Figure 4 : Cas pratiques d'alignement avec OntoRefine.....	15
Figure 5 : GLAM et Open Data en suisse (2019).....	18
Figure 6 : Ontologies applicables à une archives audiovisuelle.....	20
Figure 7 : Distribution des supports magnétiques du MJDP	24
Figure 8 : Exemple d'un enregistrement du MJDP adapter à RiC	31
Figure 9 : Diagramme de classe du nouveau modèle.....	33
Figure 10 : Extrait du fichier exporté de GraphDB	35
Figure 11 : Visualisation du graphe dans GraphDB	36
Figure 12 : Comparaison entre 7 vocabulaires d'ontologies	37
Figure 13 : Requête SPARQL pour afficher l'ensemble des propriétés d'un item	39
Figure 14 : Requête SPARQL avec des identifiants bases de données musicales.....	40
Figure 15 : Distribution des tables sélectionnées par catégories et tour de choix	41
Figure 16 : Équi-jointure des noms d'artistes et des instruments de musique.....	47

Liste des abréviations

API : Application programming interface

CSV : Comma-separated values

GLAM : Galleries, Libraries, Archives, and Museums

IRI : Internationalized Resource Identifier

JSON : JavaScript Object Notation

LD : Linked Data

LOD : Linked Open Data

MJDP : Montreux Jazz Digital Project

MJF : Montreux Jazz Festival

MMC : Metamedia centre

OWL : Web Ontology Language

R2RML : RDB to RDF Mapping Language

RDF : Resource Description Framework

RiC : Records in Contexts

SHACL : Shapes Constraint Language

SPARQL : Simple Protocol and RDF (Resource Description Framework) Query Language

SQL : Structured Query Language

Turtle : Terse RDF Triple Language

URI : Uniform Resource Identifier

URL : Uniform Resource Locator

XML : Extensible Markup Language

1. Introduction

Le présent travail de recherche a été réalisé dans le cadre du module Travail de bachelor (7061n) de la formation en information documentaire de la Haute école de Gestion de Genève (HEG-GE). Il propose d'exposer les différents résultats quantitatifs et qualitatifs obtenus, à la suite de la réalisation pratique du mandat établi en accord avec le MetaMedia Center (MMC) de L'École Polytechnique Fédérale de Lausanne (EPFL).

Au cours de la rédaction de ce travail le MMC a changé de dénomination, et se nomme à présent Cultural Heritage & Innovation Center (CHIC). Cependant, l'ancienne appellation a été conservée dans ce travail pour désigner cette entité.

Le cadre général de cette étude sont les archives numériques audiovisuelles, le crowdsourcing et les technologies du Linked Open Data (LOD) en Suisse.

Le périmètre spécifique de ce travail est le MMC de l'EPFL et le projet Montreux Jazz Digital Project (MJDP). Cette entité est responsable de la valorisation et de la conservation des archives audiovisuelles du Montreux Jazz Festival (MJF).

La **question principale de recherche** de ce travail est :

- de déterminer la faisabilité d'un export et les possibilités de synchronisation d'un jeu de données vers une plateforme de crowdsourcing.

Le développement de ce travail comprend neuf sections. L'objectif de la première section est d'établir une revue de la littérature exposant le contexte du MMC et de la plateforme de crowdsourcing Wikidata. Il est question d'explorer les procédures et outils mis à disposition par cette plateforme afin d'accomplir un versement. Des études de cas pratiques seront évoquées dans le but de présenter les perspectives de valorisation pouvant être envisagée. Les questions liées à l'alignement et la synchronisation des données avec Wikidata seront aussi discutées.

La seconde section de ce travail propose le détail des méthodologies envisagées et déployées, au cours de ce travail pour la collecte des résultats.

Les sections trois, quatre et cinq propose d'identifier les principaux acteurs responsables de l'Open Data dans le domaine des GLAM (galleries, libraries, archives, and museums) en Suisse, ainsi que les vocabulaires d'ontologie et bases de données musicales disponibles pour la description d'une archive audiovisuelle en Linked Data (LD).

La section six de ce travail concerne un état des lieux des enregistrements de la base de données relationnelle du MJDP, grâce à l'élaboration d'une série d'inventaires typologique comprenant les divers supports audiovisuels physiques et numériques du fonds. Ces inventaires seront complétés par une analyse du modèle de données du MJDP, pour enfin aboutir à la proposition d'une liste de critères quantifiables et non quantifiables pour l'évaluation.

La septième section démontre l'adaptabilité du modèle de description archivistique Record In Contexte (RiC) édité par le Comité International des Archives (CIA), à une base de données relationnelle non développée autour des standards de description usuelle adoptés par la communauté archivistique internationale. Cette section étudiera la possibilité de concevoir un modèle conceptuel simplifié en alignant les entités du MJDP avec le standard de description RiC.

La huitième section de ce travail apportera des pistes de réflexion, quant à la réalisation pratique du modèle conceptuel esquissé dans la septième section. En d'autres termes, est-il possible de réaliser une preuve de concept fonctionnelle, à partir du modèle conceptuel proposé ?

La dernière section de ce travail s'interroge principalement sur l'étude de la faisabilité du transfert d'un jeu de données. Cette section inclut également une comparaison entre les différents vocabulaires d'ontologie découverts dans la section 5 et leurs distributions dans Wikidata. Dans quelles mesures les critères définis sont-ils pertinents pour accomplir cette évaluation pratique ? L'exercice sera réalisé pour les plateformes Wikidata et opendata.swiss. Enfin, le chapitre 10.4 décrira les cas pratiques découverts pour l'alignement d'un échantillon de données avec OntoRefine.

Le concept défini par les termes alignement, mapping et reconciliation désignent des concepts similaires dans la littérature. Dans ce travail, le terme alignement a été préféré aux termes mapping et reconciliation.

2. États des lieux

2.1 Le Metamedia Center

Le Metamedia Center (MMC) de l'Ecole Polytechnique Fédérale de Lausanne (EPFL) est un centre multidisciplinaire, créé en 2010. Il est rattaché à la Vice-Présidence pour l'Innovation (VPI), principal lien de la haute école avec les milieux industriels. Depuis sa création, il poursuit le mandat de numériser, gérer et valoriser les archives du Montreux Jazz Festival (MJF).

L'objectif principal du MMC est d'encourager le développement de projets de recherches scientifiques et l'application de technologies innovantes, dans le cadre du projet Montreux Jazz Digital Project. Depuis l'ouverture du centre, plus de 35 projets ont été réalisés, en collaboration avec différents laboratoires de l'EPFL, des partenaires privés ou encore d'autres universités.

Parallèlement aux projets d'innovations, des travaux importants de conservations et de descriptions archivistiques ont été entrepris, depuis 2010. En effet, des campagnes successives d'inventaires, de descriptions et d'indexations de supports analogiques et numériques ont été réalisés par le MMC.

Depuis la création du festival en 1968, le fonds regroupe actuellement environ 14'000 bandes magnétiques, réparti en 18 formats différents représentant un total de plus de 11'000 heures d'enregistrement vidéo et audio sur les 5000 heures de concert initiale (Dufaux, Amsallem 2017). Le fonds est également constitué d'une série de plus de 100'000 négatifs photographiques.

En 2013, suite à la disparition de Claude Nobs, le fondateur du Montreux Jazz Festival (MJF), une fondation est créée pour faciliter la gestion de l'archive. Thierry Amsallem devient le président de la fondation Claude Nobs. L'archive est inscrite, la même année, au registre de la Mémoire du monde de l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO).

2.2 La plateforme de crowdsourcing Wikidata

Wikidata est une plateforme de crowdsourcing, créée en 2012 par Denny Vrandečić et la communauté du chapitre allemand de la fondation Wikimedia. L'objectif principal de la plateforme est d'offrir un répertoire centralisé des données pour les projets de Wikimedia. Selon Denny Vrandečić (Vrandečić 2013, p.90), les principales caractéristiques du projet Wikidata comprennent :

- Un accès libre aux données partagée facilitant la diffusion et réutilisation sous licence Creative Commons CC0.
- Une édition libre et participative gérée par les membres de la communauté d'utilisateurs de la plateforme.
- La création, l'édition, la recherche et la réutilisation de données multilingue.
- La possibilité de conserver une ambiguïté de sens des concepts décrit sur la plateforme.
- La réutilisation des données grâce à l'utilisation de données structurées.
- La centralisation et l'homogénéisation des concepts décrits dans les autres projets de la fondation Wikimedia.

- Un répertoire de données vérifiées pour les différents projets de Wikimedia.

Les logiciels open source MediaWiki et Wikibase sont à la base de l'infrastructure technique de la plateforme Wikidata (Thornton 2017, p.2). En effet, MediaWiki est le moteur principal des pages des différents projets soutenus par la fondation Wikimedia. Ce logiciel écrit en langage de programmation Java a servi de modèle au développement du logiciel Wikibase (Wikibase 2019). Le logiciel Wikibase se compose principalement de deux extensions du logiciel MediaWiki : Wikibase Repository (Extension:Wikibase Repository 2019) et Wikibase Client (Extension:Wikibase Client 2019). Le Wikibase Repository offre la possibilité de gérer un répertoire collaboratif ou wiki comprenant des données structurées. Le Wikibase Client est une interface, qui offre la possibilité de naviguer et de gérer un Wikibase Repository. Cette interface permet notamment la gestion du multilinguisme dans les articles de Wikidata (Wikibase 2019).

Le modèle physique de données de Wikibase (Wikibase/DataModel/Primer 2019) est composé des deux entités principales : *Item* et *Property*. L'entité *Item* correspond à une page concernant la description d'un concept sur le Wikibase Repository. Un *Item* est composé d'un identifiant unique introduit de la manière suivante : Qxxxx. Cet identifiant unique peut être décrit par un ou plusieurs *Statements*. Les *Statements* sont des catégories regroupant une propriété également identifiée par un identifiant unique et décrit sous la forme : Pxxxx. Une propriété comporte une ou plusieurs *Values* identifié par un identifiant unique défini par un autre *Item* et référencé dans la plateforme Wikidata. Si la valeur n'est pas encore existante, elle peut être entrée directement par un utilisateur sur la page de l'*Item* correspondant.

Une *Values* peut être définie par un *qualifiers*, qui peut être décrit à son tour par une ou plusieurs propriétés composées à chaque fois d'une *Values*. Une *references* est définie par une ou plusieurs *Values* introduites par une propriété construite suivant la même structure qu'un *Statements*. La qualité de l'information référencée peut être indiquée par *rank*, qui offre la possibilité de classer l'information décrite en fonction du sens le plus souvent admis pour un concept décrit par la communauté d'utilisateur de Wikidata. Une *property*, une *value* et un *qualifiers* peuvent être regroupés sous l'appellation *claim* (Wikibase/DataModel/Primer 2019).

L'ontologie de Wikibase est disponible en langage RDF (Wikibase/Indexing/RDF Dump Format 2019). Cependant, une sérialisation simplifiée de cette ontologie en langage JSON (Wikibase/DataModel/JSON 2019) est également disponible. L'ontologie RDF est employée principalement par le Wikidata Toolkit, une librairie écrite en langage Java, qui permet d'extraire des données de sauvegardes (dump) de l'ensemble des plateformes de la fondation Wikimedia (Wikimedia sans date d). Quant à la sérialisation en langage JSON, elle est principalement utilisée par l'application programming interface (API) de Wikimedia.

Une collection d'outils développée pour éditer, rechercher, visualiser et exploiter les données décrites sur Wikidata est accessible, depuis la page Tools de la plateforme (Wikidata:Tools 2019). Par ailleurs, la plateforme Wikimedia Toolforge (Wikimedia sans date g) propose, via l'utilisation du répertoire Tools Directory, la possibilité de rechercher des outils développés pour Wikidata. La plateforme Phabricator (Wikimedia sans date b) de la fondation Wikimedia permet entre autres d'assurer le suivi du développement collaboratif du modèle de données et des outils disponibles sur Wikimedia Toolforge. Cette plateforme offre à la communauté de développeur, un espace collaboratif pour le développement et la documentation des outils.

2.2.1 Procédure et outils pour accomplir un versement

La plateforme Wikimedia Toolforge propose une série d'outils spécifiques permettant d'accomplir les tâches liées au chargement, au nettoyage et à l'indexation des données, en vue d'un versement sur Wikidata.

Un workflow (voir annexe 4) proposé par la communauté GLAM est accessible sur la plateforme Wikimedia Outreach et documente les principales étapes d'un versement vers une des plateformes de la fondation Wikimédia (GLAM/Resources/Data and media partnerships workflow 2019).

Ce processus est découpé en quatre lots de tâches. Il s'agit des étapes intitulées *prés-chargement*, *chargement*, *après-chargement* et *impact*. Cependant, deux conditions préalables doivent être remplies pour déclencher la mise en œuvre du processus. En effet, une convention doit être conclue entre l'institution patrimoniale et les responsables d'un chapitre de Wikimédia. La deuxième condition préalable concerne les questions liées à l'accès aux données de l'institution culturelle elle-même (droit, ressources humaines et matérielles disponibles).

Si, les données ne sont pas directement exploitables, elles peuvent être extraites à l'aide des outils Tabula (Aristarán 2018) et/ou PAWS (PAWS 2019). Le premier de ces deux outils permet l'extraction des données disponibles sous la forme de tableau dans un fichier au format PDF. Le deuxième outil offre la perspective d'accomplir des exports directement à partir d'une base de données relationnelle comme MySQL. PAWS, un modèle de notebook basé sur le logiciel Jupyter Notebook (Project Jupyter 2019) permet d'effectuer ces opérations d'extraction, de transformation et de chargement (ETL), depuis une interface software as a service (SaaS).

La première étape du processus vise à clarifier les droits relatifs aux données partagées. Pour un versement sur Wikidata, les données devront respecter la licence CC0 (Creative Commons sans date a) des Creative Commons. Cette licence stipule que les données partagées sont libres de droits et réutilisables dans le domaine public. Le deuxième point de cette étape concerne l'évaluation des données de la base de données existante. Une liste des précédents projets (Wikidata:WikiProjects 2019) développé et soutenu par la communauté Wikidata offre une vue d'ensemble des possibilités offertes par la plateforme.

Le prochain point de cette étape est une comparaison des données locales avec celles déjà référencées sur Wikidata. Si les données partagées sont déjà existantes, il est conseillé de réutiliser les valeurs existantes plutôt que de créer de nouvelles entrées similaires. Les outils proposés pour ce point sont les moteurs de recherche PetScan (Wikimedia sans date a) et le Query Service. PetScan est une interface de recherche avancée, qui permet d'accomplir des recherches fédérées sur l'ensemble des plateformes de Wikimédia. Par ailleurs, l'annexe 5 propose une vue exhaustive des diverses possibilités de recherche d'information disponible sur la plateforme Wikidata.

Le dernier point de cette étape est la réconciliation ou alignement des données partagées avec Wikidata. Ce point peut être réalisé en ligne, à l'aide de Mix'n'match un outil manuel de réconciliation. Pour le traitement de larges jeux de données, des outils comme OpenRefine, OntoRefine et/ou le développement d'une solution spécifique seront à privilégier.

Selon Zeng, l'outil Mix'n'Match constitue une source importante et fiable de notices biographiques.

« As a tool, Mix'n'Match lists entries of hundreds of external databases in a variety of categories and allows volunteers to manually match them against Wikidata items. An exceptional feature of this resource is the number and variety external datasets: for example, dozens in the Heritage category and over 500 in the Biography category are all sourced from different countries. »
(Zeng 2019, p.12)

De Boer confirme également la découverte de bonne occurrence avec ce même outil utilisé dans le cadre du développement d'un preuve de concept pour le projet Common Thesaurus for Audiovisual Archives (GTAA) du Netherlands Institute for Sound and Vision (NISV).

« This tool [Mix'n'Match] was used to automatically match person entities based on first and last names. In total, 10,350 GTAA person concepts are linked to Wikidata entries. »
(De Boer 2019, p.214)

La deuxième étape prévue par le processus est le *chargement* des données. Il est conseillé pour cette étape d'effectuer des versements en plusieurs fois afin d'éviter les éventuelles erreurs liées à la structure des données. Les outils manuels mis à disposition pour le versement sont principalement QuickStatements (Wikimedia toolforge sans date e) et OpenRefine.

QuickStatements offre l'opportunité de copier directement des données en format csv, à partir d'un navigateur. Ce mode de versement convient principalement pour un volume de données limité. Pour des versements importants ou automatiques, il est possible de combiner le notebook PAWS avec le Pywikibot pour écrire un script en langage python. Cependant, une demande (Wikidata:Requests for permissions/Bot 2019) préalable doit être formulée auprès de la communauté de Wikidata, avant d'effectuer pratiquement le transfert vers la plateforme.

La troisième étape de la procédure intervient, à la suite du chargement des données sur la plateforme. Le premier point concerne la vérification de la structure des données téléchargée sur Wikidata. Pour ce faire, la communauté a mis à disposition l'outil EditGroups¹, qui permet de supprimer ou corriger un lot de données versé par erreur avec les outils QuickStatements ou encore OpenRefine.

Le deuxième point concerne l'enrichissement des métadonnées issues des données téléchargées sur la plateforme par un utilisateur en local ou alors par la participation des membres de la communauté Wikidata (crowdsourcing).

La dernière étape de cette procédure concerne la mesure de l'impact des données partagée sur la plateforme. Les outils BaGLAMa 2 (Wikimedia toolforge sans date a) et GLAMorgan (Wikimedia toolforge sans date b) ont été développés pour permettre un suivi du nombre des consultations d'une page spécifique ou d'une thématique précise. La Wikimedia REST API (Wikimedia sans date f) permet également d'accomplir cette mesure de la fréquentation des pages à distance.

¹ Les jeux de données téléchargé par erreur peuvent être modifié depuis l'interface EditGroups : <https://tools.wmflabs.org/editgroups/>

Enfin, une dernière alternative propose de développer dans le Wikidata Query Service une requête SPARQL spécifique, incluant les identifiants des *items* partagés pour pouvoir suivre les évolutions dans le temps.

2.2.2 Exemples de valorisation GLAM & Wikidata

Les projets *WikiProject* Corfu (Wikimedia toolforge sans date g) et Crotos (Deshayes 2019) sont deux exemples de valorisation par crowdsourcing des données issues d'institutions GLAM. Le premier exemple permet d'annoter et de géo-référencé des objets culturels sur une carte interactive. Le second propose un moteur de recherche permettant d'afficher près de 179'000 œuvres d'art. Ces deux projets utilisent simultanément les plateformes Commons et Wikidata.

Le cas pratique (Wikidata:WikiProject Performing arts/Reports/Ingesting Production Databases of the Performing Arts 2019) documente une procédure d'alignement des données du catalogue de la Swiss Theatre Collection vers la plateforme Wikidata. Ce projet a permis la création d'une nouvelle ontologie spécifique aux besoins des archives théâtrales et de la danse en Suisse. Il s'inscrit dans l'initiative internationale Wikidata:WikiProject Performing arts, qui poursuit l'objectif de partager et valoriser les données issus des domaines de la danse et du théâtre sur Wikidata.

Allison-Cassin et Scott décrivent le développement d'une preuve de concept dans le but d'enrichir les notices du catalogue de la bibliothèque de l'université de Laurentian au Canada (Allison-Cassin, Scott 2018). Les données du catalogue alignées avec Wikidata concernent des informations relatives à des musiciens canadiens. Le code source de cette preuve de concept est open source et téléchargeable à partir d'un répertoire Github.

De Boer décrit le développement d'une preuve de concept dans le cadre du projet Common Thesaurus for Audiovisual Archives (GTAA) du Netherlands Institute for Sound and Vision (NISV) (De Boer 2019). Le projet propose un alignement des données du thésaurus de cette institution avec les données accessibles, depuis Wikidata.

Van Veen partage également l'idée d'utiliser Wikidata comme une plateforme centrale en utilisant directement les identifiants de la plateforme comme identifiant unique (Van Veen 2019). Selon Van Veen, le choix de l'identifiant Wikidata permet de faciliter significativement l'alignement entre un catalogue institutionnel et la plateforme :

« Using the Wikidata identifier as resource identifier makes Wikidata the first hub. Applications therefore have in the first instance to deal with only one description model. From there, it is easy to navigate further: most information is only “one hub away,” so less prior knowledge is required to link from one source to another. »

(Van Veen 2019, p.76)

Cooley insiste sur la complémentarité des données fournies par Wikidata dans l'enrichissement des notices du catalogue dans le cadre du projet European Holocaust Research Infrastructure (EHRI) :

« By contributing their data to Wikidata, leaders in thesaurus management and authority control can improve their local metadata while also expanding the reach of that data. »

(Cooley 2019, p.96)

2.3 Alignement d'une base de données relationnelle

Le World Wide Web Consortium (W3C) propose des exemples pratiques d'alignement entre une base de données relationnelle et une base de données de graphe (World Wide Web Consortium 2010). Le *Direct Mapping* est une transposition à l'identique du modèle relationnel existant dans un modèle de graphe.

Le *Database to Ontology Mapping* permet d'adapter le modèle relationnel existant en l'alignant directement à un nouveau modèle d'ontologie. Ce modèle d'ontologie peut être défini par l'organisation elle-même ou alors adapté, à partir d'une ou plusieurs ontologies externes.

Les trois cas pratiques proposés par le W3C sont construits autour du langage de programmation R2RML (World Wide Web Consortium 2012b). Ce langage permet notamment d'intégrer des requêtes en langage SQL afin de faciliter la sélection des tables et la construction du modèle de graphe en RDF. Cependant, l'alignement de données hétérogènes est difficilement réalisable en R2RML. Cet alignement peut être effectué grâce au RDF Mapping Language (RML) (Dimou et al. 2014).

Le D2RQ Mapping Language (Freie Universität Berlin 2012) permet d'établir une connexion directe entre une base de données relationnelle et R2RML. La combinaison des langages R2RML, RML et D2RQ proposent une solution pratique de synchronisation des données entre une base de données relationnelle et un modèle de graphe. Schiavone présente une approche d'alignement de données hétérogènes utilisant RML avec la plateforme Wikidata (Schiavone et al. 2018).

Un exemple récent d'alignement a été expérimenté, puis documenté dans la littérature avec le modèle de données CIDOC-CRM et diverses ontologies de domaine (Hyvönen 2016b, p.3). Hyvönen a identifié trois étapes différentes pour la réussite d'un alignement de données culturelles. Il s'agit du *Domain neutral semantic model*, du *Metadata alignment model* et du *Shared domain ontologies*.

« 1. *Domain neutral semantic model. Agreement of using domain neutral Semantic Web standards of W3C, such as RDF(S), SKOS, and OWL.* / 2. *Metadata alignment model. Agreement on using a metadata alignment model for harmonizing the different metadata models used by different partners.* / 3. *Shared domain ontologies. Agreement of sharing domain ontologies (places, persons, etc.) whose concepts are used for populating the metadata models.* »
(Hyvönen 2016a, p.3)

Deux autres prototypes construits autour du standard RiC ont été récemment mis en ligne. Il s'agit des projets de plateforme aLod (Archives Fédérales Suisses et al. 2014) proposé par les Archives fédérales suisses (AFS) et du Pilote d'interopérabilité pour les autorités archivistiques françaises (PIAFF) (International Council on Archives 2018) des Archives nationales en France. Ces deux projets poursuivent le même objectif de tester la faisabilité d'un alignement des notices en Linked Data (LD) de catalogues de divers dépôts d'archives.

3. Méthodologie

La section méthodologique de ce travail présente le détail des diverses étapes et méthodes, qui ont permis d'établir les résultats et les conclusions présentées dans ce travail de recherche.

Open data et GLAM en Suisse

Les recherches effectuées dans le chapitre 4 ont été menées, à la suite de l'élaboration d'un plan de recherche (voir annexe 3). Le plan de recherche a permis d'orienter les recherches d'informations, en dégagant les principaux axes et sous-axes de recherche.

Les sources d'informations consultées proviennent principalement de site internet lié à des organisations et/ou associations œuvrant pour la promotion de l'Open Data en Suisse. La participation à une journée professionnelle « GLAM et Wikimedia »² organisée par l'Association des archivistes suisses (AAS) et Wikimedia CH a constitué une source de premier choix pour orienter les futures recherches d'informations exposées dans le chapitre 4.

Ontologies applicables à une archive audiovisuelle

Dans le chapitre 5, les principales sources d'informations consultées ont été les standards et recommandations proposées par diverses associations professionnelles suisses et internationales.

Bases de données musicales et vocabulaire contrôlé

Une recherche d'information sur Wikipédia, a permis la découverte de trois différentes listes (Comparison of on-demand music streaming services 2019 ; List of online digital musical document libraries 2019 ; List of online music databases 2019) répertoriant des bases de données musicales commerciales et des projets open source. Ces trois listes ont constitué la principale source d'information pour établir un état des lieux des bases de données musicales disponibles et accessibles en libre accès dans le chapitre 6.

Une deuxième source d'information est le projet WikiProject Music (Wikidata:WikiProject Music 2019), qui documente les bases de données musicales alignées avec la plateforme Wikidata.

Inventaire typologique des supports physiques

L'estimation du mètre linéaire proposée dans le chapitre 7.1 a été obtenue en relevant les dimensions de la largeur des coffrets décrite sur le site Preservation Self-Assessment Program (PSAP) (University of Illinois sans date). Ce site a consigné pour chaque support, les dimensions en pouce des coffrets contenant des bobines ouvertes ou des cassettes magnétiques. Les conversions entre les mesures en pouce et en mètre ont été effectuées avec le moteur de recherche WolframAlpha.

² Un compte rendu détaillé de cette journée est disponible dans les annexes 1 et 2 de ce travail.

Inventaire typologique des données

Dans le chapitre 7.3, l'interface *phpMyAdmin* de la base de données relationnelles MySQL à faciliter l'exploration des intitulés des tables et des colonnes du MJDP. L'option *Data dictionary* de *phpMyAdmin* permet d'afficher une vue d'ensemble des tables et des colonnes de la base de données du MJDP. Cette option permet également d'afficher les différents types de données, ainsi que les clés primaires et étrangères comprises dans le modèle de données du MJDP.

La deuxième étape de ce travail comprend l'export des intitulés des tables et des colonnes. Cet export est réalisé, à partir de la fonction export de l'interface *phpMyAdmin*. Le fichier cible obtenu est un fichier en langage SQL. À la suite de l'enregistrement du fichier SQL, un éditeur de texte évolué et une expression régulière ont été nécessaires pour la recherche de motifs spécifique sur les chaînes de caractères contenus dans le code SQL. Les résultats extraits des motifs ont permis de copier les tables et les colonnes du fichier SQL dans un tableau Excel.

Adapter le modèle de données du MJDP aux linked data

Le processus décrit sur la figure 1 s'inspire du cas pratique *Database to Ontology Mapping* du W3C discuté dans le chapitre 2.2. Ce processus comprend également la création d'un diagramme afin de créer une représentation graphique des concepts, comme le suggère la phase *Ideation* de la méthodologie *Design thinking*. (Brown 2012, p.5)

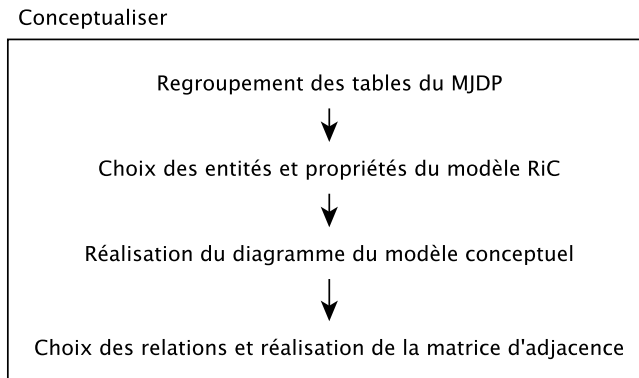
La première étape du processus décrit dans la figure 1 prévoit le regroupement thématique des tables du schéma physique de la base de données relationnelles. Le but est de proposer des intitulés génériques aux différentes entités du nouveau modèle. La principale source d'information pour réaliser ce regroupement est l'étude de la structure hiérarchique et intellectuelle de l'information des différentes interfaces graphiques de la base de données du MJDP.

La deuxième étape propose le choix et l'alignement des intitulés des nouvelles entités et de leurs propriétés avec les spécifications du modèle conceptuel RiC. Une matrice d'adjacence est créée dans le but d'offrir une vision synthétique des entités choisies en fonction de l'ensemble des possibilités décrites dans le standard RiC. (voir annexe 7)

La troisième étape est la réalisation graphique du diagramme du nouveau modèle de données conceptuelles. D'abord, les entités sont reportées de la matrice d'adjacence vers le diagramme. La matrice d'adjacence est utilisée en informatique pour structurer les données d'un graphe fini (Matrice d'adjacence 2019). Ensuite, le sens des relations entre les différentes entités est défini. Enfin, les valeurs des entités correspondantes au modèle RiC sont retranscrites sur le diagramme.

La quatrième étape prévoit le choix des relations RiC en fonction du sens entre les différentes entités. Ce choix est défini par le sens des relations et par le type des entités décrites sur le diagramme.

Figure 1 : Conceptualisation du nouveau modèle



Validation du nouveau modèle conceptuel

Le chapitre 8.2 prévoit la construction d'un prototype afin de valider théoriquement le nouveau modèle conceptuel de la base de données du MJDP. Ce prototype sera construit, à partir d'un enregistrement présent dans le schéma physique de la base relationnelle du MJDP.

L'échantillon est sélectionné à partir d'un enregistrement disponible sur l'interface graphique *play* de la base de données du MJDP. L'intitulé des valeurs de l'échantillon est ensuite reporté sur un diagramme et placé dans les entités et les relations définies précédemment par le nouveau modèle.

Réalisation du nouveau modèle conceptuel

La figure 7 propose une vue d'ensemble du processus complet, initié dans le chapitre 8 avec la figure 1. Le premier lot de tâches *Conceptualiser* concerne uniquement le chapitre 8 de ce travail.

Le deuxième lot de tâches *Réaliser* prévoit la réalisation pratique du modèle conceptuel. La première étape consiste à convertir les entités et les relations illustrée dans le diagramme du modèle conceptuel vers un diagramme de classe (figure 8). Le but est de créer un modèle physique à partir du nouveau modèle conceptuel (annexe 14). A cette occasion les quatre propriétés RiC mentionnées au début du chapitre 8 seront ajoutées à ce diagramme.

La deuxième étape consiste à compléter manuellement la matrice d'adjacence (voir annexe 7) avec les valeurs des propriétés et des relations définies sur le diagramme de la figure 7. Une fois la matrice remplie avec les valeurs issues de l'enregistrement, le résultat de cette partie est exporté dans un format de fichier interopérable.

La troisième étape de ce deuxième lot de tâches est la création d'un nouveau répertoire en prévision de l'import des données dans le logiciel GraphDB. Le menu *Repositories* dans *Setup* permet la création de ce nouveau répertoire dans GraphDB. Une fois la création du nouveau répertoire effectuée, le fichier de la matrice d'adjacence peut être importé dans la base de données GraphDB via l'interface OntoRefine.

La quatrième étape consiste à l'import des données de la matrice d'adjacence, à partir de l'interface OntoRefine (Ontotext sans date). Une fois le chargement du fichier et la sélection des colonnes effectués sur l'interface, un nouveau projet peut être créé.

La cinquième étape est effective à la suite de l'import des données et consiste au référencement des colonnes du tableau par l'attribution d'un Internationalized Resource Identifier (IRI). La valeur de l'identifiant unique IRI peut être modifiée à partir du menu *RDF*, en sélectionnant *RDF settings*.

À la suite de la configuration du IRI, la cinquième étape comprend la création automatique de deux requêtes en langage SPARQL afin de faciliter la sélection des colonnes et la construction du graphe. OntoRefine permet d'accomplir ces deux opérations en naviguant dans le menu *Data* et en sélectionnant *Generate SELECT* pour la construction automatique de cette requête de sélection. La requête de construction du graphe est activée en sélectionnant *Generate CONSTRUCT*.

Une fois les requêtes en langage SPARQL générées, elles peuvent être exécutées directement dans l'interface en sélectionnant *Run*. L'affichage des résultats de ces requêtes est à la fois disponible sous la forme d'un tableau ou en langage JSON. Il est également possible d'effectuer des filtres en sélectionnant les entêtes des colonnes ou encore d'afficher diverses statistiques.

L'étape suivante est la modification de la requête en langage SPARQL, à partir du modèle généré automatiquement par *Generate CONSTRUCT*. L'interface OntoRefine permet de tester pas à pas, la progression de la construction de la requête SPARQL et d'afficher le résultat sur le graphe. À la suite de cette démarche itérative, les données sont prêtes à être ajoutées au répertoire de la base de données GraphDB.

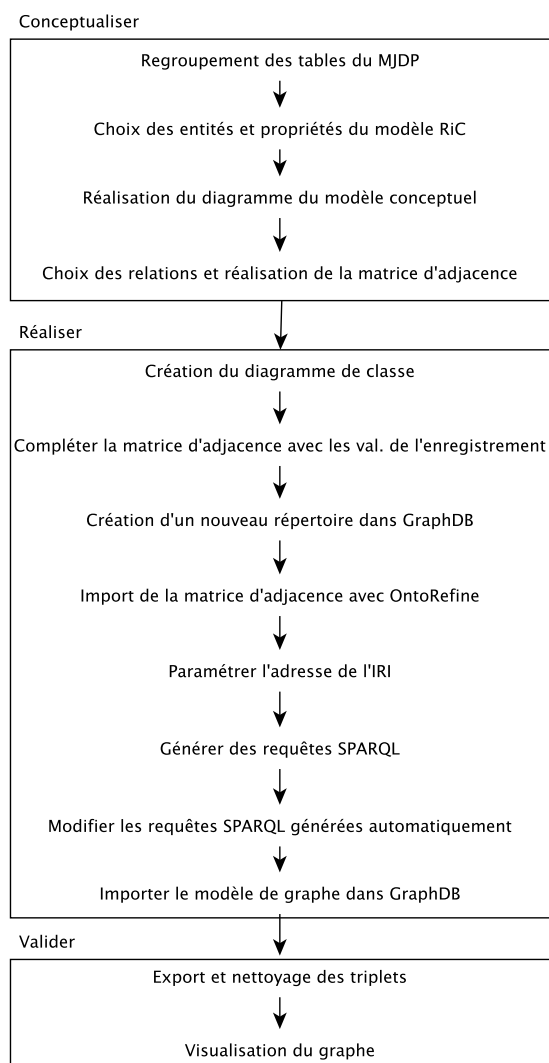
L'ultime étape de ce lot de tâches *Importer le modèle de graphe dans GaphDB* prévoit le chargement des données dans le répertoire créé, à la suite de la deuxième étape dans GaphDB. L'adresse url du *endpoint* de l'interface OntoRefine doit être relevée dans le menu *Data* en sélectionnant *Get SPARQL endpoint...*

A ce stade, l'instruction CONSTRUCT de la requête SPARQL doit être remplacée par INSERT et l'instruction SPARQL SERVICE ajoutée dans le code de la requête. Le livrable obtenu à la fin de ce lot de tâches est un fichier brut encodé en format N-Triples (figure 10).

Le dernier lot de tâches *Valider* prévoit l'export du fichier brut en format N-Triples. Ce fichier est généré à la suite de la requête INSERT. Une fois le fichier d'export téléchargé, il est vérifié et au besoin nettoyé à l'aide d'un éditeur de texte évolué et l'utilisation d'une expression régulière.

La dernière étape de ce lot de tâches est l'import du fichier nettoyé dans le répertoire de la base de données GraphDB. Le menu *Explore* permet d'observer le résultat des propriétés et des relations dans le graphe.

Figure 2 : Réalisation de la preuve de concept



Statistique sur les vocabulaires d'ontologies GLAM

Cette mesure du chapitre 10.1.1 a été effectuée à partir d'un échantillon des résultats obtenus dans le chapitre 5. Le choix des sept vocabulaires sélectionnés dans l'échantillon sont principalement issus des catégories : *spécifique à l'audiovisuelle* et *spécifique à la documentation* décrite dans la figure 6. Cependant, les vocabulaires d'ontologie RELATIONSHIP (rel) (Vocab.org 2010) décrivant les relations entre deux individus et Music Ontology (mo) (Raimond et al. 2013) relatif à des concepts liés au domaine musical ont été ajoutés à cet échantillon.

Chaque vocabulaire d'ontologie a été recherché séparément sur le répertoire Linked Open Vocabulary (LOV) et les résultats ont permis d'extraire des informations concernant le nombre de propriétés compris dans chaque vocabulaire. Le répertoire LOV permet également d'observer visuellement le nombre de liens entrant et sortant d'un vocabulaire d'ontologie en particulier. Selon Vandebussche, le répertoire LOV est un répertoire de référence pour la recherche de vocabulaires d'ontologie.

« 1) the availability of a high quality dataset of vocabularies available through multiple access methods 2) the curation by experts, making explicit for the first time the relationships between vocabularies and their version history and 3) the consideration of property semantics in term search relevance scoring.» (Vandenbussche 2014, p.14)

La première mesure effectuée sur la plateforme Wikidata a permis de comparer les vocabulaires d'ontologie possédant une occurrence dans les plateformes LOV et Wikidata. La requête SPARQL utilisée pour cette mesure est construite sur la base d'une requête proposée dans le projet Ontology de Wikidata (Wikidata:WikiProject Ontology/Mapping 2019).

Les termes sélectionnés pour la recherche proviennent de l'échantillon des vocabulaires d'ontologies et ont fait l'objet de recherche individuelle. En effet, une nouvelle requête SPARQL a été créée, puis exécutée individuellement pour chaque terme de l'échantillon. Les requêtes ont été exécutées sur l'interface du Wikidata Query Service de la plateforme Wikidata. La requête SPARQL construite afin d'accomplir ces recherches est décrite dans la figure 3.

Figure 3 : Requête SPARQL pour rechercher des ontologies sur Wikidata

```
SELECT DISTINCT * # Affiche les valeurs uniques dans
l'ensemble des colonnes du tableau de résultats
WHERE {
    {?p wdt:P2235 ?s. } # P2235: super-propriété externe
    UNION {?p wdt:P2236 ?s. } # P2236: sous-propriété externe
    UNION {?p wdt:P1628 ?s. } # P1628: propriété équivalente
    UNION {?p wdt:P1709 ?s. } # P1709: classe équivalente
    UNION {?p wdt:P2888 ?s. } # P2888: correspondance exacte
    FILTER( REGEX(STR(?s), "terme de recherche") ) # Filtre
les résultats du tableau, à partir de la chaîne de caractère
définie dans la variable ?s
}
```

(Wikidata:WikiProject_Ontology/Mapping 2019)

Dans la requête de la figure 3, le préfix *wdt* correspond à l'adresse url d'une propriété dans l'ontologie de Wikidata (Pxxx). La variable *?p* indique l'adresse de l'url correspondant à la page d'une entité ou *item* (Qxxx). La variable *?s* affiche la valeur correspondante à l'adresse url de l'ontologie externe recherchée.

Alignement avec Wikidata et d'autres bases de données

La méthode utilisée pour l'analyse des résultats extraits de l'échantillon a été en partie inspirée par la méthodologie de Thornton. Cette étude propose d'effectuer un alignement des métadonnées descriptives d'un fichier avec la plateforme Wikidata dans une optique de description (Thornton 2017).

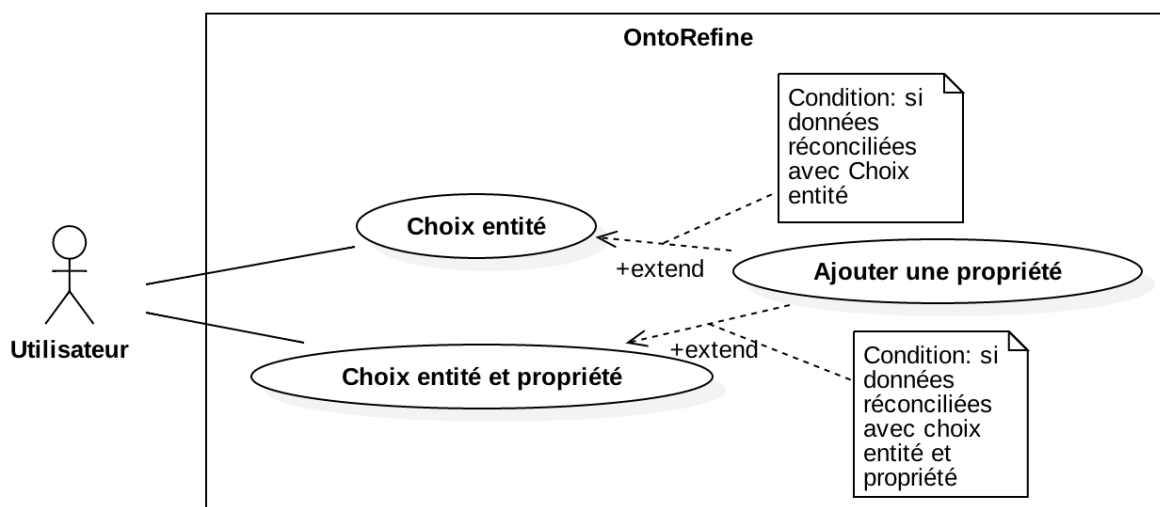
La réconciliation des données avec Wikidata présentée dans le chapitre 10.4 est réalisée avec l'API Wikidata reconciliation for OpenRefine (Wikimedia toolforge sans date d), et l'interface OntoRefine du logiciel GraphDB. Trois cas d'utilisation ont été identifiés comme illustrés sur la figure 3. Il s'agit des cas : *Choix entité*, *Choix entité et propriété* et *Ajouter une propriété*.

Le premier cas d'utilisation *Choix entité* propose l'alignement d'une série de valeurs, à partir d'une entité définie sur la plateforme Wikidata. Ce cas est celui défini par défaut dans le menu *Reconcile* dans l'interface OntoRefine.

Le deuxième cas d'utilisation *Choix entité et propriété* étend le cas d'utilisation du premier cas, en proposant d'ajouter une propriété afin de spécifier la recherche d'un concept avec Wikidata.

Le dernier cas d'utilisation *Ajouter une propriété* est uniquement accessible, à la suite d'une première réconciliation des données compris par un des deux premiers cas d'utilisations. Ce dernier cas offre la possibilité d'ajouter une ou plusieurs propriétés interne ou externe à Wikidata.

Figure 4 : Cas pratiques d'alignement avec OntoRefine



4. Open data et GLAM en Suisse

« [...] new emerging practices have been subsumed under the term “OpenGLAM” – the equivalent of “open government”, applied to the cultural heritage sector (the acronym “GLAM” stands for galleries, libraries, archives, and museums). » (Estermann 2018, p.2)

Ce chapitre propose d'identifier les principaux acteurs et groupes de travail actifs dans le domaine de l'Open Data et des GLAM (galleries, libraries, archives, and museums), en Suisse. Les résultats obtenus ont été consignés sous la forme d'une carte conceptuelle (figure 5), avant de faire l'objet d'une rédaction dans ce rapport.

Wikimedia CH (Wikimedia CH 2019) est une sous-entité régionale de la fondation Wikimedia. La fondation Wikimedia a été créée en 2003 par Jimmy Wales aux États-Unis et rassemble l'ensemble des membres de la communauté Wikimedia. La fondation comprend actuellement 12 projets actifs de plateformes de crowdsourcing. Les divers projets de crowdsourcing proposés par les communautés d'utilisateurs de Wikimedia s'intéressent, à divers domaines de la connaissance et ont pour la plupart une portée encyclopédique.

Les principaux projets participatifs de la fondation Wikimedia (Wikimedia sans date e) incluent actuellement les projets : Wikipédia, Wikibooks, Wikiversité, Wikinews, Wiktionnaire, Wikisource, Wikiquote, Wikivoyage, Wikimedia Commons, Wikidata, Wikispecies et MediaWiki.

Wikimedia CH est le chapitre suisse de Wikimedia. Cette sous-entité ou chapitre (selon la terminologie employée par la plateforme) de Wikimedia a été fondé en 2006, à Zurich. Cette sous-entité représente et coordonne l'ensemble des projets réalisés en Suisse sur les diverses plateformes de Wikimedia dans les quatre langues nationales suisses (Allemand, Français, Italien et Romanche). La fondation organise régulièrement des conférences et ateliers thématiques, avec pour objectif, de former de futurs contributeurs aux diverses plateformes de crowdsourcing de Wikimedia.

La fondation Wikimedia CH participe également, à la coordination des projets GLAM pour la Suisse, et collabore avec d'autres initiatives internationales proposées par d'autres chapitres de Wikimedia. Les rencontres GLAM & Wikimedia CH (Wikidata:Events/GLAM & Wikimedia CH 2019) sont des événements publics organisés dans le but de favoriser les retours d'expériences et les échanges de bonnes pratiques, entre les membres de la communauté Wikimedia CH et les publics externes à la fondation – institutions patrimoniales et personnes individuelles intéressées par les thématiques culturelles en Suisse.

Les différents projets GLAM en Suisse et à l'international ont en commun le projet GLAM-WIKI (GLAM 2019). Ce projet a été initié par l'entité Wikimedia Outreach, l'organe chargé des activités liées au lobbying, à la formation et à la communication de la fondation Wikimedia.

Les initiatives GLAM sont principalement présentes sur les plateformes Wikipédia, Wikimedia Commons et Wikidata. La communauté francophone de Wikipedia:GLAM regroupe les chapitres des communautés d'utilisateurs belges, canadiens, québécois, français, tunisiens et suisses. Les plateformes Wikimedia Commons et Wikidata possèdent également une affiliation au mouvement GLAM : le Commons:GLAM (Commons:GLAM 2019) et le Project Cultural heritage (Wikidata:WikiProject Cultural heritage 2019) .

L'Open Knowledge Switzerland (Open Knowledge foundation sans date b) est une sous-entité régionale de la Open Knowledge Foundation (OKFN). La OKFN a été fondée en 2004 et poursuit des activités de lobbying dans le domaine de l'Open Data. La mission principale poursuivie par cette fondation est d'offrir une sensibilisation aux enjeux et utilisations potentielles des Open Data par la société civile, les entreprises et les gouvernements. Cette fondation administre un répertoire international offrant la possibilité de publier des jeux de données en open data, le datahub (Open Knowledge Foundation sans date a). Un phénomène plus communément désigné par le terme Open Gouvernement Data (OGD).

L'Open Knowledge Switzerland a été fondé en 2011, sous l'appellation OpenData CH. Cette sous-entité de la Open Knowledge Foundation (OKFN) pour la Suisse. Un groupe de travail a été créé en 2013, pour répondre aux besoins du mouvement GLAM pour la Suisse, il s'agit de l'OpenGlam CH. La création de l'OpenGLAM CH Working Group (Open Knowledge Foundation sans date c) fait directement suite à une première initiative internationale, les OpenGLAM Principles (Open Knowledge Foundation 2013) visant à faciliter la diffusion et la réutilisation des données partagées dans le domaine culturel. L'OpenGLAM CH Working Group organise des événements communs avec la fondation Wikimedia CH, sous l'appellation OpenGlam.

Les Archives fédérales suisses (AFS) ont administré de 2016 à 2019, la plateforme opendata.swiss. Cette plateforme n'est pas directement liée au mouvement Open Data, mais appartient à une sous-catégorie de l'Open Data : le Open Gouvernement Data (OGD). L'OGD est un mouvement international qui a pour mission principale de rendre accessible les données produites par des administrations publiques. En Suisse, les données liées au OGD comprennent les jeux de données produit par l'ensemble des autorités communales, cantonales et fédérales. Les données partagées sur le portail opendata.swiss (Office fédéral de la statistique sans date) sont principalement issues des administrations publiques suisses. Cependant, il est néanmoins possible pour des organisations privées d'effectuer, sous certaines conditions, des versements vers cette plateforme.

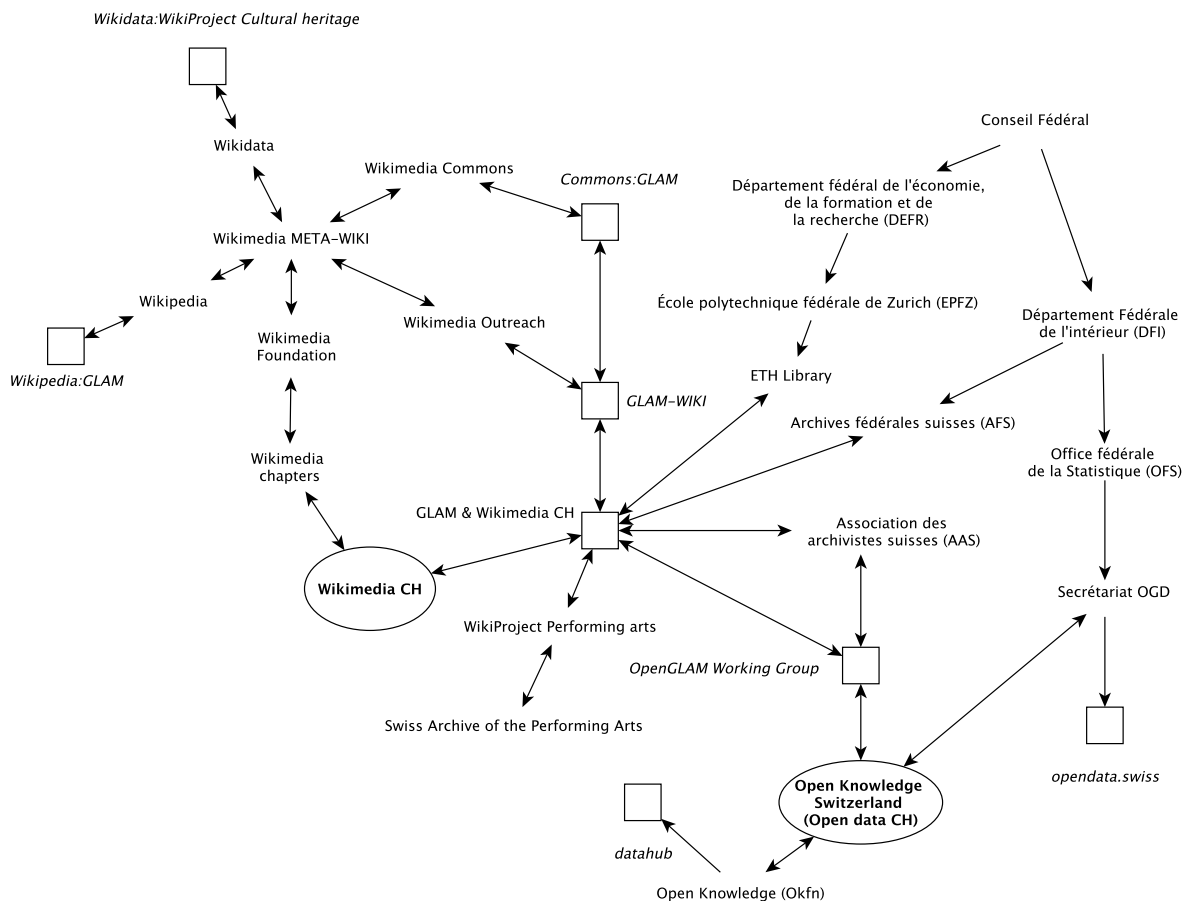
Selon la stratégie en matière de libre accès aux données publiques en Suisse pour les années 2019 à 2023, la gouvernance de la plateforme opendata.swiss changera dès 2019 (Conseil fédéral 2018, p.868), au profit d'un nouveau secrétariat dirigé par l'Office fédéral de la statistique (OFS). Le nouveau secrétariat OGD poursuivra la maintenance de la plateforme opendata.swiss et développera de nouvelles interfaces afin de valoriser les jeux de données existants et actuellement non exploités sur la plateforme.

Depuis 2017, l'association des Archivistes suisses (AAS) (Association des archivistes suisses 2017) s'est progressivement rapprochée de Wikimedia CH en organisant des activités communes, autour de la thématique GLAM.

La bibliothèque de l'École Polytechnique Fédérale de Zurich (EPFZ) a exporté sur la plateforme Wikimedia Commons, un important fonds photographique (Gasser 2019). À ce jour, le projet a permis le versement automatique de 50'000 photographies vers la plateforme. À terme, l'ensemble des versements devrait comptabiliser plus de 130'000 photographies, et constituer une des contributions les plus importantes d'une institution patrimoniale suisse sur Wikimedia Commons. Par ailleurs, la bibliothèque de l'EPFZ a effectué des travaux similaires pour des fonds d'archives photographiques de la bibliothèque nationale suisse (BN) et des AFS.

Ce premier chapitre a permis d'établir qu'il existait des synergies entre les acteurs associatifs issus du mouvement Open Data et le monde des archives patrimoniales en Suisse. En effet, une communauté d'utilisateur est présente sur le plan national et organisée en diverses associations. Certaines de ces associations espèrent également tendre à une prochaine professionnalisation de la pratique du crowdsourcing, en Suisse (voir annexe 1).

Figure 5 : GLAM et Open Data en suisse (2019)



5. Ontologies applicables à une archive audiovisuelle

Dans cette section, il est question d'effectuer un état des lieux des vocabulaires d'ontologies existants concernant les domaines de la documentation, de la conservation et de l'audiovisuel. Les vocabulaires d'ontologies généralistes c'est-à-dire ceux applicables à différents domaines de la connaissance seront également évoqués au cours de cette section.

Les différentes ontologies découvertes, à la suite des recherches sont visibles sur la figure 6 et ont été organisées en trois catégories thématiques : *spécifique à l'audiovisuelle*, *spécifique à la documentation* et *généraliste*. Le jeu de données (data set) est l'unité centrale de ce diagramme de Venn. La liste des ontologies proposées dans ce chapitre est loin d'être exhaustive. En effet, elle a été établie à partir d'une sélection restreinte d'organisations et/ou d'associations professionnelles œuvrant dans les domaines de la documentation et de l'audiovisuelle.

Ce chapitre ne traite pas des modèles de données décrivant des métadonnées techniques, comme par exemple les standards audiovisuels proposés par le groupe de travail de l'Organisation internationale de normalisation (ISO), le Moving Picture Experts Group (MPEG) ou ceux de la Society of Motion Picture and Television Engineers (SMPTE).

La première catégorie présentée sur la figure 6 (*spécifique à l'audiovisuelle*) concerne les ontologies proposées par des associations professionnelles de standardisation dans l'audiovisuel. On retrouve principalement les vocabulaires d'ontologies proposées par l'European Broadcasting Union (EBU), l'EBUCore (European broadcasting union sans date) et l'Ontology for Media Resources (ma-ont) (World wide web consortium 2012a) du World Wide Web Consortium (W3C). Ces deux ontologies permettent d'extraire les métadonnées techniques disponibles dans les modèles de données de plusieurs types de formats de fichiers audiovisuels et de les publier sur le web sémantique. Des possibilités d'alignement entre ces deux ontologies existent et sont documentées dans les spécifications de l'ontologie ma-ont.

La deuxième catégorie comprend les modèles d'ontologie concernant le domaine de la documentation et sont principalement utilisés dans le cadre des services d'information documentaire (SID), des archives et dans le monde muséal.

On retrouve dans cette catégorie les ontologies proposées par l'International Committee for Documentation (CIDOC), le CIDOC-CRM (International Committee for Documentation 2011), le Record in Context (RiC) (Internet archive sans date b) du Conseil International des Archives (CIA), l'Europeana Data Model (EDM) (Europeana foundation sans date) de la Europeana Foundation et l'ontologie BIBFRAME (Library of Congress 2017) de la Library of Congress (LOC). La Library of Congress (LOC) est l'auteure d'une série de recommandations concernant la description d'une collection audiovisuelle avec le modèle BIBFRAME, le BIBFRAME AV (Library of Congress 2015). Ces quatre ontologies sont principalement adaptées, à la description et à la documentation d'objet physique et sont très génériques.

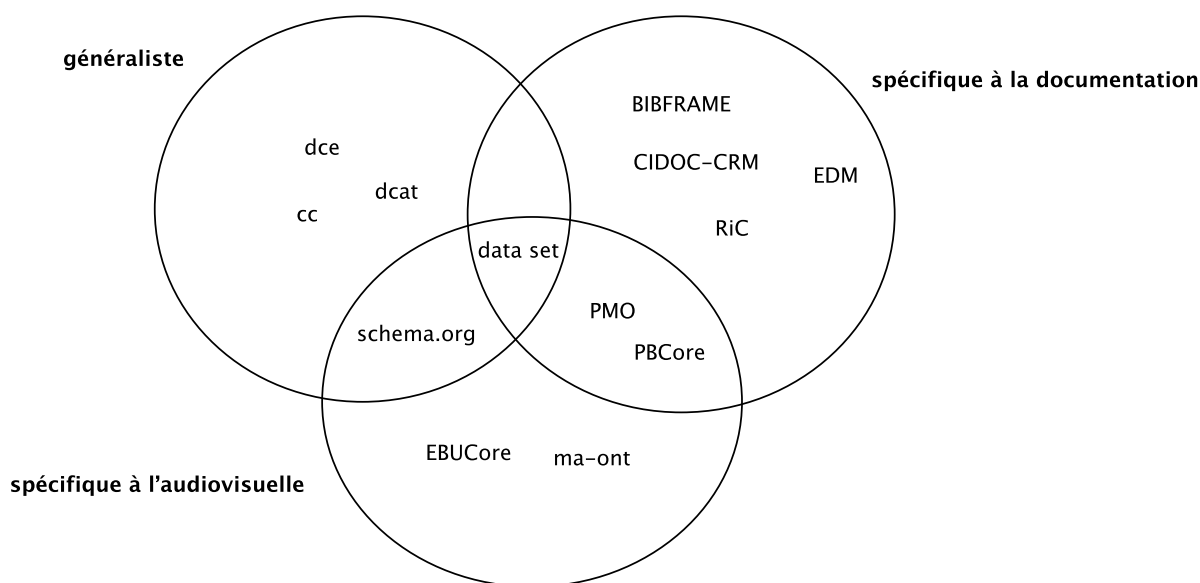
La conjonction des catégories *spécifique à l'audiovisuelle* et *spécifique à la documentation* inclut les ontologies de la Performed Music Ontology (PMO) (Stanford university sans date), une variation de l'ontologie BIBFRAME, spécialement adaptée à la description de performance musicale. Le deuxième modèle de données est le PBCore (Pbc core sans date) de la Public broadcasting communities. Ce modèle possède un vocabulaire de termes contrôlés liés à la

description d'objet audiovisuel et offre des possibilités d'alignement avec les modèles de données BIBFRAME et EBUCore.

La troisième catégorie inclut quatre ontologies généralistes employées dans divers contextes de la représentation de la connaissance et présent dans bon nombre d'ontologies citées précédemment dans ce chapitre. Il s'agit de l'ontologie de la Dublin Core Metadata Initiative, la dce (Dublin Core Metadata Initiative 2012), la cc (Creative Commons sans date b) de la fondation Creative Commons, du schema (Schema.org sans date) de schema.org et de la dcat (World Wide Web Consortium 2014a) du W3C. La première permet de décrire un document, la deuxième de préciser la nature des droits d'auteurs liés à la réutilisation d'une œuvre, la troisième au référencement sur les moteurs de recherche et la dernière facilite l'interopérabilité entre les catalogues de données sur le web.

Les ontologies citées dans ce chapitre permettent de contextualiser des jeux de données en décrivant de nombreux concepts liés aux pratiques audiovisuelles et musicales. L'utilisation d'un vocabulaire standardisé facilite l'alignement des données en Linked Open Data (LOD). La réutilisation d'une ou plusieurs ontologies évite un développement long et complexe et facilite l'alignement de termes et/ou autres folksonomies (indexation personnelle) décrites par des modèles de données de base de données en local. Enfin, l'utilisation d'ontologie apporte une meilleure gestion du multilinguisme.

Figure 6 : Ontologies applicables à une archives audiovisuelle



6. Bases de données musicales et vocabulaires contrôlés

Les résultats obtenus dans cette partie ont été consignés dans le tableau 1. Ce tableau présente une sélection de bases de données musicales commerciales et/ou open source gérée par diverses communautés d'utilisateur et référencé par Wikidata. En effet, il existe des identifiants de type *external-id* pour chacune des bases de données répertoriées dans la liste du tableau 1. Par ailleurs, une liste de vocabulaire contrôlé est également accessible, depuis la page d'accueil du projet WikiProject Music.

Tableau 1 : Liste des bases de données musicales

Base de données musicales
All About Jazz
AllMusic
Amazon Music
Discogs
Genius
Google Play Music
International Standard Musical Work Code
International Standard Recording Code
iTunes
Last.fm
Library of Congress JukeBox
MusicBrainz
RollDaBeats
Rate Your Music
Shazam
Songfacts
Spotify
Tidal
WhoSampled

(Comparison of on-demand music streaming services 2019 ; List of online digital musical document libraries 2019 ; List of online music databases 2019 ; Wikidata:WikiProject Music 2019)

7. Évaluation du fonds MJDP

Cette section propose d'établir une série de trois inventaires typologiques, à partir des enregistrements de la base de données du MJDP. Le premier de ces trois inventaires concerne les supports physiques du fonds. Cet inventaire proposera d'établir une estimation de l'encombrement du volume physique du fonds (chapitre 7.1). Le second de ces inventaires (chapitre 7.2) proposera d'établir une vue d'ensemble des formats et encodage utilisé par les enregistrements numériques audiovisuelles du fonds. Le dernier inventaire (chapitre 7.3) donnera une indication sur les divers types de données observés dans le modèle physique de la base de données relationnelles du MJDP.

L'analyse des inventaires permettra de constituer une liste de critères d'évaluation quantifiable et non quantifiable en prévision d'un futur versement vers une plateforme de crowdsourcing. Il s'agit dans ce chapitre de constituer une liste de critères exhaustifs par rapport au contexte d'une archive audiovisuelle. Les critères définis dans ce chapitre seront testés pratiquement dans les chapitres 10.2.1 et 10.3.1 de ce travail.

7.1 Inventaire typologique des supports physiques

Un recensement récent des supports physiques du fonds a permis de mettre à jour, 18 types (Dufaux, Amsallem 2017) de support audiovisuel différents. Les dates extrêmes comprises par ces supports physiques s'étendent de 1968 à 2016 (Dufaux, Amsallem 2017).

Les résultats du tableau 2 proposent une vue synthétique des différentes typologies de supports physiques du fonds, en fonction du nombre de pièces par support. Le choix du regroupement des supports magnétiques en fonction de la largeur de bande, a été établi à la suite de la lecture des recommandations de Memoriav (Office fédérale de la culture 2017, p.6, p.15).

La source des données de ce tableau est une série de statistiques extraite directement des enregistrements, de la base de données du MJDP.

La première colonne du tableau 2, *Intitulé support* mélange les catégories supports spécifiques et non spécifiques décrites par Memoriav (Office fédéral de la culture 2017, p.24) concernant la classification des supports audiovisuels. Memoriav désigne par la catégorie *spécifique* une famille générique de support, tandis que la catégorie *non spécifique* comprend l'ensemble des déclinaisons des supports d'une même famille générique.

La deuxième colonne *Largeur de bande [pouce]* donne une information sur la largeur de la bande magnétique. Cette unité de mesure est exprimée en pouce. Les colonnes *Bobine ouverte*, *Cassette magnétique*, *Disque optique* et *Disque Dure* concerne les types de supports d'enregistrements audiovisuels. La dernière colonne *Estimation du mètre linéaire* décrit une unité de grandeur exprimée en mètres concernant le volume occupé par les supports physiques magnétiques.

La somme des colonnes des pièces inventoriées dans le tableau 2 comprend un total de 17'385³ pièces. La colonne *Bande magnétique* regroupe 3'409 pièces et représente les 1/5 des supports physiques du fonds. La colonne *Cassette magnétique* (13'143 pièces) correspond, quant à elle aux 3/4 des supports physiques. Les deux dernières colonnes de ce tableau *Disque optique* et *Disque dur* incluent les supports physiques les moins fréquents, avec un peu moins de 5 %.

Il est probable que les chiffres décrits dans le tableau 2 soient prochainement revus à la hausse. En effet, de nouvelles pièces du fonds non encore inventoriées dans la base de données sont actuellement, en cours de traitement. Il s'agit principalement de cassettes magnétiques 1/4 de pouce de type DVCAM et MiniDV. On peut donc prévoir un accroissement du total de la colonne *Cassette magnétique*. Le même constat peut également être fait pour les supports physiques. En effet, notre inventaire a permis la découverte de deux nouvelles typologies de supports, le DVCAM et le MiniDV. Les contenus dans ces supports ne font pas partie des archives officielles des concerts du Montreux Jazz Festival, mais documentent les coulisses de l'événement. L'article évoqué en début de chapitre décrit uniquement la typologie des supports physiques des enregistrements des concerts officiels du MJF (Dufaux, Amsellem 2017).

L'annexe 6 propose une table de conversion de la largeur des coffrets en mètre. En l'état actuel de l'inventaire, le volume occupé par les supports magnétiques et les disques optiques s'élève selon cette estimation à environ 353 mètres linéaire.

Tableau 2 : Inventaire typologique des supports physiques

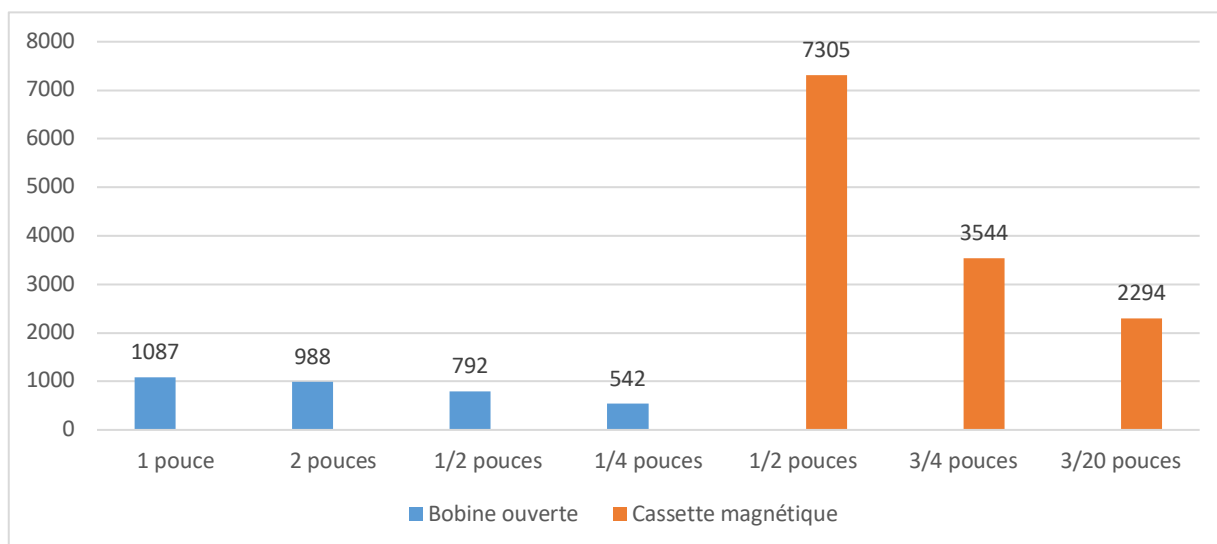
Intitulé support	Largeur de bande [pouce]	Bobine ouverte	Cassette magnétique	Disque optique	Disque dur	Estimation du mètre linéaire
Quadruplex audio et video	2"	988				100,38
1-inch Type B audio et video; 1-inch Type C audio et video; 1-inch HD video	1"	1087				34,51
Bande magnétique audio	1/2"	792				25,15
Bande magnétique audio	1/4"	542				10
U-matic audio et video ; Digital D-2	3/4"		3544			119,72
Betacam; Betacam SP; Digital Betacam; Betamax; HDCAM; Video Cassette Recording (VCR); Digital D-5; LTO-4; LTO-6	1/2"		7305			31,44
DVCAM; MiniDV (en cours d'inventaire)	1/4"					
Digital Audio Tape (DAT)	3/20"		2294			24,087
Mackies					64	
Professional Disc (PFD)				769		8,07

(Montreux Jazz Digital Project 2019a)

³ Les statistiques ont été extraites de la base de données du MJDP, le 5 mai 2019.

Le graphique de la figure 7 illustre la distribution des supports physiques en fonction des supports magnétiques du fonds. Ce graphique est basé sur les résultats compris dans les colonnes *Bobine ouverte* et *Cassette magnétique* du tableau 2. Les bandes magnétiques se présentent soit sous la forme de bobine ouverte ou alors de cassette. On constate que les cassettes magnétiques, qui possèdent une largeur de bandes de 1/2 pouces sont les plus fréquentes dans le fond du MJDP. Les bobines ouvertes d’une largeur de bande de 1 pouce arrivent en tête de la catégorie de ce type de support physique.

Figure 7 : Distribution des supports magnétiques du MJDP



(Montreux Jazz Digital Project 2019b)

7.2 Inventaire typologique des enregistrements numériques

Les métadonnées descriptives en langage XML des enregistrements numériques de chaque fichier audiovisuel sont directement consultables dans une zone de texte sur l'interface *play* de la base de données du MJDP. Certaines informations comme l'intitulé de l'extension, la durée, la date de création ou encore la taille du fichier font l'objet de champs individuels sur l'interface *play*. Pour l'instant, ces champs sont la seule utilisation concrète de métadonnées descriptives obtenues à partir de fichiers audiovisuels du fonds.

L'inventaire décrit dans le tableau 3 a été établi, à la suite de l'analyse des métadonnées descriptives de l'inventaire des extensions de fichiers audiovisuels du fonds. La possibilité offerte par l'interface *play* de consulter directement les métadonnées descriptives a permis de gagner un temps considérable pour accomplir cette tâche d'évaluation. En effet, cette option a permis d'éviter une utilisation manuelle du logiciel MediaInfo (Mediaarea 2019), permettant également d'accomplir cette lecture des métadonnées descriptives. Les extensions et les codecs découverts ont été ensuite recherchés sur la base de connaissance Sustainability of Digital Formats (Library of Congress 2019) de la Library of Congress (LOC).

Les extensions principales des fichiers observés dans le fonds numérique des archives du MJDP sont principalement pour l'audio l'extension wav. Pour les fichiers audiovisuels, les extensions avi, mov, mkv, mxf, mp4 et wmv. Les fichiers photographiques sont composés principalement des extensions jpg, tif et eps.

Les résultats du tableau 3 ont permis de mettre en évidence les standards proposés par le groupe de travail de l'Organisation internationale de normalisation (ISO), le Moving Picture Experts Group (MPEG). Ce groupe de travail est spécialisé dans la compression des formats audiovisuels comme le MPEG-4 (H-264) ou encore le MPEG-2 (H.262). Une autre organisation à l'origine de divers standards de métadonnées descriptives fréquemment employés en audiovisuel est la Society of Motion Picture and Television Engineers (SMPTE). Cette organisation propose un encodage de la base de temps des enregistrements audiovisuels, le SMPTE timecode. Ce code permet la synchronisation de l'image avec l'audio.

Le principal format de fichier conteneur découvert dans le domaine de l'audio est le Resource Interchange File Format (RIFF). Cependant, une plus grande diversité est constatée au niveau des fichiers conteneurs liés à l'audiovisuelle. En Effet, les formats conteneurs suivants ont pu être relevés, au cours de cet inventaire : l'Audio Video Interleaved (AVI), le QuickTime File Format, le Matroska Multimedia Container, le Material Exchange Format (MXF) ou encore le Advanced Systems Format (ASF). Le fonds photographique comporte également un format de fichier conteneur, le Encapsulated PostScript (EPS). Le format de fichiers principal pour l'audio est le format Waveform Audio File Format (WAVE). Les fichiers audiovisuels sont principalement représentés sous la forme des fichiers conteneurs.

Les principaux codecs rencontrés pour l'encodage des données audio sont le Pulse code modulation (PCM), le Linear Pulse Code Modulated Audio (LPCM) et le Advanced Audio Coding (AAC). L'encodage des données audiovisuelles est principalement réalisé avec le Advanced Audio Coding (AAC) et le Video Picture Encoding (V210).

Les formats de compression des données des fichiers audiovisuelles sont le MPEG-4 (H-264) et le MPEG-2 (H.262).

Les fréquences d'échantillonnage préférées pour l'audio sont le 48 Khz et le 96 Khz. L'espace colorimétrique des enregistrements des séquences est le YUV (Dufaux, Amsallem 2017). Cet espace chromatique varie entre une structure de sous-échantillonnage comprise entre 4:2:2 pour les formats les moins compressés et 4:2:0 pour les enregistrements, les plus compressés du fonds.

Tableau 3 : Format des enregistrements numériques du fonds

Format	Enregistrement (technique; encodage / standard; échantillonnage)
Waveform Audio File Format (WAVE)	Pulse code modulation (PCM); Linear Pulse Code Modulated Audio (LPCM); 48 kHz.
Waveform Audio File Format (WAVE)	Pulse code modulation (PCM); Linear Pulse Code Modulated Audio (LPCM), EBU Broadcast Wave Format 2; 48 kHz.
JPEG	Exif Exchangeable Image File Format (Exif 2.2)
TIFF, Revision 6.0	
Resource Interchange File Format (RIFF)	Waveform Audio File Format (WAVE)
Audio Video Interleaved (AVI)	Video Picture Encoding (V210); YUV 4:2:2
Encapsulated PostScript File Format (EPS)	
QuickTime File Format	Advanced Audio Coding (AAC); MPEG-4 (H-264); 48 kHz.
QuickTime File Format	Advanced Video Coding (AVC); MPEG-4 (H-264); YUV 4:2:0
QuickTime File Format	Video Picture Encoding (V210), YUV 4:2:2
QuickTime File Format	Pulse code modulation (PCM); Linear Pulse Code Modulated Audio (LPCM); 48 kHz.
Matroska Multimedia Container	Pulse code modulation (PCM); Linear Pulse Code Modulated Audio (LPCM); 48 kHz.
Matroska Multimedia Container	Advanced Video Coding (AVC), MPEG-4 (H-264); YUV 4:2:0
Material Exchange Format (MXF)	MPEG-2 (H.262); YUV 4:2:0 ou YUV 4:2:2
Material Exchange Format (MXF)	Pulse code modulation (PCM); Linear Pulse Code Modulated Audio (LPCM); 48 kHz.
Material Exchange Format (MXF)	Advanced Video Coding (AVC), MPEG-4 (H-264); YUV 4:2:2
Material Exchange Format (MXF)	Pulse code modulation (PCM); Linear Pulse Code Modulated Audio (LPCM); 48 kHz.
Advanced Systems Format (ASF)	Windows Media Video (WMV); YUV 4:2:0 ou YUV 4:2:2

(Montreux Jazz Digital Project 2019c)

7.3 Inventaire typologique des données

Le schéma de la base de données relationnelle du Montreux Jazz Digital Project (MJDP) se compose d'un total de 160 tables comprenant 957 colonnes différentes. La somme des lignes ou enregistrement est d'environ 10 millions et représente un volume de données de 7,2 Go. La table *file* compose environ 80% du volume de l'ensemble des données de la base de données⁴.

⁴ Statistiques extraites le 12 juin 2019 de la base de données du MJDP, à l'aide de l'option *Data dictionary* de phpMyAdmin

Les types de données observés dans le modèle physique de la base de données relationnelles du MJDP sont au nombre de 11. Ces types de données concernent principalement des données de type textuelle, numérique et date. La typologie des formats de données textuelles inclut les types : *varchar*, *longtext*, *mediumtext* et *text*. Les données numériques sont principalement des types *int*, *tinyint*, *bigint* et *double*. Enfin, les données concernant des dates sont principalement du type *date*, *datetime* et *time*.

7.4 Liste des critères

Pour faciliter le choix des tables et des colonnes de la base de données du MJDP, une liste de critères quantifiables a été établie à partir des différents inventaires réalisés, au cours de cette section 7. Les critères non quantifiables ont été établis à partir de la nature contextuelle et historique du fonds.

Les critères quantifiables et non quantifiables découverts sont au nombre de 27 et sont présentés dans les tableaux 4 et 5. Les critères ont été regroupés en deux catégories, les critères quantifiables (tableau 4) et ceux non quantifiables (tableau 5). La première catégorie de critères définit principalement des critères d'ordre technique, liée à des caractéristiques physiques et numériques des pièces du fonds. La seconde catégorie quant à elle, regroupe des caractéristiques contextuelles et non quantifiables.

Les intitulés des critères décrits dans les tableaux 4 et 5 sont accompagnés d'exemples afin d'éviter d'éventuelles ambiguïtés de sens. Certains intitulés ont été directement choisis à partir des définitions proposées par Memoriav. (Office fédéral de la culture 2017)

Tableau 4 : Liste des critères quantifiables du fonds

Intitulé du critère	Exemple de critère mesurable
Moyen de stockage	Cassette magnétique
Largeur de bande du moyen de stockage	1 pouce
Variation d'un même modèle	LTO-4; LTO-6
Nature de l'enregistrement	Analogique / Numérique
Technique d'enregistrement magnétique	Helical scanning / Linéaire
Extension informatique	mov; wav
Compression	MPEG-2 Video Encoding (H.262)
Qualité de la compression	4:4:4; 4:2:2
Débit binaire	221 Mbps
Format image	16:9; 4:3
Résolution de l'image	SD; Full HD
Type de balayage	Entrelacé / Progressif
Fréquence	50 Hz; 60 Hz
Echantillonnage	25 fps; 29.97 fps
Standard	PAL (Phase Alternating Line)
Identifiant externe WD	P154; P31
Item WD	Q83299; Q32883
Date	1969-06-22; 2009-07-09

Tableau 5 : Liste des critères non quantifiables du fonds

Intitulé du critère	Exemple de critère non mesurable
Concert	Benny Carter Orchestra; Etta James
Personne	Stanley Clarke; Dizzy Gillespie
Contrat	Stanley Clarke
Lieux	Casino Montreux; MDH Club
Instrument	Alto flute; Harp
Chanson	Downstairs Blues; Redemption Song
Genre musical	Blues; Jazz; Rock
Label	Atlantic; Jazz Horn Music Corporation
Format numérique	Fichier / Container

Ce chapitre a permis d'établir un aperçu quantitatif et qualitatif de la nature des données disponibles dans le fond audiovisuel du MJDP. Les critères quantifiables et non quantifiables établis à la suite de ce chapitre serviront de base à la sélection des données, en prévision d'un futur versement vers une plateforme de crowdsourcing.

8. Adapter le modèle de données du MJDP aux linked data

L'objectif principal de cette section est de proposer une nouvelle modélisation du schéma relationnel de la base de données du MJDP. Cette nouvelle modélisation est basée sur les recommandations du standard Records in Contexts (RiC), proposé par le Conseil International des Archives (CIA) et des recommandations du World Wide Web Consortium (W3C) concernant l'alignement d'une base de données relationnelle vers une base de données de graphe (World Wide Web Consortium 2010).

Cette section offre des pistes de réflexion sur l'adaptabilité du modèle de description archivistique RiC. En montrant notamment la faisabilité de transposer théoriquement les entités, les propriétés et les relations de RiC vers un modèle de données relationnelle non construit autour des standards de description archivistique usuelle.

8.1 Choix des entités, propriétés et relations

La nouvelle modélisation illustrée dans l'annexe 14 est le résultat de la troisième étape de la procédure de la figure 1. Les entités représentées dans l'annexe 14 sont au nombre de 30 et proposent une vision simplifiée de l'actuel modèle relationnel du MJDP. En effet, certains concepts du modèle initial ont été volontairement écartés dans cette nouvelle modélisation, comme les éléments liés à la gestion administrative (comptes utilisateurs) ou encore à l'activité administrative (record management) du centre MetaMedia Center (MMC). La partie décrite par cette nouvelle modélisation cherche à documenter le contexte de l'archive définitive du Montreux Jazz Festival.

L'entité centrale de ce nouveau modèle *Concert* décrit un événement unique et distinct, à la fois dans le temps et géographiquement. Cette entité possède des liens directs avec les entités *Date Concert*, *Place Concert*, *Playlist*, *Recording*, *Group* et *Person*. Ces entités précisent des éléments contextuels de l'entité *Concert*.

Les entités *Date Concert* et *Place Concert* définissent respectivement la date unique d'un concert et l'emplacement géographique d'une salle de concert du Montreux Jazz Festival. *Person* et *Group* détaillent la formation musicale, qu'il s'agisse d'un musicien isolé et/ou d'un orchestre. L'entité *Playlist* propose une liste des morceaux joués et autres événements décrits, au cours de la performance (applaudissement, discours, réglage d'instrument, etc..). Enfin, *Recording* regroupe les différentes versions des enregistrements audiovisuels disponibles par concert.

L'entité *Recording* est la deuxième entité la plus importante de ce modèle. Elle possède des liens hiérarchiques avec les entités *LTO tape*, *Video tape*, *Audio tape* et *Concert files*. Les trois premières entités décrivent les différents supports physiques disponibles dans le fond. *Concert files* comprend la description des supports numériques du fonds et diverses informations relatives, à l'acquisition et documentation des enregistrements numériques du fonds.

L'entité *Concert files* possède des relations hiérarchiques avec les entités *Date Files*, *Photo file*, *Video file* et *Audio file*. *Photo file* décrit des photographies numérisées composées de divers formats. Les trois entités *Photo file*, *Video file* et *Audio file* documentent les métadonnées descriptives extraites des enregistrements audiovisuels. Les données renseignées par ces entités sont le résultat de la décompression des fichiers containers et des fichiers informatiques générés, à la suite du processus d'acquisition.

L'entité *Contract* définit les différentes modalités concernant les droits d'auteurs et de diffusion d'un enregistrement de l'entité *Recording* du modèle. Elle possède principalement des liens avec les entités *Person* et *Concert*. *Date Contract* indique la date de signature d'un contrat. *Contract* pointe uniquement en direction de l'entité *Concert*.

Les deux entités *Festival* et *Archives* apportent des éléments contextuels, liés, à la gestion administrative du fonds d'archives. Elles comprennent l'ensemble des entités du nouveau modèle en documentant le contexte de création et d'accroissement des différentes unités intellectuelles du fonds.

Le nouveau modèle comprend 9 des 14 types d'entités définis par le modèle conceptuel du standard Records in Contexts (RiC) (International Council on Archives 2016, p.13-19). Cependant, ce nombre s'élève à 17, si les sous-entités *RiC-E4 Agent (of type Person)*, *RiC-E4 Agent (of type Delegate-Agent)* et *RiC-E4 Agent (of type Corporate Body and Corporate Body Assumed Identity)* sont incluses dans ce calcul.

La distribution des entités RiC dans le modèle est la suivante : 8 entités sont du type *RiC-E1 Record* et 7 du type *RiC-E12 Date*. Les entités du type *RiC-E3 Record Set* sont présentes 6 fois et *RiC-E13 Place*, 4 fois. *RiC-E4 Agent (type Person)*, *RiC-E4 Agent (type Group)*, *RiC-E5 Occupation*, *RiC-E9 Activity* et *RiC-E10 Mandate* apparaissent une seule fois dans le nouveau modèle.

Records in Contexts (RiC) définit 69 propriétés (International Council on Archives 2016, p. 21-38) réparties en moyenne 9 fois dans chaque entité. Les 4 premières propriétés du modèle RiC sont des propriétés communes à l'ensemble des entités du modèle. Il s'agit des propriétés suivantes : *RiC-P1 Global Persistent Identifier*, *RiC-P2 Local Identifier*, *RiC-P3 Name* et *RiC-P4 General Note*. Dans notre modélisation ces propriétés ont été incluses au nouveau modèle et apparaissent uniquement dans le diagramme de classe (figure 9).

Les 30 entités décrites dans le nouveau modèle incluent 57 des 69 propriétés définies par le standard RiC. Les propriétés comprises entre *RiC-P5* et *RiC-P31* apparaissent 7 fois, ainsi que *RiC-P61* et *RiC-P62*.

Le nombre de relations défini entre les différentes entités du nouveau modèle représente moins de 2% des 792 relations théoriques prévues par le standard RiC (International Council on Archives 2016, p.40-90).

8.2 Validation du nouveau modèle conceptuel

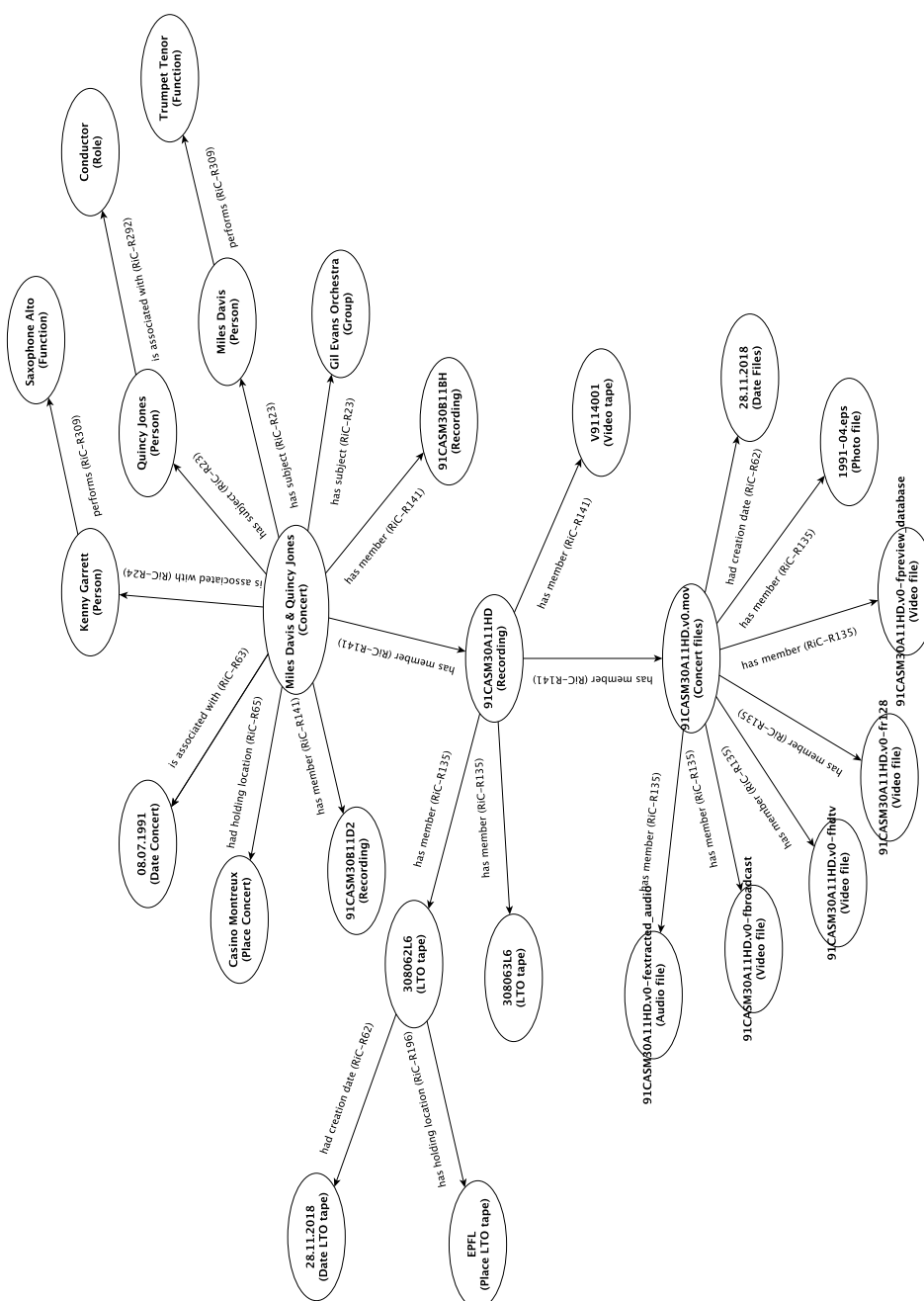
La figure 8 présente le diagramme développé dans le but de valider le nouveau modèle conceptuel de données. Il se compose de 17 des 30 entités définies précédemment par le nouveau modèle de l'annexe 14 et comprend 8 types d'entités différentes sur les 9 définies dans le nouveau modèle. Quelques 10 types de relations différentes sont incluses dans cet exemple sur les 14 initialement prévu par le nouveau modèle. Ce diagramme intègre également 14 colonnes réparties dans 10 tables, ce qui correspond à un peu plus de 6 % du volume des 160 tables de la base relationnelle du MJDP.

Une recherche manuelle de chacun des termes de l'exemple de la figure 8 sur Wikidata a permis de confirmer une occurrence pour 11 valeurs d'entités sur 17. Le diagramme de l'annexe 15 décrit les entités de la figure 7, qui paraissent être d'excellents candidats pour un alignement avec la plateforme Wikidata.

De manière générale, on constate des occurrences élevées concernant les valeurs des entités correspondantes à des lieux géographiques ou des instruments de musique. En revanche, il est difficile de trouver des occurrences concernant les dates ou les intitulés de concerts.

Les entités représentées en traitillé sur le diagramme de l'annexe 15 indiquent des valeurs recherchées, mais actuellement non disponibles dans Wikidata. Cela signifie que ces valeurs n'ont pas encore été ajoutées ou décrites par un utilisateur de la communauté de Wikidata et qu'en l'état, un alignement n'est pas envisageable.

Figure 8 : Exemple d'un enregistrement du MJDP adapter à RiC



9. Réalisation du nouveau modèle conceptuel

Cette section propose la réalisation pratique du diagramme présentée dans le chapitre 8.2 afin de démontrer la faisabilité pratique du modèle illustré dans l'exemple de la figure 8.

Les résultats obtenus ont montré la faisabilité d'un modèle partiel développé à partir d'un enregistrement spécifique de la base de données du MJDP. Néanmoins, on notera le caractère encore expérimental de ce livrable, en effet les fonctions administratives du modèle de données initial n'ont pas été prises en compte dans le développement de cette solution. De plus, un alignement avec d'autres ontologies autres que RiC n'a pas été réalisé.

9.1 Diagramme de classe

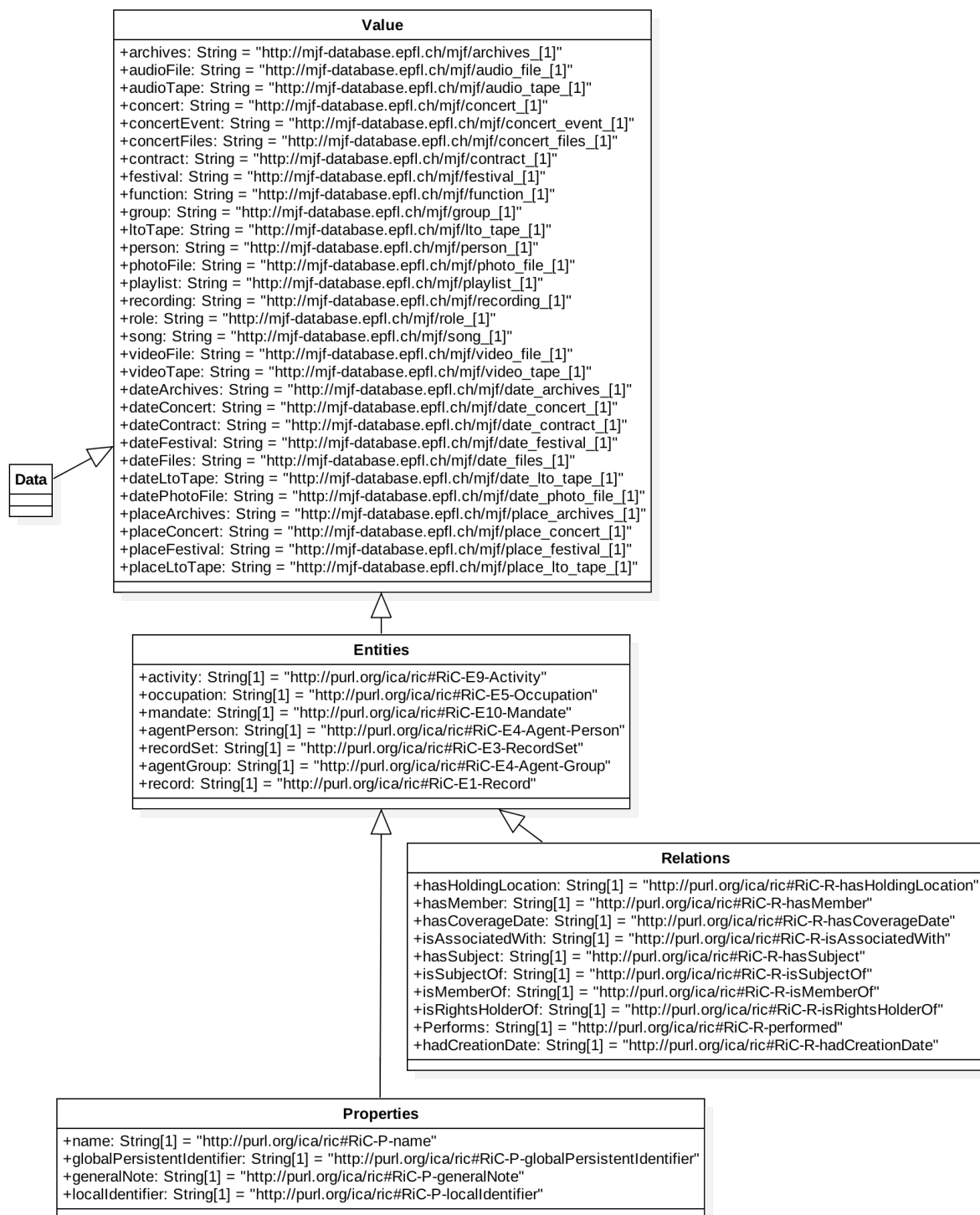
Le diagramme de classe de la figure 9 se compose des cinq classes nécessaires à l'implémentation physique du nouveau modèle conceptuel de données (annexe 14). Ce diagramme est composé des classes *Value*, *Entites*, *Properties*, *Relations* et *Data*. Les attributs du modèle de la figure 9 utilisent des chaînes de caractère (String) comme type de données.

La classe *Value* est composée des attributs correspondants aux entités décrites dans le modèle conceptuel et définis par une adresse Uniform Resource Identifier (URI). Chaque URI se termine par une valeur numérique unique incrémentée à chaque nouvel enregistrement.

Les attributs des classes *Entites*, *Properties* et *Relations* définissent les différentes entités du modèle Records in Contexts (RiC) sélectionnées pour la modélisation du nouveau modèle. Chaque attribut de ces trois classes se compose d'une adresse Internationalized Resource Identifier (IRI) défini par l'ontologie RiC (Internet Archive sans date b) et accessible à distance, depuis le répertoire PURL (Internet Archive sans date a).

Des relations de généralisation sont définies entre les différentes classes du modèle, et permettent d'introduire une notion de hiérarchisation entre les classes. L'entité *Value* est la classe la plus générique de ce modèle, elle possède deux relations de généralisation, qui pointent vers elle. Il s'agit des classes *Data* et *Entities*. Les classes *Properties* et *Relations* possèdent également une relation de généralisation vers la classe *Entities* et partagent des propriétés communes avec *Entities*.

Figure 9 : Diagramme de classe du nouveau modèle



9.2 Matrice d'adjacence

La structure de la matrice d'adjacence proposée sur le tableau 6 est composée par les valeurs uniques des IRI définies par les attributs de la classe *Value*. Les colonnes *MJF_table.colonne = Entities.attribut*, *MJF_table.colonne = Properties.attribut* et *MJF_table.colonne = Relations.attribut* incluent les valeurs uniques des attributs des entités *Entities*, *Properties* et *Relations*.

Dans le tableau 6, *Value.attribut* indique à la fois la valeur d'un nœud ou d'un champ du futur graphe (figure 11). Cette valeur sera interprétée comme un nœud, si elle se trouve dans la première colonne du tableau 6. Cependant, si elle se trouve à l'intersection d'une ligne et d'une colonne, elle sera interprétée soit comme la valeur d'un champ ou d'un nœud. Chaque ligne du tableau 6 correspond à un enregistrement unique construit à partir d'un IRI défini dans la base de données de graphe.

Le résultat de la matrice d'adjacence créée, à partir du nouveau modèle conceptuel comprend un total de 7 colonnes pour la catégorie *MJF_table.colonne = Entities.attribut*, 10 colonnes pour la catégorie *MJF_table.colonne = Relations.attribut* et 4 colonnes pour la catégorie *MJF_table.colonne = Properties.attribut*. Un total théorique de 21 colonnes⁵ est nécessaire pour réaliser la matrice d'adjacence du modèle conceptuel. Cependant, ce chiffre est fortement dépendant du nombre de colonnes choisi dans chaque table de la base de données relationnelles.

Tableau 6 : Structure de la matrice d'adjacence du graphe

	MJF_table.colonne = Entities.attribut	MJF_table.colonne = Properties.attribut	MJF_table.colonne = Relations.attribut
Value.attribut	Value.attribut	Value.attribut	Value.attribut

La matrice d'adjacence créée manuellement et adaptée à l'exemple de la maquette de la figure 8 comprend un total de 23 colonnes : 17 colonnes concernent la catégorie *MJF_table.colonne = Entities.attribut* et 6 colonnes la catégorie *MJF_table.colonne = Relations.attribut*. Le graphe de la figure 11 n'intègre pas les nœuds concernant les valeurs des 4 colonnes de la catégorie *MJF_table.colonne = Properties.attribut*. Ces valeurs ont été écartées volontairement de la modélisation.

9.3 Fichier de triplet

La requête en langage SPARQL (voir annexe 8) permet de construire le graphe de la figure 11. Cette requête est composée des préfix *rdfs*, *ric*, *mjf* et *spif*, définis par des adresses IRI distantes employées pour la construction de cette requête. Le préfix *ric* comprend l'IRI externe de l'ontologie RiC et *mjf* celui de l'adresse IRI des valeurs des entités *Value.attribut* du nouveau modèle. Le préfix *spif* (Spin sans date) est utilisé pour optimiser certaines opérations sur les valeurs des propriétés.

⁵ Le résultat énoncé ci-dessus prend en compte une situation idéale, où l'on sélectionne uniquement une colonne par table pour la réalisation pratique de la matrice d'adjacence.

L'instruction CONSTRUCT permet la construction du graphe, à partir des 23 colonnes définies par la matrice d'adjacence de l'exemple. La structure d'un triplet introduit par la variable *?base* permet la création d'un IRI unique pour chaque ligne d'enregistrement de la matrice d'adjacence.

L'instruction WHERE offre la possibilité de sélectionner les 23 colonnes de la matrice d'adjacence. Le triplet introduit par la variable *?row* sélectionne toutes les colonnes définies par les catégories *MJF_table.colonne = Entities.attribut*, *MJF_table.colonne = Properties.attribut* et *MJF_table.colonne = Relations.attribut*.

L'instruction SERVICE de la requête permet de définir l'adresse IRI local du graphe dans la base de données GraphDB et l'instruction BIND, l'encodage de l'adresse IRI.

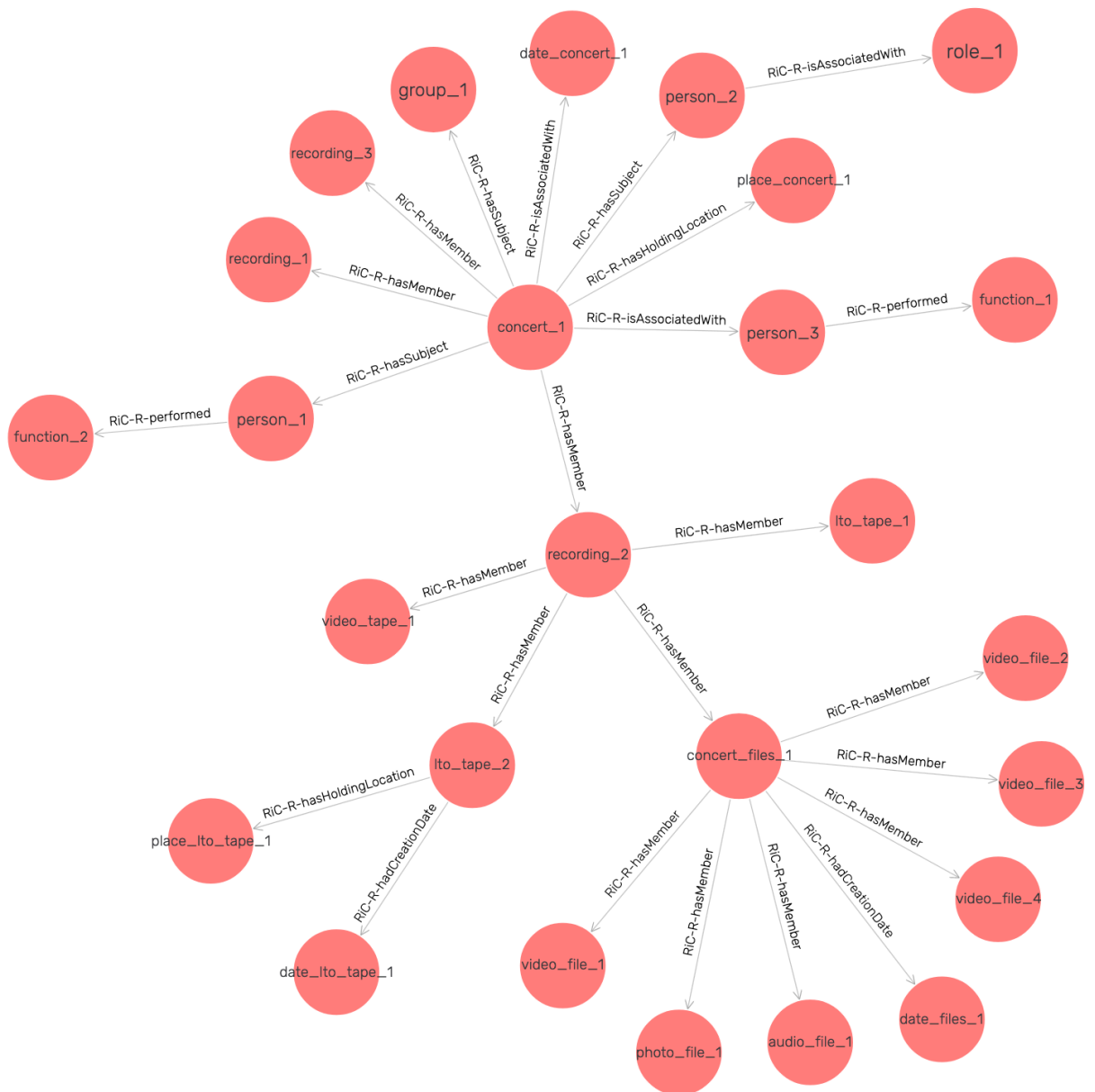
La requête qui permet la mise à jour de la base de données de graphe conserve la même structure que la requête CONSTRUCT (voir annexe 8). L'instruction CONSTRUCT a été simplement remplacée par INSERT.

Le résultat de la requête INSERT permet la visualisation du résultat du graphe de la figure 11 sur l'interface de la base de données de graphe GraphDB. Cette visualisation utilise la librairie Graphviz (Github 2019). La représentation graphique de la figure 11 est basée sur le code source de l'export du fichier de triplet au format N-Triples de GraphDB, illustré dans la figure 10.

Figure 10 : Extrait du fichier exporté de GraphDB

```
<http://mjf-database.epfl.ch/mjf/concert_1>
<http://purl.org/ica/ric#RiC-E3-RecordSet> "Miles Davis &
Quincy Jones" .
<http://mjf-database.epfl.ch/mjf/recording_1>
<http://purl.org/ica/ric#RiC-E3-RecordSet> "91CASM30B11D2" .
<http://mjf-database.epfl.ch/mjf/recording_2>
<http://purl.org/ica/ric#RiC-E3-RecordSet> "91CASM30A11HD" .
<http://mjf-database.epfl.ch/mjf/recording_3>
<http://purl.org/ica/ric#RiC-E3-RecordSet> "91CASM30B11BH" .
```

Figure 11 : Visualisation du graphe dans GraphDB



10. Partager les données du MJDP en linked open data

10.1 Exploration et sélection des données externes sur Wikidata

L'enjeu de cette partie est d'effectuer une série de mesures quantitatives afin de sélectionner, puis choisir manuellement des vocabulaires d'ontologies GLAM et des propriétés de type *external-id* disponibles sur la plateforme de Wikidata.

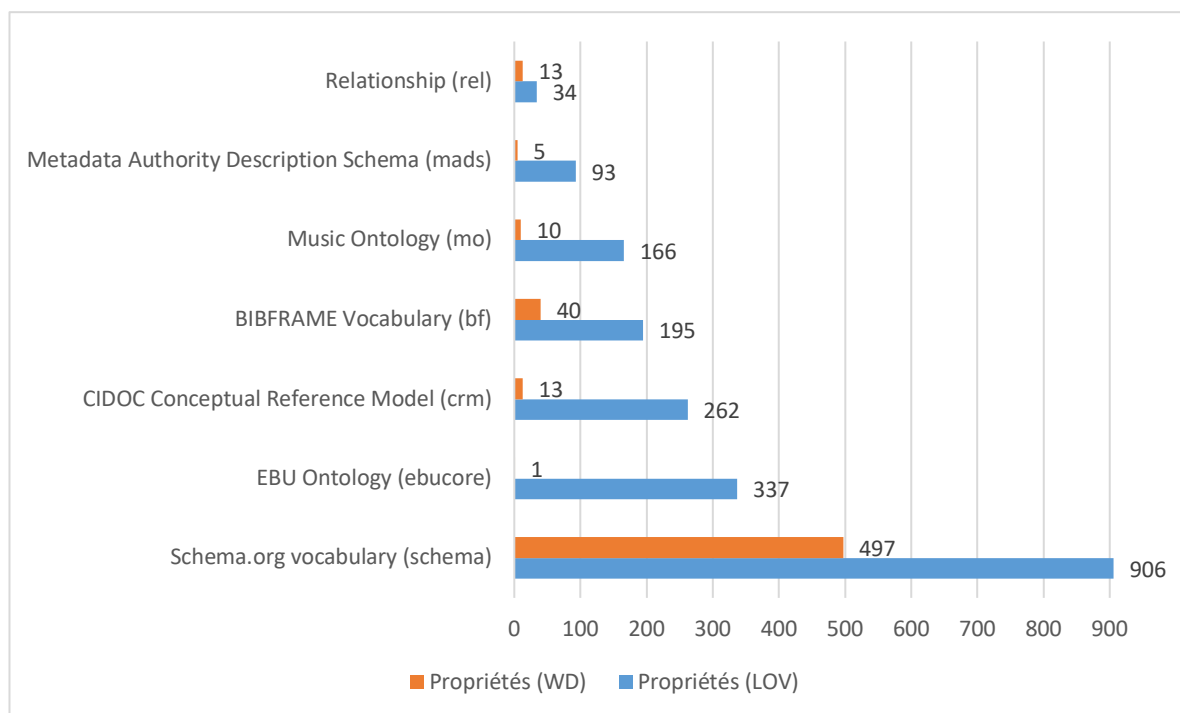
10.1.1 Statistique sur les vocabulaires d'ontologies GLAM

Le répertoire LOV référençait au mois de mai 2019, date à laquelle ces différentes mesures ont été réalisées, un total de 666 vocabulaires d'ontologies et plus de 63'000 propriétés différentes.

On observe sur la figure 12 qu'une faible proportion des propriétés sont définies sur la plateforme Wikidata pour chaque vocabulaire d'ontologie testé dans l'échantillon. En effet, le vocabulaire d'ontologie qui possède le plus de propriétés décrites sur la plateforme Wikidata est le vocabulaire schema.

De manière générale, la fréquence d'utilisation d'un vocabulaire d'ontologie peut être facilement obtenue en dénombrement le nombre de liens pointant vers un vocabulaire d'ontologie, à partir de la visualisation⁶ disponible pour chaque vocabulaire sur LOV. Cette information offre une excellente piste de réflexion pour le choix d'un vocabulaire d'ontologie ou un argument supplémentaire, quand il s'agit de choisir entre deux vocabulaires similaires.

Figure 12 : Comparaison entre 7 vocabulaires d'ontologies



(Linked Open Vocabularies 2019)

⁶ L'exemple de *dcterms* sur le répertoire Linked Open Vocabularies (LOV) donne un bon aperçu des relations directes entrantes et sortantes de ce vocabulaire d'ontologie : <https://lov.linkeddata.es/dataset/lov/vocabs/dcterms>

10.1.2 Choix et sélection manuelle de propriétés

Un inventaire (voir annexe 9) non exhaustif des propriétés de la plateforme Wikidata a été établi, à partir d'une sélection manuelle des propriétés concernant spécifiquement le contexte du MJDP. Les principales sources d'informations consultées afin d'établir cet inventaire ont été principalement la page du WikiProject Music et la liste exhaustive des propriétés de Wikidata (Wikidata:Database reports/List of properties/all 2019). Le critère de sélection des propriétés a été opéré en fonction des bases de données musicales recensées précédemment dans la section 6.

Une première requête SPARQL (voir annexe 10) regroupe un total de 35 identifiants de type *external-id* et a été créée à partir de l'inventaire (voir annexe 9). Ces 35 identifiants de l'échantillon appartiennent aux quelque 3'900 propriétés du même type disponible sur Wikidata. Le résultat obtenu à partir de cette requête est de 415'394 *items* sur un total de 57'300'000⁷ *items*, ce qui représente moins d'un 1 % de l'ensemble des *items* disponibles sur la plateforme Wikidata.

À partir de cette première mesure, 5 autres requêtes SPARQL ont été créées, à partir des 35 identifiants de l'échantillon de départ. Les requêtes concernent les sujets suivants : les morceaux, les albums, les maisons de production, les artistes et les identifiants spécifiques à la base de données *MusicBrainz*.

Le résultat de la requête (voir annexe 10) concernant les artistes montre que les identifiants *MusicBrainz artist ID*, *Discogs artist ID*, *AllMusic artist ID* comprennent 90 % des résultats des *items* présents sur la plateforme Wikidata. Les 10 % restants sont divisés entre les 14 autres identifiants compris dans le résultat de cette requête.

Dans la requête (voir annexe 10) concernant les morceaux, les identifiants *MusicBrainz work ID*, *AllMusic song ID* et *Spotify track ID* composent 97 % des résultats. Cependant, *MusicBrainz work ID* se détache significativement du reste des résultats. Il représente un peu plus de 87 % des résultats. Les identifiants *Spotify album ID*, *MusicBrainz recording ID* et *iTunes album ID* regroupent 89 % des résultats de la requête (voir annexe 10) concernant les albums. L'identifiant *MusicBrainz label ID* occupe plus de 75 % des résultats dans la requête (voir annexe 11) sur les maisons de production.

Les identifiants *MusicBrainz artist ID*, *MusicBrainz release group ID* et *MusicBrainz work ID* (voir annexe 11) sont présents dans près de 89 % des résultats des identifiants spécifiques, à la base de données musicales *MusicBrainz*.

À la suite des résultats des différentes requêtes, une liste des identifiants de type *external-id* susceptibles d'offrir le plus grand nombre d'occurrences dans Wikidata (tableau 7) a été créée afin de faciliter le choix d'identifiant externe.

⁷Ce chiffre a été établis à partir des statistiques de Wikidata du 6 juin 2019 : <https://grafana.wikimedia.org/d/000000167/wikidata-datamodel?refresh=30m&orgId=1>

Tableau 7 : Liste des identifiants de bases de données musicales

Choix identifiants Wikidata
MusicBrainz artist ID
Discogs artist ID
AllMusic artist
MusicBrainz work ID
AllMusic song ID
Spotify track
Spotify album ID
MusicBrainz recording ID
iTunes album ID
MusicBrainz label ID
MusicBrainz artist ID
MusicBrainz release group ID
MusicBrainz work ID

(Wikidata:Database reports/List of properties/all 2019)

10.1.3 Validation du choix d'un identifiant externe

Notre échantillon est composé de la valeur de l'*item* Q188969, qui correspond à la page de description du musicien B.B King sur la plateforme Wikidata. La page de cet *item* comprend un total de 200 propriétés différentes, dont 60 correspondent à des propriétés de type *external-id*. Ce résultat a été obtenu à partir de la requête de la figure 13.

Figure 13 : Requête SPARQL pour afficher l'ensemble des propriétés d'un item

```
SELECT DISTINCT ?p
WHERE {wd:Q188969 ?p ?o}
```

(Wikidata:SPARQL query service/queries/examples 2019)

La requête suivante (figure 14) a permis d'effectuer des recherches, à partir d'une sélection d'identifiants de type *external-id* définis dans le tableau 7. Le tableau 8 décrit les résultats obtenus pour la valeur des trois premiers identifiants compris dans la requête de la figure 14 de l'*item* Q188969 (B.B King).

Figure 14 : Requête SPARQL avec des identifiants bases de données musicales

```

SELECT *
WHERE {
  {wd:Q188969 wdt:P434 ?o .} #MusicBrainz artist ID
  UNION {wd:Q188969 wdt:P1953 ?o .} #Discogs artist ID
  UNION {wd:Q188969 wdt:P1728 ?o .} #AllMusic artist
  UNION {wd:Q188969 wdt:P435 ?o .} #MusicBrainz work ID
  UNION {wd:Q188969 wdt:P1730 ?o .} #AllMusic song ID
  UNION {wd:Q188969 wdt:P2207 ?o .} #Spotify track
  UNION {wd:Q188969 wdt:P2205 ?o .} #Spotify album ID
  UNION {wd:Q188969 wdt:P4404 ?o .} #MusicBrainz recording ID
  UNION {wd:Q188969 wdt:P2281 ?o .} #iTunes album ID
  UNION {wd:Q188969 wdt:P966 ?o .} #MusicBrainz label ID
  UNION {wd:Q188969 wdt:P436 ?o .} #MusicBrainz release group
  UNION {wd:Q188969 wdt:P435 ?o .} #MusicBrainz work ID
}

```

(Wikidata:SPARQL query service/queries/examples 2019)

Tableau 8 : Extrait du résultat des valeurs des propriétés de type *external-id*

Intitulé	Valeur	Url bd externe
Discogs artist ID	37729	https://www.discogs.com/artist/\$1
MusicBrainz artist ID	dcb03ce3-67a5-4eb3-b2d1-2a12d93a38f3	https://musicbrainz.org/artist/\$1
AllMusic artist ID	mn0000059156	https://www.allmusic.com/artist/\$1

(Wikidata sans date a)

10.2 Préparation d'un versement vers Wikidata

10.2.1 Choix et mise en pratique des critères d'évaluation

Le choix des critères d'évaluation pour un versement sur la plateforme Wikidata a été établi, à partir des deux listes de critères définis par les tableaux 4 et 5 du chapitre 7.4.

Des 27 critères théoriques définis initialement, seulement 5 critères ont été retenus pour l'évaluation pratique du fonds. En effet, une large majorité des critères définis théoriquement ont été jugés trop spécifiques pour les besoins de cette évaluation pratique du fonds. Ils ont été volontairement écartés de l'analyse, à la suite d'une première tentative de regroupement.

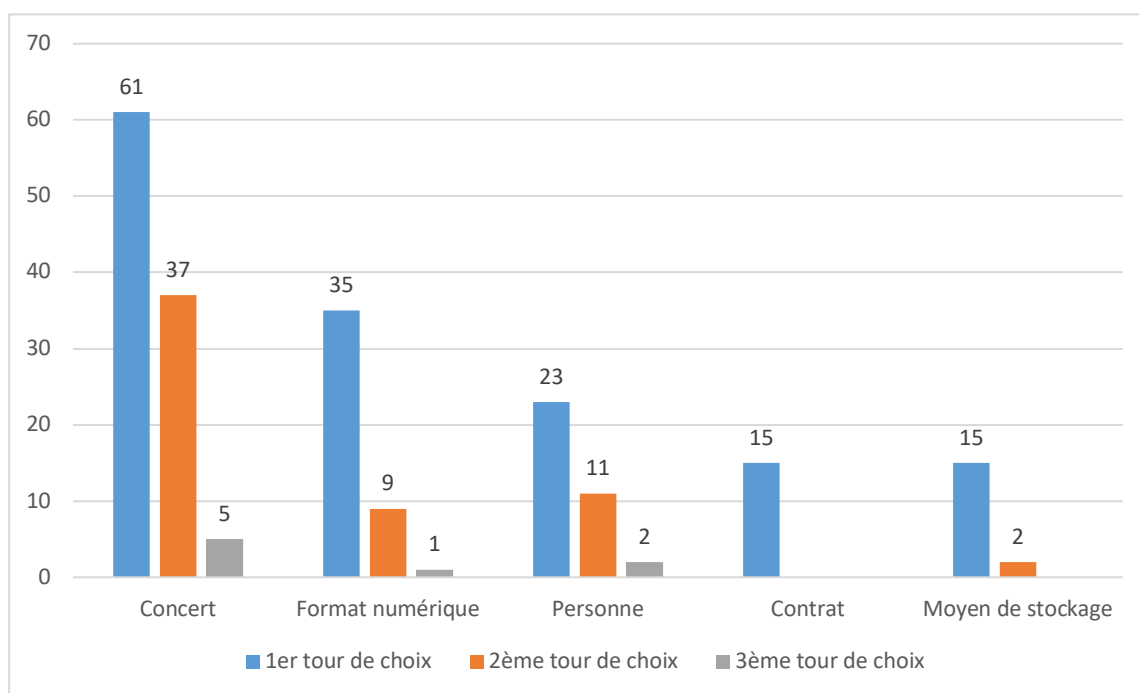
Le choix des tables et des colonnes de la base de données relationnelles s'est déroulé en trois tours de choix (figure 15). Le premier tour de choix a permis le regroupement thématique des 149 tables de la base de données relationnelles du MJDP⁸. Il est intéressant de relever que

⁸ Extraction effectuée à partir du fichier SQL de base de données du MJDP, le 20 mars 2019.

11 tables n'ont pas pu être regroupées sous l'un des critères définis pour l'analyse. Les tables non regroupées ont été directement écartées, car elles ne correspondaient pas à nos besoins pour ce transfert de données.

Le deuxième tour de choix a permis de réduire le nombre de tables de 149 à 59. On constate que les tables regroupées avec le critère *Contrat* n'ont pas été retenues, à la fin de ce second tour. En effet, la nature de certains contrats ne rentre pas dans le champ des données directement partageables librement dans le domaine public sous licence CC0.

Figure 15 : Distribution des tables sélectionnées par catégories et tour de choix



(Montreux Jazz Digital Project 2019d)

Le troisième tour de choix a réduit le volume des tables de 59 à 8. Cette fois-ci, on constate que deux catégories supplémentaires ont disparu (*Événement* et *Moyen de stockage*). Ce tour a donc permis de faire émerger un total de 9 colonnes du type *varchar* de la base de données relationnelle du MJDP. Le résultat de cette évaluation (tableau 9) montre que 5% des tables de la base de données du MJDP pourraient être des candidats potentiels, à un alignement avec la plateforme Wikidata. Ce chiffre se réduit à moins de 1%, si l'on se base uniquement sur les colonnes.

Tableau 9 : Intitulé des tables et colonnes du MJDP sélectionnées pour un transfert

Intitulé du critère	Table MJDP	Colonne MJDP
Concert	composition	title
Concert	concert	name
Concert	concert	concert_date
Concert	instrument	name
Concert	location	name
Concert	song	title
Format numérique	phototag	name
Personne	artist	computed_public_name
Personne	band	name

(Montreux Jazz Digital Project 2019e)

Une sélection manuelle des identifiants des propriétés de Wikidata a été effectuée, à partir des résultats du tableau 9. Cela a permis de définir quatre identifiants différents décrivant des concepts internes à la plateforme Wikidata. Le tableau 10 offre un aperçu des choix de propriété interne et externe possible sur Wikidata en fonction des colonnes sélectionnées pour le transfert. Ce tableau a été créé à partir des identifiants de propriété interne et externe compris dans le tableau 7 et l'annexe 9. Un exemple du cas pratique *Direct Mapping* en langage R2RML et format Turtle (World Wide Web Consortium 2014b), a été adapté à partir des tables du tableau 9 (voir annexe 13).

On constate sur les 11 propriétés que seules 4 propriétés incluent des valeurs décrites par crowdsourcing par les membres de la communauté d'utilisateur de Wikidata. En effet, la majorité des propriétés proposée dans cette grille concerne principalement des propriétés de type *external-id* et fait référence à des identifiants d'autres bases de données musicales.

Tableau 10 : Choix des propriétés sélectionnées sur Wikidata pour un transfert

	instrument (P1303)	inception (P571)	official name (P1448)	title (P1476)	AllMusic artist ID	Discogs artist ID	MusicBrainz artist ID	AllMusic song ID	Spotify track ID	MusicBrainz recording ID	MusicBrainz work ID
composition_title								x	x	x	x
concert_name					x	x	x				
instrument_name	x										
concert_date		x									
song_title								x	x	x	x
location_name			x								
phototag_name				x							
artist_computed_public_name					x	x	x				
band_name					x	x	x				

(Montreux Jazz Digital Project 2019e ; Wikidata:Database reports/List of properties/all 2019)

10.2.2 Préparation du fichier d'export de données

Les neuf fichiers exportés en format csv de la base de données du MJDP sont une « mise à plat » des colonnes sélectionnées dans le tableau 10. À la suite de cet export, chaque ligne correspond à un seul enregistrement et la possibilité d'avoir une combinaison de plusieurs enregistrements par ligne est perdue, comme cela peut être le cas dans une base de données relationnelle.

Le fichier *computed_public_name* est un export de la table *artist* et correspond au nom officiel d'un musicien. Cette colonne est une concaténation du prénom et du nom d'un musicien. Il est composé d'un total de 29'554 enregistrements. Le fichier *band_name* désigne le nom d'une formation musicale incluant un ou plusieurs musiciens et se compose de 2'662 lignes d'enregistrement.

Les fichiers *song_title* et *composition_title* comprennent le titre d'une œuvre musicale. La somme des enregistrements de ces deux fichiers s'élève à 83'277 lignes et représente près de 66 % des enregistrements proposés au versement sur Wikidata.

Le nom et la date des concerts sont indiqués par les intitulés des fichiers *concert_name* et *concert_date* et se compose chacun de 4'902 enregistrements. Un instrument est défini par le fichier *instrument_name* (n = 340) et un lieu géographique (n = 52) par le fichier *location_name*.

Le fichier *phototag_name* (n = 180) désigne un ensemble de termes contrôlés ou thésaurus employé pour le classement des photographies du fonds du MJDP. Les valeurs contenues dans ce fichier concernent des données hétérogènes comme des noms de personnalités, des lieux géographiques ou d'autres concepts liés à la pratique photographique.

L'ensemble des enregistrements cumulés compris dans les neuf fichiers représente un total de 125'870 lignes d'enregistrements. Ce chiffre n'inclut pas les éventuelles données dupliquées ou redondantes, entre les différentes colonnes. Ce chiffre nous donne une indication sur le nombre d'enregistrements potentiellement partageable avec la plateforme Wikidata.

10.3 Préparation d'un versement vers opendata.swiss

10.3.1 Choix et mise en pratique des critères d'évaluation

La sélection des tables et des colonnes en vue d'un versement sur la plateforme opendata.swiss reprend presque la totalité des colonnes définie dans le tableau 9, à l'exception de la colonne *phototag_name*. À la différence de la précédente sélection, cette fois-ci le choix s'arrête au niveau des tables. En effet, aucune action d'alignement des données n'est prévue avec cette plateforme. Les données seront directement rendues accessibles sur le répertoire sans traitement préalable.

10.3.2 Préparation du fichier de description de données partagées

Les fichiers sont exportés dans des fichiers interopérables en format csv. Après une dernière vérification concernant l'encodage et la structure principale des données, les fichiers sont prêts à être partagés sur l'interface de la plateforme opendata.swiss. Le futur répertoire du Montreux Digital Project sur opendata.swiss devrait regrouper 7 jeux de données différents, donnant accès à la consultation de plus de 60 tables de la base de données relationnelle actuelle.

Le transfert des données ne peut être directement effectué vers la plateforme opendata.swiss. En effet, une demande doit au préalable, être formulée par courriel auprès des Archives fédérales suisses (AFS), avant l'ouverture d'un compte institutionnel sur la plateforme. Si la demande est acceptée par les AFS, c'est-à-dire si l'institution remplit le critère concernant la poursuite d'un mandat jugé d'utilité publique en Suisse. Un accès à l'interface de gestion du compte est rendu disponible. L'interface de gestion du compte utilise le Content Management System (CMS) Wordpress.

L'interface de configuration telle que décrite par le manuel d'utilisation de la plateforme (Office fédéral de la statistique, 2016a) permet à l'utilisateur de naviguer dans six menus différents : *Dashboard*, *Medien*, *Datensätze*, *Organisationen*, *Benutzer* et *Werkzeuge*. Cependant, seuls les menus *Benutzer* et *Datensätze* sont évoqués dans le manuel d'utilisateur pour procéder à la description des métadonnées liées au partage d'un jeu de données sur opendata.swiss.

Le menu *Benutzer* permet à l'utilisateur d'accéder à un formulaire permettant de gérer la configuration du compte utilisateur. Le deuxième menu, *Datensätze* offre la possibilité de créer, gérer et publier des métadonnées de description afin de documenter les jeux de données. Le résultat de ces descriptions sera par la suite publié sur la page principale de l'institution sur le portail opendata.swiss.

Le formulaire de versement des métadonnées peut être complété manuellement, depuis un formulaire disponible sur l'interface Wordpress de la plateforme ou alors directement en langage RDF. Cependant, le fichier RDF produit devra impérativement respecter le standard DCAT-AP for Switzerland format (Office fédéral de la statistique 2016b) développé pour interagir avec les données partagées sur la plateforme. Un exemple de la structure des balises RDF est d'ailleurs disponible sur le site de la plateforme (Office fédéral de la statistique 2016c).

L'exemple proposé (voir annexe 12) a été adapté au contexte du Montreux Jazz Digital Project (MJDP). Le fichier a été construit à partir de l'exemple du fichier RDF, décrivant un jeu de métadonnées mis à disposition par l'Office fédéral de topographie swisstopo⁹.

Dans notre exemple (annexe 12) la balise <dc:Catalog> définit un nouveau catalogue. Le préfixe *dc* renvoie au vocabulaire d'ontologie Data Catalog Vocabulary (DCAT) (World Wide Web Consortium 2014a). Ce vocabulaire a été spécialement développé par le World Wide Web Consortium (W3C) afin de faciliter l'interopérabilité entre les catalogues de données sur le web. Le modèle de données du Data Catalog Vocabulary (DCAT), se compose de 4 entités : *Catalog*, *Dataset*, *CatalogRecord* et *Distribution*. La plateforme opendata.swiss utilise uniquement trois des quatre entités décrites dans le modèle : *Catalog*, *Dataset* et *Distribution*.

La classe *Catalog* est composée uniquement de la propriété *dataset*. La classe *Dataset* utilise 4 propriétés des 15 propriétés définies dans le modèle de données DCAT. Il s'agit des propriétés : *contactPoint*, *theme*, *keyword* et *landingPage*. La classe *Dataset* comprend 7 propriétés définies par le vocabulaire d'ontologie DublinCore (Dublin Core Metadata Initiative 2012) sur les 55 propriétés disponibles dans ce modèle. Il s'agit des propriétés : *title*, *description*, *issued*, *publisher*, *language*, *temporal* et *accrualPeriodicity*.

La classe *Distribution* comporte dans cet exemple deux propriétés sur les 11 définies par le modèle de données DCAT. Ces deux propriétés sont *accessURL* et *mediaType*. Deux propriétés du modèle DublinCore sont également présentes dans cette classe *Distribution*, il s'agit des propriétés *issued* et *rights*.

10.4 Alignement avec Wikidata et d'autres bases de données

L'échantillon de départ se compose de quatre colonnes distinctes et comprend dix valeurs « mise à plat », à la suite de l'export des données. Le choix des valeurs a été effectué manuellement, mais aléatoirement¹⁰ dans chaque colonne. Les valeurs des colonnes présentées dans l'échantillon sont extraites du fichier précédemment décrit dans le chapitre 10.2.2.

Les deux cas d'utilisations *Choix entité* et *Choix entité et propriété* seront discutés à la suite de l'alignement semi-automatique avec Wikidata.

La valeur des données de la colonne *instrument_name* de l'échantillon a été identifiée par l'API Wikidata reconciliation comme étant des concepts proches de l'entité : *human(Q5)*. Cette même proposition d'entité a également été proposée pour les valeurs des colonnes *artist_computed_public_name* et *band_name*. L'entité *single(Q134556)* est apparue pour les valeurs de la colonne *song_title*.

Dans les colonnes *instrument_name* et *song_title*, les choix proposés automatiquement par l'API n'ont pas pu être suivis, car jugés trop éloignés et non similaires aux valeurs comprises dans ces colonnes. Cependant, les propositions de l'API ont pu être appliquées pour les colonnes *artist_computed_public_name* et *band_name*. On constate que les noms de personnes sont généralement reconnus par l'API comme appartenant à l'entité *human(Q5)*.

⁹L'exemple peut être téléchargé depuis la plateforme opendata.swiss : https://handbook.opendata.swiss/samples/ogdch_dcatap_import.rdf

¹⁰ L'ordre alphabétique n'a pas été respecté pour choisir les valeurs comprises dans l'échantillon

Les résultats obtenus dans le tableau 11 indiquent que les valeurs de la colonne *instrument_name* ont une occurrence avec 80 % des valeurs de Wikidata. Les résultats des colonnes *artist_computed_public_name* et *band_name* affichent respectivement un taux d'occurrence de 60 % et 40 %. La colonne *song_title* compte seulement 10 % de résultats similaires.

Tableau 11 : Résultats de l'alignement des données du premier échantillon

artist_computed_public_name	wd_artist_computed_public_name	band_name	wd_band_name
Maythem Ghazi	-	Benjamin Clementine	Q16203707
Julio Padron	-	Cornershop	-
Taboo	Q145480	Fleet Foxes	Q852335
Ben Moore	-	Jim Lauderdale	Q1169621
Jimi Mbaye	Q3844075	Magnus Lindgren Quartet	-
Eric Gardner	-	Robert Plant & The Sensational Space Shifters	-
Julie Hugo	Q3189221	Steve Winwood & his band	-
Professor Longhair	Q918447	Vercoquin	-
Toninho Geraes	Q18708133	Theo Croker	Q7781439
Michael Franti	Q715181	Take off Big band	-

(Montreux Jazz Digital Project 2019f)

Une deuxième tentative d'alignement des colonnes de l'échantillon a été réalisée, à partir des données alignées avec le logiciel OntoRefine. Le but de cette mesure est de découvrir, si les données possèdent des identifiants de base de données externes référencés sur la plateforme Wikidata. Cette mesure correspond au cas d'utilisation *Ajouter une propriété*. Les propriétés de Wikidata de type *external-id* : MusicBrainz artist ID (P434), Discogs artist ID (P1953), AllMusic artist ID (P1728) ont été appliquées aux colonnes *artist_computed_public_name* et *band_name*. Les propriétés MusicBrainz work (P435), AllMusic song (P1730) et Spotify track (P2207) ont été utilisées pour aligner la colonne *song_title*.

Les résultats obtenus à la suite de l'alignement de la colonne *artist_computed_public_name* montrent (tableau 12) que la propriété MusicBrainz artist possède une occurrence avec 90 % des résultats alignés. Ce résultat est de 30 % pour Discogs artist ID et AllMusic artist ID. On constate également deux identifiants MusicBrainz ID différents pour un même enregistrement dans la colonne *artist_computed_public_name*. Il s'agit de la valeur « Michael Franti » qui correspond à deux enregistrements distincts dans la base de données de MusicBrainz. En revanche la valeur « Taboo » possède un seul et unique enregistrement sur MusicBrainz.

Tableau 12: Résultats du cas pratique *Ajouter une propriété* pour la colonne `artist_computed_public_name`

artist_computed_public_name	MusicBrainz artist ID	Discogs artist ID	AllMusic artist ID
-	-	-	-
-	-	-	-
Taboo	a813f779-77e2-42a8-a34a-cbd581fc0313	413557	mn0000013325
Taboo	68aeb6d8-f862-4162-b245-fc4e4f64fa6f	-	-
-	-	-	-
Jimi Mbaye	d42f3094-8274-4c56-8c82-30f579bdf648	-	-
-	-	-	-
Julie Hugo	-	-	-
Professor Longhair	c6177e77-151d-422d-bc87-57ae21d2780e	18394	mn0000369270
Toninho Geraes	-	-	-
Michael Franti	650012e9-f732-4295-936d-fdd99bbc6258	66041	mn0000001241
Michael Franti	e23eac42-f904-47da-8599-823ff431bea4	-	-

(Montreux Jazz Digital Project 2019f)

La deuxième mesure présentée dans ce chapitre est basée sur un échantillon obtenu, à partir de la fusion des résultats exportés de la base de données du MJDP (tableau 13). Cette deuxième mesure donne un aperçu de l'application du cas d'utilisation *Ajouter une propriété*. Les résultats du tableau 13 ont été obtenus, à l'aide de la requête SQL de la figure 16. Cette requête a permis d'effectuer la réunion des colonnes concernant le nom officiel d'un musicien en fonction d'un instrument de musique utilisé par ce même musicien. Une requête similaire a été effectuée afin d'extraire les résultats des colonnes associant les noms des groupes (*band_name*) en fonction des noms des instruments (*band_instrument_name*) joués par les membres de ce groupe. Le résultat de la première requête comprend un total de 21'316 enregistrements uniques et celui de la seconde 16'331. Une sélection manuelle et aléatoire (non alphabétique) a permis de constituer un échantillon comprenant un total de dix enregistrements visibles dans le tableau 13.

Figure 16 : Équijointure des noms d'artistes et des instruments de musique

```
SELECT DISTINCT artist.computed_public_name AS artist_name,
map_concert_artist.instrument AS instrument_name
FROM artist
INNER JOIN map_concert_artist ON map_concert_artist.artist_id
= artist.id
```

Tableau 13 : Valeurs initiales du cas pratique *Choix entité et propriété*

artist_name	artist_instrument_name	band_name	band_instrument_name
Adrian Cunningham	Saxophone	James Vincent McMorrow	keyboards
Al Green	lead voc	The Good, The Bad And The Queen	guitar
Aram Bajakian	Guitar	Norah Jones	keyboards
Baptiste Trotignon	piano	Of Monsters and Men	guitar
Bastian Baker	Guitare	The World Quintet	clarinet
Joy Denalane	vocals	The Whip	bass
Karl Denson	sax	The Two	Voice/Guitar
Lily Allen	vocals	The Temperance Movement	Drums
Luca Leombruni	bass	The Roots	Trumpet
Lucien Zerad	guitar	The Egg	drums

La technique employée pour la réconciliation semi-automatique des données du tableau 13 est similaire, à la mesure discutée précédemment. Cependant, les colonnes comprenant des noms d'instruments ont été ajoutées, à l'aide d'une propriété de Wikidata pour contextualiser et préciser la valeur du terme recherché. Cette propriété est ajoutée, dans notre échantillon, à la valeur de l'entité human (Q5). La propriété P1303 définit la valeur d'un instrument de musique sur la plateforme Wikidata et a été choisie pour joindre les colonnes *artist_name* et *artist_instrument_name*, puis *band_name* à *band_instrument_name*. Enfin, l'alignement des identifiants des bases de données externes (tableau 14) a été effectué comme précédemment dans le cas *Ajouter une propriété*.

Tableau 14 : Résultats du cas pratique *Ajouter une propriété* pour la colonne *artist_computed_public_name*

artist_name	MusicBrainz artist ID	Discogs artist ID	AllMusic artist ID
-	-	-	-
Q313260	fb7272ba-f130-4f0a-934d-6eeea4c18c9a	25261	mn0000607448
Q15126523	-	-	-
Q775611	51f3cf89-9a44-4313-8633-13a6584c3504	690626	mn0000146606
Q810599	d8477d11-b8f7-44ae-aa2f-be2a1dccc442	-	mn0002793827
Q543531	05ec340f-c772-4995-9256-092c6931d098	28044	mn0000293700
Q1700847	e841dc87-7d2f-4ee2-ac70-e0941a7954c4	12629	mn0000854407
Q58735	6e0c7c0e-cba5-4c2c-a652-38f71ef5785d	554419	mn0000562850
-	-	-	-
-	-	-	-

La dernière section de ce travail a proposé une méthode d'évaluation en trois temps basée sur les critères quantifiables et non quantifiables définis, à la suite du chapitre 7.4. La démonstration de cette méthode a montré qu'il était possible d'établir une sélection restreinte et efficace des tables et des colonnes de la base de données relationnelle utilisée en vue d'un versement sur une plateforme de crowdsourcing.

La méthode décrite devrait être transposable dans d'autres contextes similaires ou éloignés à celle d'une archive audiovisuelle sans trop de modifications majeures. Cependant, l'approche pratique décrite est fortement dépendante du contexte technique, de la volumétrie et de la complexité de la base de données relationnelles.

Le résultat obtenu montre avant tout que l'ordre de grandeur des données partageables à la suite de l'évaluation se situe aux alentours de 5 % de l'ensemble des données dans ce contexte. Il en résulte qu'un effort important de préparation et de conditionnement des données doit être effectué en amont avant un transfert vers une plateforme de crowdsourcing.

Pour le moment aucun versement pratique n'a encore été effectué vers la plateforme Wikidata, il est donc impossible de discuter de ce point en particulier. La mesure de l'impact d'un versement n'a pas pu être effectuée sur ce fonds en particulier.

L'alignement manuel d'un échantillon de données du MJDP avec la plateforme Wikidata a donné des résultats encourageants, quant à un futur alignement automatique des données. Les résultats ont montré de très bonnes occurrences pour des concepts uniques relatifs à des noms de personne, d'instruments ou encore des lieux géographiques. Le choix manuel de concept à l'aide de propriété paramétrée par l'utilisateur augmente sensiblement le score de l'indice de similarité¹¹ entre les données comparées. En revanche, le système est peu performant concernant un alignement mélangeant divers concepts dans une même requête.

¹¹ Aussi connu en statistique sous l'appellation indice et distance de Jaccard.

11. Conclusion

Ce travail de recherche a permis de mettre en exergue quelques éléments nécessaires à la réussite d'un futur alignement vers Wikidata pour les archives audiovisuelles du MJDP. Ces différents points sont les suivants :

- La création d'un identifiant unique sur Wikidata afin de centraliser l'ensemble des ressources concernant les archives du MJF sur cette plateforme.
- La mise à disposition en local d'un identifiant unique pour chaque enregistrement de la base de données du MJDP.
- Un alignement de la base de données relationnelle du MJDP vers le modèle Wikibase afin de faciliter la synchronisation des données.
- L'alignement à partir de Wikidata avec d'autres thésaurus et bases de données spécialisées dans le domaine musical.
- La combinaison de plusieurs jeux de données hétérogènes afin de préciser le contexte d'une information comme présenté dans le chapitre 10.4 avec le cas *Choix entité et propriété*.
- Le partage de données contrôlées et inédites sur Wikidata.

Les aspects pratiques du versement et de la synchronisation des données avec Wikidata sont les principales limitations de ce travail. La possibilité d'un alignement avec d'autres ontologies spécifiques à la documentation comme par exemple BIBFRAME ou CIDOC-CRM n'a pas été effectuée dans ce travail.

La publication prochaine de la version définitive du standard de description Records in Contexts (RiC) dans le milieu des archives ouvre de nombreuses perspectives. La première de ces perspectives est sans doute une facilitation du partage et de la diffusion des données liées aux notices d'inventaire entre les centres d'archives, grâce à la standardisation des vocabulaires d'ontologies. Puis, dans un second temps l'alignement de ces données avec d'autres ontologies GLAM, dans le but d'unifier les problèmes liés à l'hétérogénéité entre les données comme le formalise Bergman (Bergman 2006) dans une classification des sources d'hétérogénéités ou encore comme le souligne Zeng et Llanes-Padrón.

« Yet another complicated situation is that their documentation has followed special professional standards and best practices defined and implemented in different professions involving libraries, archives, museums, and other cultural heritage institutions for a long time. »
(Zeng 2019, p.28)

« More and more theoretical and practical lines are being opened up which separate out the work procedures in archives, libraries, museums and other cultural institutions. There is no doubt that conceptual models, ontologies and formats are the way forward to compatibility and total interoperability of heritage and cultural information in the context of the semantic web. »

(Llanes-Padrón et Pastor-Sánchez 2017, p.19)

Le développement d'un nouveau langage le Shapes Constraint Language (SHACL) devrait faciliter l'alignement entre des concepts similaires décrit entre des ontologies issues d'un même domaine. En effet, SHACL facilite la construction de règles d'inférence à l'intérieur des requêtes SPARQL (Knublauch 2017).

*« [...] SHACL shape graphs are used to validate that data graphs satisfy a set of conditions they can also be viewed as a description of the data graphs that do satisfy these conditions. »
(World Wide Web Consortium 2017)*

La seconde perspective offerte par l'utilisation de l'ontologie RiC et du web sémantique est sans doute, une amélioration de la recherche d'informations, grâce à l'utilisation de vocabulaire contrôlé et l'association d'éléments de description interne et/ou externe au dépôt d'archives. Ces éléments améliorent significativement la description du contexte et du contenu d'un document.

Ce nouveau changement de paradigme dans la description archivistique offre également une plus grande flexibilité, quant au stockage et à la gestion des notices. Ces notices peuvent alors être hébergées dans des bases de données de graphe en local ou distantes dans le cloud. L'avenir du stockage des Linked Open Data (LOD) est-il à la centralisation ou au contraire à la décentralisation des entrepôts de données ? L'étude de la plateforme de crowdsourcing Wikidata et des exemples tirés des répertoires opendata.swiss ou encore datahub ont montré que les deux solutions étaient actuellement envisageables d'un point de vue technique et technologique.

La très répandue base de données relationnelle MySQL utilisée dans bon nombre de contexte GLAM propose dans une version récente (Oracle 2019), la possibilité de gérer simultanément une base de données relationnelle et une base de données de graphe. Cette solution hybride ne serait-elle pas un excellent compromis en faveur d'un transfert progressif des notices de description archivistiques vers l'ontologie RiC et le LOD ? Une alternative évitant de ce fait un transfert massif de données vers des plateformes et autres répertoires de données LOD distants ? Cette solution permettrait de bénéficier des avantages importants introduit par le Linked Data (LD) avec un alignement possible à d'autres bases de données de graphes centralisées et/ou décentralisées distantes.

Aujourd'hui, tous les scénarios liés aux partages des LOD restent ouverts dans le domaine GLAM. Cependant, comme démontré dans ce mémoire, les contraintes technologiques et techniques ne semblent plus à l'heure actuelle être un frein majeur, à l'accomplissement du Linked Data et du Linked Open Data.

Bibliographie

ARCHIVES FÉDÉRALES SUISSES et al., 2014. *aLOD.ch* [en ligne]. 2019. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://www.alod.ch>

ARISTARÁN, Manuel, et al. 2018. *Tabula* [en ligne]. 2018. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://tabula.technology>

ASSOCIATION DES ARCHIVISTES SUISSES, 2017. *Journée suisse des archives 2017 – Aperçu* [en ligne]. 2017. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://vsa-aas.ch/fr/journee-des-archives-2017/archivtag-2017-uebersicht/>

BERGMAN, Michael K., 2006. *Sources and Classification of Semantic Heterogeneities. AI3: Adaptive Information* [en ligne]. 6 juin 2006. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://www.mkbergman.com/232/sources-and-classification-of-semantic-heterogeneities/>

Commons:GLAM. *Wikimedia Commons* [en ligne]. Dernière modification de la page le 30 Janvier 2019 à 14:10. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://commons.wikimedia.org/wiki/Commons:GLAM>

DESHAYES, Benoît, 2019. *Crotos* [en ligne]. Dernière modification de la page le 17 mai 2019. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://zone47.com/crotos/>

DUBLIN CORE METADATA INITIATIVE, 2012. *DCMI Metadata Terms* [en ligne]. 14 juin 2012. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://dublincore.org/specifications/dublin-core/dcmi-terms/2012-06-14/?v=terms>

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, 2018. *MetaMedia Center MMC* [en ligne]. 14 décembre 2018. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://metamedia.epfl.ch/>

EUROPEAN BROADCASTING UNION, [sans date]. *Ontology* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.ebu.ch/metadata/ontologies/ebucore/>

EUROPEANA FOUNDATION, [sans date]. *Europeana Data Model* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://pro.europeana.eu/resources/standardization-tools/edm-documentation>

INTERNATIONAL COMMITTEE FOR DOCUMENTATION, 2011. *Definition of the CIDOC Conceptual Reference Model* [en ligne]. Novembre 2011. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://www.cidoc-crm.org/html/5.0.4/cidoc-crm.html>

INTERNATIONAL COUNCIL ON ARCHIVES, 2018. *Sémantisation et visualisation de métadonnées archivistiques : mise en ligne du prototype français PIAAF* [en ligne]. 4 mai 2018. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.ica.org/fr/semantisation-et-visualisation-de-metadonnees-archivistiques-prototype-PIAAF>

INTERNATIONAL COUNCIL ON ARCHIVES, 2016. *Records in Contexts - Conceptual Model* [en ligne]. 29 août 2016. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.ica.org/sites/default/files/RiC-CM-0.1.pdf>

GITHUB, 2019. *Graphviz* [en ligne]. 5 juillet 2019. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://github.com/xfr6/graphviz>

GOOGLE, [sans date]. *OpenRefine* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://openrefine.org/>

FREIE UNIVERSITÄT BERLIN, 2012. *The D2RQ Mapping Language* [en ligne]. 12 mars 2012. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://d2rq.org/d2rq-language>

MEDIAAREA, 2019. *MediaInfo* [en ligne]. 2019. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://mediaarea.net/fr/MediaInfo>

ONTOTEXT, [sans date]. *Loading data using OntoRefine* [en ligne]. 19 juin 2019. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://graphdb.ontotext.com/documentation/free/loading-data-using-ontorefine.html>

ORACLE, 2019. *MySQL Document Store* [en ligne]. 2019. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.mysql.com/products/enterprise/document_store.html

PBCORE, [sans date]. *PBCore* [en ligne]. 2019. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://pbcore.org/>

PROJECT JUPYTER, 2019. *Jupyter* [en ligne]. 12 avril 2019. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://jupyter.org/>

RAIMOND, Yves et al. 2013. *The Music Ontology – Specification* [en ligne]. 22 juillet 2013. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://musicontology.com/specification/>

SCHEMA.ORG, [sans date]. *Data Model* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://schema.org/docs/datamodel.html>

SPIN, [sans date]. *SPIN SPARQL Inferencing Notation* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://spinrdf.org/>

STANFORD UNIVERSITY et al., [sans date]. *PerformedMusicOntology* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://performedmusicontology.org/ontologies/PerformedMusicOntology.html>

UNIVERSITY OF ILLINOIS, [sans date]. *Preservation Self-Assessment Program* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://psap.library.illinois.edu/collection-id-guide#audiovisual>

VOCAB.ORG, 2010. *RELATIONSHIP: A vocabulary for describing relationships between people* [en ligne]. 19 avril 2010. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://vocab.org/relationship/>

WOLFRAMALPHA, 2019. *Wolframalpha* [en ligne]. 2019. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.wolframalpha.com/input/?i=1+inch>

Creative commons

CREATIVE COMMONS, [sans date a]. *CC0 1.0 Universal (CC0 1.0) Public Domain Dedication* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://creativecommons.org/publicdomain/zero/1.0/>

CREATIVE COMMONS, [sans date b]. *Describing Copyright in RDF* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://creativecommons.org/ns>

Internet archive

INTERNET ARCHIVE, PURL administration [sans date a]. *PURL Administration* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://archive.org/services/purl/>

INTERNET ARCHIVE, PURL administration [sans date b]. *Records in Contexts Ontology* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://skos.um.es/TR/ric/>

Illustration, graphique

MONTREUX JAZZ DIGITAL PROJECT, 2019a. Inventaire typologique des supports physiques : constituer à partir de statistiques de la base de données du MJDP. 5 Mai 2019.

MONTREUX JAZZ DIGITAL PROJECT, 2019b. Distribution des supports magnétiques : constituer à partir de statistiques de la base de données du MJDP. 5 Mai 2019.

MONTREUX JAZZ DIGITAL PROJECT, 2019c. Format des enregistrements numériques recensés dans le fonds : constituer à partir de statistiques de la base de données du MJDP. 5 Mai 2019.

MONTREUX JAZZ DIGITAL PROJECT, 2019d. Distribution du nombre de tables sélectionnées par catégories à l'issue de chaque tour de choix : constituer à partir des tables et colonnes de la base de données du MJDP. 5 Mai 2019.

MONTREUX JAZZ DIGITAL PROJECT, 2019e. Tables et colonnes sélectionnées à partir de la base de données du MJDP pour un transfert de données sur la plateforme Wikidata : constituer à partir des tables et colonnes de la base de données du MJDP. 5 Mai 2019.

MONTREUX JAZZ DIGITAL PROJECT, 2019f. Fichier au format csv extrait de la base de données du MJDP. 5 Mai 2019.

Library of congress

LIBRARY OF CONGRESS, 2017. *BIBFRAME Ontology* [en ligne]. 4 mai 2017. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://id.loc.gov/ontologies/bibframe.html>

LIBRARY OF CONGRESS, 2015. *BIBFRAME AV Assessment: Technical, Structural, and Preservation Metadata* [en ligne]. 23 septembre 2015. Mise à jour le 4 janvier 2016. [Consulté le 8 juillet 2019]. Disponible à l'adresse : www.loc.gov/bibframe/docs/pdf/bf-avtechstudy-01-04-2016.pdf

LIBRARY OF CONGRESS, 2019. *Sustainability of Digital Formats: Planning for Library of Congress Collections* [en ligne]. 23 avril 2019. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.loc.gov/preservation/digital/formats/index.html>

MediaWiki

API:Main page. *MediaWiki* [en ligne]. Dernière modification de la page le 29 novembre 2018 à 22:35. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.mediawiki.org/wiki/API:Main_page

Extension:Wikibase Client. *MediaWiki* [en ligne]. Dernière modification de la page le 25 juin 2019 à 12:05. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.mediawiki.org/wiki/Extension:Wikibase_Client

Extension:Wikibase Repository. *MediaWiki* [en ligne]. Dernière modification de la page le 25 juin 2019 à 12:12. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.mediawiki.org/wiki/Extension:Wikibase_Repository

Manuel:Pywikibot. *MediaWiki* [en ligne]. Dernière modification de la page le 13 juin 2019 à 11:21. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.mediawiki.org/wiki/Manual:Pywikibot/fr>

PAWS. *MediaWiki* [en ligne]. Dernière modification de la page le 11 mai 2019 à 16:13. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.mediawiki.org/wiki/PAWS>

Wikibase. *MediaWiki* [en ligne]. Dernière modification de la page le 27 avril 2019 à 17:29. [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.mediawiki.org/wiki/Wikibase>

Wikibase/API. *MediaWiki* [en ligne]. Dernière modification de la page le 27 mai 2019 à 11:43. [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.mediawiki.org/wiki/Wikibase/API>

Wikibase/DataModel. *MediaWiki* [en ligne]. Dernière modification de la page le 29 mai 2019 à 05:16. [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.mediawiki.org/wiki/Wikibase/DataModel>

Wikibase/DataModel/JSON. *MediaWiki* [en ligne]. Dernière modification de la page le 5 juillet 2019 à 11:43. [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.mediawiki.org/wiki/Wikibase/DataModel/JSON>

Wikibase/DataModel/Primer. *MediaWiki* [en ligne]. Dernière modification de la page le 1 août 2018 à 14:02. [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer>

Wikibase/Indexing/RDF Dump Format. *MediaWiki* [en ligne]. Dernière modification de la page le 2 mai 2019 à 21:49. [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format

Office fédéral de la statistique

OFFICE FÉDÉRAL DE LA STATISTIQUE, opendata.swiss [sans date]. *opendata.swiss* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://opendata.swiss/fr>

OFFICE FÉDÉRAL DE LA STATISTIQUE, opendata.swiss, 2016a. *Benutzerhandbuch opendata.swiss* [en ligne]. 20 juin 2016. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://handbook.opendata.swiss/de/library/opendataswiss-userguide.html>

OFFICE FÉDÉRAL DE LA STATISTIQUE, opendata.swiss, 2016b. *DCAT-AP for Switzerland format* [en ligne]. 9 février 2016. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://handbook.opendata.swiss/en/library/ch-dcat-ap.html>

OFFICE FÉDÉRAL DE LA STATISTIQUE, opendata.swiss, 2016c. *Harvester* [en ligne]. 3 août 2016. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://handbook.opendata.swiss/de/publish/harvester.html>

Open knowledge foundation

OPEN KNOWLEDGE FOUNDATION, OpenGLAM, 2013. *OpenGLAM Principles* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://openglam.org/principles/>

OPEN KNOWLEDGE FOUNDATION, datahub, [sans date a]. *datahub* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://old.datahub.io/>

OPEN KNOWLEDGE FOUNDATION, Opendata.ch [sans date b]. *Opendata.ch* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://fr.opendata.ch/>

OPEN KNOWLEDGE FOUNDATION, Opendata.ch [sans date c]. *Open GLAM Working Group* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://glam.opendata.ch>

Articles

ALLISON-CASSIN, Stacy, SCOTT, Dan, 2018. Wikidata : a platform for your library's linked open data [en ligne]. 4 mai 2018. *Code4lib Journal*. Vol. 40. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://journal.code4lib.org/articles/13424>

BROWN, Tim, 2008. Design Thinking [en ligne]. 2008. Harvard business review. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://fusesocial.ca/wp-content/uploads/sites/2/2018/06/Design-Thinking.pdf>

CONSEIL FÉDÉRALE, 2018. Stratégie en matière de libre accès aux données publiques en Suisse pour les années 2019 à 2023 [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.admin.ch/opc/fr/federal-gazette/2019/855.pdf>

COOEY, Nancy, 2019. Leveraging Wikidata to Enhance Authority Records in the EHRI Portal. 2019. *Journal of Library Metadata*. Vol. 19 pp. 83-98. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://doi.org/10.1080/19386389.2019.1589700>

DE BOER, Victor et al., 2019. The Benefits of Linking Metadata for Internal and External Users of an Audiovisual Archive. In: Metadata and Semantic Research. *Springer International Publishing*, pp. 212-223. Communications in Computer and Information Science. ISBN 978-3-030-14401-2.

DIMOU, Anastasia et al., 2014. RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data. In: Proceedings of the 7th Workshop on Linked Data on the Web (LDOW2014), Seoul, Korea [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : http://events.linkedata.org/ldow2014/papers/ldow2014_paper_01.pdf

DUFAUX, Alain, et AMSALLEM, Thierry, 2017. Le Montreux Jazz digital project : La sauvegarde des archives audiovisuelles du Montreux Jazz Festival, un patrimoine pour l'innovation et la recherche. *Ressi* [en ligne]. 20 décembre 2017. [Consulté le 8 juillet 2019]. Disponible à l'adresse : http://www.ressi.ch/num18/article_138

ESTERMANN, Beat, 2018. Development paths towards open government – an empirical analysis among heritage institutions. 2018. *Government Information Quarterly*. Vol. 35. n°4, pp. 599-612. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://doi.org/10.1016/j.giq.2018.10.005>

GASSER, Michael, 2019. Commons-Massupload von Digitalisaten aus dem Bildarchiv der ETH-Bibliothek: Umsetzung, Erfahrungen, Ausblick [en ligne]. 28 mars 2019. *Research Collection*. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://doi.org/10.3929/ethz-b-000334687>

HYVÖNEN, Eero, 2016a. Cultural Heritage Linked Data on the Semantic Web: Three Case Studies Using the Sampo Model. In : ESPAGNE. VIII Encounter of Documentation Centres of Contemporary Art: *Open Linked Data and Integral Management of Information in Cultural Centres Artium, Vitoria-Gasteiz October 19-20, 2016* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://seco.cs.aalto.fi/publications/submitted/hyvonon-vitoria-2017.pdf>

HYVÖNEN, Eero et al., 2016b. *Publishing Second World War History as Linked Data Events on the Semantic Web* [en ligne]. 2016. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://seco.cs.aalto.fi/publications/2016/hyvonon-et-al-warsa-dh2016.pdf>

KNUBLAUCH, Holger, 2017. *From SPIN to SHACL* [en ligne]. 1 août 2017. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.spinrdf.org/spin-shacl.html>

LLANES-PADRÓN, Dunia et PASTOR-SÁNCHEZ, Juan-Antonio, 2017. Records in contexts: the road of archives to semantic interoperability. *Program*. 10 octobre 2017. Vol. 51, n° 4. pp. 387-405. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://doi.org/10.1108/PROG-03-2017-0021>

OFFICE FÉDÉRALE DE LA CULTURE, Memoriav, 2017. *Recommandations Memoriav L'archivage numérique des films et vidéos*. 2017. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://memoriav.ch/video/recommandations-video/archivage-numerique-des-films-et-des-vidéos/?lang=fr>

SCHIAVONE, Luisa et al., 2018. The CoBiS Linked Open Data Project and Portal. *EPJ Web of Conferences*. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.epj-conferences.org/articles/epjconf/pdf/2018/21/epjconf_lisaviii2018_12013.pdf

THORNTON, Katherine et al., 2017. Modeling the Domain of Digital Preservation in Wikidata. 2017. *ACM International Conference Proceeding Series*. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://production-scientifique.bnf.fr/sites/default/files/7katherine-thornton.pdf>

VANDEBUSSCHE, Pierre-Yves et al., 2015. Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. *Stanford Center for Biomedical Informatics Research* [en ligne]. 2015. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://www.semantic-web-journal.net/system/files/swj1178.pdf>

VAN VEEN, Theo, 2019. Wikidata: From “an” Identifier to “the” Identifier. *Information technology and libraries* [en ligne]. 2019. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://ejournals.bc.edu/index.php/ital/article/download/10886/9483/>

VRANDEČIĆ, Denny, 2013. The Rise of Wikidata. 2013. *IEEE Intelligent Systems*. Vol. 28. n°4. pp. 90-95. ISSN 1541-1672. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <10.1109/MIS.2013.119>

ZENG, Marcia Lei, 2019. Semantic enrichment for enhancing LAM data and supporting digital humanities. *El profesional de la información*, Vol. 28, n° 1. eISSN: 1699-2407 [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://doi.org/10.3145/epi.2019.ene.03>

Wikimedia

WIKIMEDIA, [sans date a]. *PetScan* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://petscan.wmflabs.org/>

WIKIMEDIA, [sans date b]. *Phabricator* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://phabricator.wikimedia.org/>

WIKIMEDIA, [sans date c]. *MediaWiki API help* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.wikidata.org/w/api.php>

WIKIMEDIA, [sans date d]. *Wikimedia Downloads* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://dumps.wikimedia.org/>

WIKIMEDIA, [sans date e]. *Wikimedia Foundation* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://wikimediafoundation.org/>

WIKIMEDIA, [sans date f]. *Wikimedia REST API* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://wikimedia.org/api/rest_v1/#/

WIKIMEDIA, [sans date g]. *Wikimedia Toolforge* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/>

Wikimedia Meta-Wiki

Mix'n'match/Manuel. *Wikimedia Meta-Wiki* [en ligne]. Dernière modification de la page le 30 juin 2018 à 15:10. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://meta.wikimedia.org/wiki/Mix%27n%27match/Manual/fr>

PetScan. *Wikimedia Meta-Wiki* [en ligne]. Dernière modification de la page le 8 décembre 2018 à 09:23. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://meta.wikimedia.org/wiki/PetScan/fr>

Wikimedia CH. *Wikimedia Meta-Wiki* [en ligne]. Dernière modification de la page le 16 juin 2019 à 20:30. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://meta.wikimedia.org/wiki/Wikimedia_CH

Wikimedia Outreach

GLAM. *Wikimedia Outreach* [en ligne]. Dernière modification de la page le 10 avril 2019 à 08:43. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://outreach.wikimedia.org/wiki/GLAM>

GLAM/Resources/Data and media partnerships workflow. *Wikimedia Outreach* [en ligne]. Dernière modification de la page le 22 mars 2019 à 17:53. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://outreach.wikimedia.org/wiki/GLAM/Resources/Data_and_media_partnerships_workflow

Wikimedia Toolforge

WIKIMEDIA TOOLFORGE, [sans date a]. *BaGLAMa 2* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/glamtools/baglama2/>

WIKIMEDIA TOOLFORGE, [sans date b]. *GLAMorgan* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/glamtools/glamorgan.html>

WIKIMEDIA TOOLFORGE, [sans date c]. *Mix'n'match* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/mix-n-match/#/>

WIKIMEDIA TOOLFORGE, [sans date d]. *OpenRefine-Wikidata interface* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/openrefine-wikidata>

WIKIMEDIA TOOLFORGE, [sans date e]. *QuickStatements* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/quickstatements/#/>

WIKIMEDIA TOOLFORGE, [sans date f]. *Wikimedia toolforge* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/>

WIKIMEDIA TOOLFORGE, [sans date g]. *WikiProject Corfu* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/projector/#/Corfu>

Wikidata

WIKIDATA, [sans date a]. *Wikidata Query Service* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://query.wikidata.org/>

WIKIDATA, [sans date b]. *Wikidata Query Service* [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://query.wikidata.org/sparql>

Wikidata:Database reports/List of properties/all. *Wikidata* [en ligne]. Dernière modification de la page le 20 juin 2019 à 15:05. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.wikidata.org/wiki/Wikidata:Database_reports/List_of_properties/all

Wikidata:Events/GLAM & Wikimedia CH. *Wikidata* [en ligne]. Dernière modification de la page le 29 juin 2019 à 16:41. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.wikidata.org/w/index.php?title=Wikidata:Events/GLAM_%26_Wikimedia_CH

Wikidata:Requests for permissions/Bot. *Wikidata* [en ligne]. Dernière modification de la page le 4 juillet 2019 à 13:01. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.wikidata.org/wiki/Wikidata:Requests_for_permissions/Bot

Wikidata:SPARQL query service/queries/examples. *Wikidata* [en ligne]. Dernière modification de la page le 29 juin 2019 à 18:07. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples

Wikidata:Tools. *Wikidata* [en ligne]. Dernière modification de la page le 18 janvier 2019 à 11:57. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.wikidata.org/wiki/Wikidata:Tools>

Wikidata:WikiProjects. *Wikidata* [en ligne]. Dernière modification de la page le 27 mai 2019 à 18:31. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.wikidata.org/wiki/Wikidata:WikiProjects>

Wikidata:WikiProject Cultural heritage. *Wikidata* [en ligne]. Dernière modification de la page le 5 juillet 2019 à 17:28. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.wikidata.org/wiki/Wikidata:WikiProject_Cultural_heritage

Wikidata:WikiProject Music. *Wikidata* [en ligne]. Dernière modification de la page le 31 mai 2019 à 11:06. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.wikidata.org/wiki/Wikidata:WikiProject_Music

Wikidata:WikiProject Ontology/Mapping. *Wikidata* [en ligne]. Dernière modification de la page le 13 novembre 2019 à 08:13. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.wikidata.org/wiki/Wikidata:WikiProject_Ontology/Mapping

Wikidata:WikiProject Performing arts/Reports/Ingesting Production Databases of the Performing Arts. *Wikidata* [en ligne]. Dernière modification de la page le 30 juin 2019 à 06:53. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.wikidata.org/wiki/Wikidata:WikiProject_Performing_arts/Reports/Ingesting_Production_Databases_of_the_Performing_Arts

Wikipédia

Comparison of on-demand music streaming services. Wikipédia : l'encyclopédie libre [en ligne]. Dernière modification de la page le 8 juillet 2019 à 00:50. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://en.wikipedia.org/w/index.php?title=Comparison_of_on-demand_music_streaming_services&oldid=905266680

List of online digital musical document libraries. Wikipédia : l'encyclopédie libre [en ligne]. Dernière modification de la page le 4 mai 2019 à 14:20. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://en.wikipedia.org/w/index.php?title=List_of_online_digital_musical_document_libraries&oldid=895474660

List of online music databases. Wikipédia : l'encyclopédie libre [en ligne]. Dernière modification de la page le 8 juillet 2019 à 00:04. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://en.wikipedia.org/w/index.php?title=List_of_online_music_databases&oldid=905261415

Matrice d'adjacence. Wikipédia : l'encyclopédie libre [en ligne]. Dernière modification de la page le 17 février 2019 à 15:00. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://fr.wikipedia.org/wiki/Matrice_d'adjacence

World Wide Web Consortium

WORLD WIDE WEB CONSORTIUM, 2010. *Use Cases and Requirements for Mapping Relational Databases to RDF* [en ligne]. 2 juin 2010. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.w3.org/2001/sw/rdb2rdf/use-cases/>

WORLD WIDE WEB CONSORTIUM, 2012a. *Ontology for Media Resources 1.0* [en ligne]. 9 février 2012. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.w3.org/TR/mediaont-10/>

WORLD WIDE WEB CONSORTIUM, 2012b. *R2RML: RDB to RDF Mapping Language* [en ligne]. 27 septembre 2012. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.w3.org/TR/r2rml/>

WORLD WIDE WEB CONSORTIUM, 2014a. *Data Catalog Vocabulary (DCAT)* [en ligne]. 16 janvier 2014. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.w3.org/TR/vocab-dcat/>

WORLD WIDE WEB CONSORTIUM, 2014b. *RDF 1.1 Turtle* [en ligne]. 25 février 2014. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.w3.org/TR/turtle/>

WORLD WIDE WEB CONSORTIUM, 2017. *Shapes Constraint Language (SHACL)* [en ligne]. 20 juillet 2017. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.w3.org/TR/shacl/>

Annexe 1 : « GLAM & Wikimedia » 2019 à Berne

La journée professionnelle « GLAM & Wikimedia » s'est déroulée le 28 mars 2019, à la bibliothèque nationale suisse (BN) de Berne. Cette conférence a été organisée conjointement par l'Association des archivistes suisses (AAS) et Wikimedia CH.

Les retours d'expériences de cette journée concernaient des projets GLAM (Galleries, Libraries, Archives and Museums) réalisés en Suisse, mais également à l'étranger dans le cadre d'institutions patrimoniales. Les sujets des conférences ont principalement décrit des projets de valorisations avec les outils de Wikimedia : Wikipédia, Commons et Wikidata.

Journée des archives 2017

Gilliane Kern, conseillère en gestion d'information – Association des archivistes suisses (AAS)

Les ateliers organisés dans le cadre de cette journée ont permis la création d'un modèle d'article pour les institutions d'archives sur Wikipédia :

*Cite archive*¹² (Wikipédia 2018). Ce modèle est disponible en trois langues (français, allemand, anglais). Il a été construit à partir des champs de la norme générale et internationale de description archivistique ISAD(G) du conseil international des archives (CIA).

Plus de 50 centres d'archives en suisses ont participé à cette journée. Cette première expérience a permis notamment la création de synergies entre les associations Wikimedia CH, Open Data.ch et l'Association des archivistes suisses (AAS).

Retour d'expérience du Musée cantonal de zoologie, Lausanne

Chantal Ebongué-Pittet, chargée de communication – Musée cantonal de zoologie

Exemple de la valorisation et de la description d'un fonds d'archives photographiques avec les outils Wikipédia et Commons.

L'objectif pour le musée est d'accroître la visibilité de ses collections, auprès du grand public. La formation du photographe a été prise en charge par un membre de l'association Wikimedia CH.

Le choix et la curation des photographies partagées sur Commons a été réalisé suivant deux axes : l'architecture (Palais de Rumine) et la taxidermie (une collection en particulier). Le volume des photographies versé sur Commons est d'environ 500 photographies. Une catégorie concernant la taxidermie et deux catégories sur l'architecture ont été créées sur Commons pour le classement des fichiers versés.

La description des métadonnées des photographies comprend le sujet, le nom du photographe, le support de l'image et le type de licence de partage.

Un projet de téléchargement automatique des collections a été imaginé, mais n'a jamais été réalisé. De nouveaux versements vers la plateforme sont prévus à l'avenir.

¹² Modèle:Cite archive. Wikipédia [en ligne]. Dernière modification de la page le 11 avril 2018 à 18:03. [Consulté le 7 avril 2019]. Disponible à l'adresse : https://fr.wikipedia.org/w/index.php?title=Mod%C3%A8le:Cite_archive&oldid=147430674

La difficulté de suivre l'impact réel de la visibilité des fonds versés sur la plateforme et peu d'implication de la part du personnel du musée sont quelques difficultés rencontrées, au cours du projet.

Retour d'expérience des Archives de l'État de Neuchâtel

Géraldine Galfetti, archiviste – Archives de l'État de Neuchâtel

L'objectif pour les archives de l'État de Neuchâtel est de valoriser la description des notices des fonds, déjà disponibles sur la plateforme institutionnelle et attirer de nouveaux publics via Wikipédia.

La formation à Wikipédia a été acquise en ligne via des tutoriels sur la plateforme et suite à la participation à des *wikipermanences* (ateliers pratiques d'initiations aux outils de Wikimedia).

La création d'une page et d'un compte institutionnel sur la Wikipédia sont conseillés pour authentifier sans ambiguïté la source de la contribution.

Le choix des thématiques traitées dans les articles partagés sur Wikipédia concerne principalement des fonds privés et des données librement consultables sur la plateforme institutionnelle. Le but est d'offrir des articles de vulgarisations sur divers sujets historiques.

L'ajout de liens de références (url) dans les articles sur Wikipédia permet des renvois directs vers la description des notices des fonds d'archives sur le portail des archives de l'État de Neuchâtel. L'édition des *infobox* sur Wikipédia a été effectué avec le modèle *Cite archive*.

Généralement, les articles consultables sur Wikipédia sont subjectifs et/ou orientés par les utilisateurs en fonction des langues. Les contributeurs non institutionnels ont donc une connaissance partielle du contexte historique d'un fonds ou d'un événement. La conséquence est une description lacunaire et non exhaustive des notices. En plus, les articles peuvent potentiellement être mis à jour régulièrement par la communauté. Ce cas de figure n'est pas possible dans un contexte institutionnel car il n'y a pas de révision fréquente des notices de description. L'anonymat de la contribution permet d'élargir la participation à une plus large audience et pas uniquement aux pairs.

Suite à la publication des articles, peu de corrections ont été constatées sur les sujets partagés sur la plateforme. Certains sujets ont suscité des discussions au sein de la communauté sans pour autant provoquer de polémiques.

Quelques projets futurs :

Poursuivre la contribution des archives sur la plateforme.

Proposer des traductions semi-automatiques des articles en français, via l'outil *Traduction de contenu* (MediaWiki 2018).

Versements de nouvelles photographies vers Commons.

Dans Wikidata décrire les entités des fonds disponibles sur Wikipédia en français et effectuer des liens vers la propriété *archivé par* (P485).

Poursuivre les collaborations avec l'Université de Neuchâtel concernant la rédaction d'articles sur Wikipédia par des étudiants.

Commons-Massenupload von Digitalisaten aus dem Bildarchiv der ETH-Bibliothek: Umsetzung, Erfahrungen, Ausblick

Michael Gasser, ETH-Bibliothek / Archiv

Plus de 130'000 photographies sous licences Creative Commons CC0 ont été versées automatiquement sur la plateforme Commons.

La plateforme *e-pics* de l'ETH-Bibliothek utilise les licences Creative Commons CC0 et CC-BY-SA 4.0.

Le suivi de l'impact des collections partagées est effectué avec les outils *Cassandra*¹³ et *BaGLAMa 2*¹⁴.

Projets futurs :

Crowdsourcing sur les images versées sur Commons.

Versement d'un fonds photographique des Archives fédérales suisses (AFS).

Rédaction de recommandations pour la bibliothèque nationale suisse (BN) concernant la numérisation de certains supports photographiques.

Descriptions et intégration des métadonnées de Commons dans Wikidata.

Veranstaltung im PTT-Archiv am Archivtag, Kooperation mit der Uni Bern, Historisches Institut

Diego Hättenschwiler et Jonas Veress, Wikimedia CH et PTT-Archiv

Dans le cadre du projet *Oral history* des archives des PTT, créations et éditions d'articles Wikipédia par des étudiants en histoire de l'Université de Berne.

Le constat est le suivant, suite à la mise en ligne des articles : l'impact est important au moment de la publication des articles (effet de *Buzz*) et accroît significativement la visibilité du projet et des liens, qui pointe vers le projet grâce à l'excellent référencement de Wikipédia (cinquième site mondiale). Cependant, il est difficile de mobiliser une audience sur le long terme. La participation des étudiants aux articles a diminué significativement, suite à la publication des articles sur Wikipédia et la fin du cours.

Organisation d'ateliers Wikipédia en 2015, articles Wikipédia affichés dans l'OPAC

Béatrice Perret Anadi, directrice adjointe – Bibliothèque de la ville de Bienne

Intégration de Wikipédia au catalogue de la bibliothèque de Bienne. Le but est d'enrichir les notices du catalogue de la bibliothèque avec des informations issues de parties d'articles de Wikipédia.

Les articles sont disponibles en allemand, français et sont affichés en fonctions des recherches effectuées dans le catalogue. Il n'y a pas de gestion des homonymes et de nombreuses incohérences sémantiques sont observées dans les résultats de recherches.

¹³ WIKIMEDIA SWITZERLAND, 2019. GLAM stat tool (Cassandra) [en ligne]. [Consulté le 7 avril 2019]. Disponible à l'adresse : <http://stats.wikimedia.swiss/>

¹⁴ WIKIMEDIA, 2019. BaGLAMa 2 [en ligne]. [Consulté le 7 avril 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/glamtools/baglamma2/index.html>

La bibliothèque a également organisé plusieurs *wikipermanences* avec le constat suivant : il est difficile de créer et de fidéliser une communauté d'utilisateurs. L'outil a difficilement été accepté par le personnel de la bibliothèque.

CoBiS : coordinamento delle biblioteche speciali e specialistiche di area metropolitana torinese

Ilario Valdelli, chargé de projet – Wikimedia CH

CoBiS est un réseau de 66 bibliothèques spécialisées dans la région de Turin en Italie.

La problématique principale du projet CoBiS sont les hétérogénéités matérielles et bibliographiques des différents catalogues.

Présentation des étapes du projet :

Transformation des données initiales en triplet RDF. Le choix des modèles de données *BIBFRAME* et *Schema.org* a été décidé pour le mapping, car ils intègrent les réalités métiers et commerciales du web sémantique. Le modèle *BIBFRAME* est développé par la Library of Congress (LOC) et *Schema.org* par un consortium dirigé par Google.

Les données des différentes bases de données disponibles ont été exportées en format csv. Des requêtes en SPARQL ont été développées pour réaliser des conversions en RDF ou JSON. Les libraires en langage Java des projets open source *TARQL* et *JARQL* ont été utilisées. Ces deux projets exploitent l'environnement Apache *Maven*.

Le logiciel de *Virtuoso* a été implémenté pour exploiter des endpoints pour les requêtes SPARQL.

L'alignement des diverses bases de données a été effectué, depuis Wikidata. Les identifiants des œuvres (entités) partagées par d'autres catalogues comme *viaf*, *bnf* et *dbpedia* ont été alignés avec le catalogue de *CoBiS*.

Le résultat est l'intégration de notices biographiques et bibliographiques directement sur le catalogue de la plateforme de *CoBiS*.

Wikidata dans le paysage des données ouvertes et liées

Anne Chardonnens, doctorante – Université libre de Bruxelles

La problématique principale des précédents projets de Wikimedia : une décentralisation des données entre les différents projets. Le projet Wikidata permet une recentralisation des données et la réutilisation d'un concept unique et exploitable par l'ensemble des projets de Wikimedia.

Le modèle de données de Wikidata est composé d'une seule entité. Une entité peut être composée de zéro ou plusieurs propriétés. La structure d'un triplet est la suivante : sujet ; prédicat ; objet.

Le Linked Open Data cloud (LODC) a subi une transformation importante, en l'espace d'une décennie. Dbpedia était la première grande base de données du web sémantique. A présent, Wikidata est devenu la base de données principale du web sémantique. L'objectif de Wikidata est de devenir un hub central pour les données du Linked Open Data (LOD).

Wikidata Query Service permet de développer des requêtes en langage SPARQL et de visualiser directement le résultat dans un navigateur. Ces visualisations peuvent prendre la forme de liste, de galerie d'image, de cartographie ou encore de chronologie.

Les propriétés sont ajoutées suite à l'approbation de la communauté. Elles ne sont jamais figées et peuvent potentiellement évoluer en fonction des itérations successives de la communauté.

Wikimedia, Wikidata et Institutions patrimoniales (GLAM)

Benoît Deshayes, Wikimedia France

Rappel des différents projets de la fondation Wikimedia :

Wikimedia Commons pour le partage de fichier et la description de métadonnées.

Wikipédia pour le partage des informations.

Wikidata pour le partage des données.

Le wikicode permet de décrire le contenu d'une page dans les projets de Wikimedia. Les utilisateurs peuvent s'appuyer sur des modèles existants ou créer un modèle en fonction de leurs besoins.

Le problème de la catégorisation dans Commons : actuellement la situation est peu viable d'où l'idée du projet Wikidata pour la création d'identifiant unique.

Principales limites de Commons : les données sont hétérogènes et non structurées ; les logiques de classement sont manuelles, peu intuitives et fortement dépendantes d'un utilisateur ; l'anglais est la langue principale ; outils de rangement plus que de recherches.

Les éléments de Wikidata dans Wikipédia sont principalement les informations issues des *infobox*.

Wikidata permet la description de réalités contradictoires d'un même sujet : exemple deux entités différentes sur un même sujet, d'où l'importance de vérifier les références.

L'ontologie de Wikidata est collaborative et évolue constamment. Il y a beaucoup de discussions au niveau de la communauté à ce sujet.

Les versements sur Wikidata peuvent être effectués automatiquement, par bot ou manuellement.

Wikidata offre la possibilité de faire des recherches sémantiques à partir des propriétés décrites dans les entités. Par exemple les tableaux qui ont pour sujet la neige. Il est également possible de faire des regroupements de plusieurs versions d'une même œuvre ou encore de rechercher des monuments sur une carte. L'intégration des modules de visualisations *Mirador* et *Universal Viewer* de l'International Image Interoperability Framework (IIIF) seront bientôt disponibles sur la plateforme.

Le résultat des recherches est fortement dépendant de la qualité de la description des entités par les utilisateurs.

Afin d'augmenter la qualité des données partagées tout en évitant le Digital labor, une solution serait d'augmenter la participation rémunérée à Wikidata par des professionnelles GLAM.

Exemple d'un projet de valorisation avec Wikidata: *Crotos*, un metamoteur d'œuvre d'art construit à partir des collections du Louvre et d'autres institutions muséales en Europe.

Bibliographie

CHANTAL.EBONGUE, 2019. Retour d'expérience du Musée cantonal de zoologie, Lausanne [en ligne]. 25 mars 2019. [Consulté le 7 avril 2019]. Disponible à l'adresse : https://commons.wikimedia.org/wiki/File:Journ%C3%A9e_GLAM_28.3.19_%C3%A0_Berne_pr%C3%A9sentation_du_mus%C3%A9e_cantonal_de_zoologie_Lausanne_C._Ebongu%C3%A9.pdf

CHARDONNENS, Anne, 2019. Wikidata dans le paysage des données ouvertes et liées [en ligne]. 27 mars 2019. [Consulté le 7 avril 2019]. Disponible à l'adresse : https://commons.wikimedia.org/wiki/File:GLAMwikiCH_Introduction_Wikidata.pdf

GALFETTI, Géraldine, 2019. Retour d'expérience des Archives de l'État de Neuchâtel [en ligne]. 28 mars 2019. [Consulté le 7 avril 2019]. Disponible à l'adresse : https://commons.wikimedia.org/wiki/File:Galfetti_GLAMwikiCH_pr%C3%A9sentation_wikipedia_2019.pdf

GASSER, Michael, 2019. Commons-Massenupload von Digitalisaten aus dem Bildarchiv der ETH-Bibliothek: Umsetzung, Erfahrungen, Ausblick [en ligne]. ETH-Bibliothek. 28 mars 2019. [Consulté le 7 avril 2019]. Disponible à l'adresse : <https://www.research-collection.ethz.ch/handle/20.500.11850/334687>

HADI, 2019. Das PTT-Archiv und Wikipedia [en ligne]. 25 mars 2019. [Consulté le 7 avril 2019]. Disponible à l'adresse : https://commons.wikimedia.org/wiki/File:PTT-Archiv_-_Uni_Bern_-_Wikipedia.pdf

ILARIO, 2019. A Case Study : CoBiS : coordinamento delle biblioteche speciali e specialistiche di area metropolitana torinese [en ligne]. 28 mars 2019. [Consulté le 7 avril 2019]. Disponible à l'adresse : https://commons.wikimedia.org/wiki/File:A_case_study.pdf

KERN, Gilliane, 2019. Journée suisse des archives 2017 [en ligne]. 28 mars 2019. [Consulté le 7 avril 2019]. Disponible à l'adresse : https://commons.wikimedia.org/wiki/File:Bazak_Pr%C3%A4sentation_VSA_Archivtag_v2.pdf

MEDIAWIKI, 2018. Traduction de contenu [en ligne]. [Consulté le 7 avril 2019]. Disponible à l'adresse : https://www.mediawiki.org/wiki/Content_translation/fr

Modèle:Cite archive. *Wikipédia* [en ligne]. Dernière modification de la page le 11 avril 2018 à 18:03. [Consulté le 7 avril 2019]. Disponible à l'adresse : https://fr.wikipedia.org/w/index.php?title=Mod%C3%A8le:Cite_archive&oldid=147430674

Modèle:Infobox Archives. *Wikipédia* [en ligne]. Dernière modification de la page le 1 février 2018 à 22:33. [Consulté le 7 avril 2019]. Disponible à l'adresse : https://fr.wikipedia.org/w/index.php?title=Mod%C3%A8le:Infobox_Archives&oldid=145091266

SHONAGON, 2019. Wikimedia, Wikidata et Institutions patrimoniales (GLAM) [en ligne]. 28 mars 2019. [Consulté le 7 avril 2019]. Disponible à l'adresse : [https://commons.wikimedia.org/wiki/File:Wikimedia,_Wikidata_et_Institutions_patrimoniales_\(GLAM\).pdf](https://commons.wikimedia.org/wiki/File:Wikimedia,_Wikidata_et_Institutions_patrimoniales_(GLAM).pdf)

WIKIMEDIA, 2019. *BaGLAMa 2* [en ligne]. [Consulté le 7 avril 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/glamtools/baglama2/index.html>

WIKIMEDIA SWITZERLAND, 2019. GLAM stat tool (Cassandra) [en ligne]. [Consulté le 7 avril 2019]. Disponible à l'adresse : <http://stats.wikimedia.swiss/>

WIKIMEDIA SWITZERLAND, 2019. Wikidata:Events/GLAM & Wikimedia CH – Wikidata [en ligne]. [Consulté le 7 avril 2019]. Disponible à l'adresse : [https://www.wikidata.org/w/index.php?title=Wikidata:Events/GLAM %26 Wikimedia CH](https://www.wikidata.org/w/index.php?title=Wikidata:Events/GLAM_%26_Wikimedia_CH)

Annexe 2 : Wikidata:Events/Atelier 2019 à Lausanne

Cet atelier pratique s'est déroulé le 29 mars 2019, à l'espace Dickens de Lausanne. L'objectif de cette journée était d'établir une introduction théorique et pratique à la recherche et à la contribution sur Wikidata. La plateforme *Crotos*¹⁵ développée par l'intervenant de cet atelier, Benoît Deshayes (Wikimedia France), a servi de cas pratique au cours de cette journée.

Paramétrage d'un compte utilisateur

- Paramétrage des préférences d'un compte.
 - Il est possible de renseigner et paramétrer diverses options dans cette rubrique, comme des informations personnelles sur l'utilisateur du compte ou encore d'ajouter des Gadgets pour faciliter la recherche sur la plateforme.
- Activation d'un Gadgets dans les préférences de Wikidata.
 - Une liste des différents outils proposés sur la plateforme est disponible sur *Wikidata:Outils*.
 - L'outil *EasyQuery* permet de créer directement des requêtes SPARQL sur la page d'une entité sans utiliser l'interface *Query Service*.
- Ajouter un script en JavaScript pour l'automatisation de certaines tâches
 - Le code JavaScript *Common.js* permet de personnaliser certaines fonctionnalités d'un compte utilisateur.
 - Le code JavaScript *useful.js* permet une aide à la recherche de propriétés.

Rechercher une entité ou une propriété sur Wikidata

- Un répertoire exhaustif des propriétés est disponible sur la page *Wikidata:List of properties*. Ce répertoire est régulièrement mis à jour.
- La recherche simple s'effectue dans le champ recherche de l'interface de Wikidata. Une recherche avancée est également disponible.
- La recherche peut être effectuée via l'utilisation d'une API.
- Il est possible de configurer une librairie en python *Pywikibot* afin d'automatiser certaines tâches.

Présentation de l'outil de recherches en SPARQL

- Les requêtes en langage SPARQL peuvent être exécutées dans l'interface Query Service. Il existe un catalogue d'exemples¹⁶ permettant de construire facilement une requête à partir de contexte similaire.
- Il est possible de faire appel à la communauté pour éditer des requêtes complexes.
- Différents modes de visualisations sont disponibles pour afficher les résultats des requêtes sous forme de tableaux, galerie d'image, graphique de relations ou sous forme d'une cartographie.
- Le téléchargement des résultats des requêtes est disponible en plusieurs formats : csv, json, etc..
- Il est possible de convertir et de copier directement sur l'interface la requête dans divers langages de programmations (java, javascript, php, python, R, etc..). Une option utile pour faciliter la construction de requêtes, à partir d'une API distante.

¹⁵ DESHAYES, Benoît, [s.d]. Crotos [en ligne]. [Consulté le 8 avril 2019]. Disponible à l'adresse : <https://zone47.com/crotos/>

¹⁶ Wikidata:SPARQL query service/queries/examples. Wikidata [en ligne]. Dernière modification de la page le 23 mars 2019 à 13 h 21. [Consulté le 8 avril 2019]. Disponible à l'adresse : https://www.wikidata.org/wiki/Wikidata:Request_a_query

Ajouter des fichiers sur Commons

- Rechercher un modèle pour décrire un fichier dans Commons
 - Une liste exhaustive des différents modèles de description est disponible sur la page *Wikipedia:List of infoboxes*.
- Modification d'un modèle
 - Les modèles sont écrits en wikicode et peuvent être modifiés manuellement par l'utilisateur après l'import.
- Import d'un lot de fichiers
 - *Commons:Upload tools* propose une série d'outils pour le téléchargement d'un lot de fichiers vers Commons.
 - *Commons:GLAMwiki Toolset* propose une série d'outils à destination des institutions GLAM pour faciliter le téléchargement d'un grand volume de fichiers et pour la description.
- Copier des modèles dans Wikidata
 - L'outil *PLtools: Harvest Templates* permet de parcourir l'ensemble des pages rattaché à un modèle spécifique et de créer automatiquement les entités dans Wikidata.

Ajouter une entité ou une propriété sur Wikidata

- L'ajout manuel d'une nouvelle entité s'effectue dans le champ « Créer un nouvel élément » directement sur Wikidata.
- L'ajout des propriétés et de leurs descriptions peut être effectué manuellement sur la page de l'entité.
- L'outil *PLtools: Descriptioner* permet d'automatiser la description d'un lot d'entités, à l'aide d'une requête SPARQL.

Bibliographie

API:Main page. *MediaWiki* [en ligne]. Dernière modification de la page le 29 novembre 2018 à 22 h 35. [Consulté le 8 avril 2019]. Disponible à l'adresse : https://www.mediawiki.org/wiki/API:Main_page

API:Web APIs hub. *MediaWiki* [en ligne]. Dernière modification de la page le 7 août 2018 à 20 h 18. [Consulté le 8 avril 2019]. Disponible à l'adresse : https://www.mediawiki.org/wiki/API:Web_APIs_hub

Commons:GLAMwiki Toolset. *Wikimedia Commons* [en ligne]. Dernière modification de la page le 2 novembre 2018 à 19 h 52. [Consulté le 8 avril 2019]. Disponible à l'adresse : https://commons.wikimedia.org/wiki/Commons:GLAMwiki_Toolset

Commons:Upload tools. *Wikimedia Commons* [en ligne]. Dernière modification de la page le 11 mars 2019 à 20 h 05. Disponible à l'adresse : https://commons.wikimedia.org/wiki/Commons:Upload_tools

DESHAYES, Benoît, [s.d]. *Crotos* [en ligne]. [Consulté le 8 avril 2019]. Disponible à l'adresse : <https://zone47.com/crotos/>

Help:Description. *Wikidata* [en ligne]. Dernière modification de la page le 1 février 2019 à 5 h 30. [Consulté le 8 avril 2019]. Disponible à l'adresse : <https://www.wikidata.org/wiki/Help:Description>

Manual:Pywikibot. *MediaWiki* [en ligne]. Dernière modification de la page le 21 novembre 2018 à 18 h 42. [Consulté le 8 avril 2019]. Disponible à l'adresse : <https://www.mediawiki.org/wiki/Manual:Pywikibot>

MediaWiki:Common.js. *Wikipedia* [en ligne]. Dernière modification de la page le 19 mars 2019 à 15 h 20. [Consulté le 8 avril 2019]. Disponible à l'adresse : <https://en.wikipedia.org/w/index.php?title=MediaWiki:Common.js&oldid=888496901Page>

User:Magnus Manske/wikidata useful.js. *Wikidata* [en ligne]. Dernière modification de la page le 17 novembre 2017 à 9 h 47. [Consulté le 8 avril 2019]. Disponible à l'adresse : [https://www.wikidata.org/wiki/User:Magnus Manske/wikidata useful.js](https://www.wikidata.org/wiki/User:Magnus_Manske/wikidata_useful.js)

Wikidata:List of properties. *Wikidata* [en ligne]. Dernière modification de la page le 3 février 2019 à 16 h 56. [Consulté le 8 avril 2019]. Disponible à l'adresse : https://www.wikidata.org/wiki/Wikidata:List_of_properties

Wikidata:Outils. *Wikidata* [en ligne]. Dernière modification de la page le 27 décembre 2018 à 18 h 27. [Consulté le 8 avril 2019]. Disponible à l'adresse : <https://www.wikidata.org/wiki/Wikidata:Tools/fr>

Wikidata:Request a query. *Wikidata* [en ligne]. Dernière modification de la page le 8 avril 2019 à 16 h 28. [Consulté le 8 avril 2019]. Disponible à l'adresse : https://www.wikidata.org/wiki/Wikidata:Request_a_query

Wikidata:SPARQL query service/queries/examples. *Wikidata* [en ligne]. Dernière modification de la page le 23 mars 2019 à 13 h 21. [Consulté le 8 avril 2019]. Disponible à l'adresse : https://www.wikidata.org/wiki/Wikidata:Request_a_query

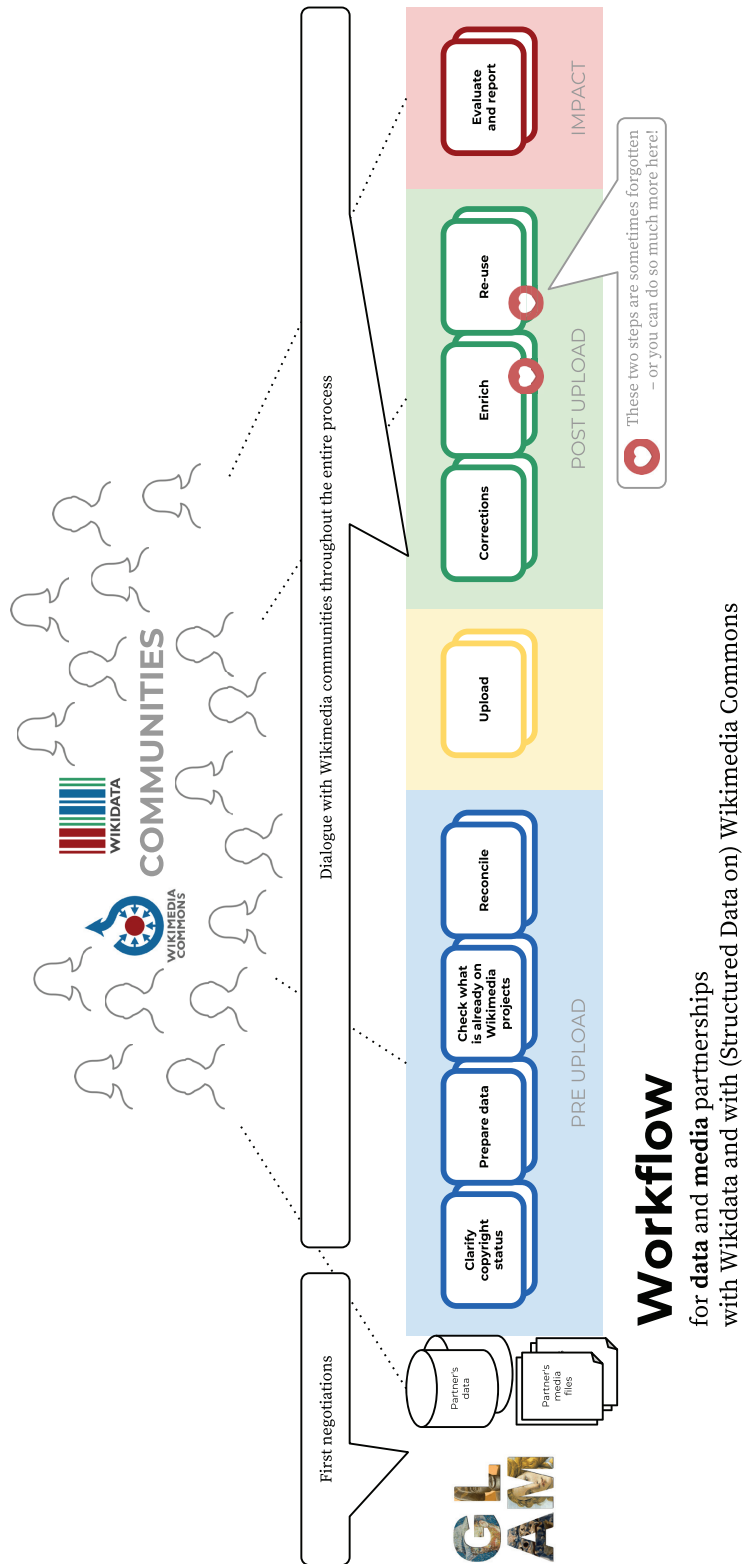
Wikipedia:List of infoboxes. *Wikipedia* [en ligne]. Dernière modification de la page le 5 mars 2019 à 23 h 43. [Consulté le 8 avril 2019]. Disponible à l'adresse : https://en.wikipedia.org/w/index.php?title=Wikipedia:List_of_infoboxes&oldid=886386640

WIKIMEDIA TOOLFORGE, [s.d]. *glam2commons* [en ligne]. [Consulté le 8 avril 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/glam2commons/>

Annexe 3 : Plan de recherches

Axe	Sous-Axe	Enjeux	Type de source
Parties prenantes	Acteurs publics et associatifs de l'Open data Suisse	Étude des documents stratégiques	Moteurs et metamoteur de recherche. Site institutionnel
	Acteurs publics, privés et associatifs de l'audiovisuel	Inventaire et présentation succincte des normes, recommandations et standards	Moteurs et metamoteur de recherche. Site institutionnel
	GLAM (Galleries, Bibliothèques, Archives et Musées)	Inventaire et présentation succincte des standards de description spécifique à l'audiovisuel	Moteurs et metamoteur de recherche. Revue spécialisée
Standard de préservation numérique dans l'audiovisuel	Formats, conteneurs, codec et fichier pérennes	Inventaire et présentation succincte des modèles de données et métadonnées	Moteurs et metamoteur de recherche. Site institutionnel
Linked Open Data et audiovisuel	Associations professionnelles et communauté	Inventaire des vocabulaires et ontologies dans le domaine de l'audiovisuel	Moteurs et metamoteur de recherche. Revue spécialisée
	Standards de qualité	Normes et standards pour l'utilisation et la réutilisation des Open data	Moteurs et metamoteur de recherche. Revue spécialisée
	Exemples de projets d'enrichissement des notices par crowdsourcing	Retour d'expériences	Moteurs et metamoteur de recherche. Revue spécialisée

Annexe 4 : Workflow GLAM pour un versement



(GLAM/Resources/Data and media partnerships workflow 2019)

Annexe 5 : La recherche d'information sur Wikidata

La recherche d'information sur la plateforme Wikidata est effectuée par défaut, à l'aide de l'interface du Wikibase Client par un utilisateur non authentifié, comme décrit dans le cas d'utilisation *Recherche non authentifiée* de la figure 1. L'ensemble des cas d'utilisations décrit sur la figure 1 permettent d'effectuer des recherches, à distance à partir d'adresses url (endpoints) mises à disposition pour l'interrogation d'une application programming interface (API).

Le cas d'utilisation *Recherche authentifiée* est une extension du cas précédent. En effet, l'utilisateur doit d'abord être identifié sur la plateforme pour pouvoir utiliser cette deuxième possibilité de recherche. Cependant, ces deux cas d'utilisations incluent par défaut, le cas d'utilisation *Interface* qui permet d'effectuer des recherches sur le Wikibase Repository, via l'interface du Wikibase Client. Le Wikibase Client propose deux interfaces de recherche différentes. Le premier est un champ de recherche directement accessible, depuis la page d'accueil de Wikidata et le second est accessible, à partir du Wikidata Query Service (Wikidata, sans date a).

Dans le cas d'utilisation *Interface*, les modalités de recherche disponibles dans le champ de recherche principale sont la recherche simple, avancée et à facette. La recherche simple permet d'effectuer directement une recherche en mode plein texte. La recherche avancée augmente la précision de la recherche plein texte, grâce à l'utilisation d'opérateurs booléens. Enfin, la recherche à facette est une extension de la recherche avancée, qui offre la possibilité de sélectionner une ou plusieurs entités nommées. Ces entités nommées ont été définies à partir du modèle de données de Wikibase.

Les deux cas d'utilisation *Distante sans robot* et *Distante avec un robot* sont des extensions des cas d'utilisations *Recherche authentifiée* et *Recherche non authentifiée*. Le premier cas d'utilisation est disponible via la configuration et le paramétrage d'un script permettant l'exécution manuelle d'une API. Le deuxième cas d'utilisation prévoit également une utilisation distante de l'API et l'exécution manuelle ou automatique d'un script, à la suite de la configuration d'un robot.

Le cas d'utilisation *Alignement des données* est à la fois une extension du cas *Distante sans robot* et une inclusion du cas *Interface*. Dans le cas d'utilisation *Recherche non authentifiée*, le cas *Alignement des données* propose à l'utilisateur une recherche sur l'API du Wikidata reconciliation for OpenRefine (Wikimedia toolforge sans date b) en utilisant soit les cas *Interface* ou *Distante sans robot*. Dans le premier cas, un import de données externe peut être effectué directement dans le navigateur, à partir par exemple de l'interface Mix'n'match (Wikimedia toolforge sans date a). Un jeu de données sous forme tabulaire peut être copié pour lancer la recherche. Dans le second cas, l'API du Wikidata reconciliation for OpenRefine peut être interrogée à l'aide d'un script.

Le cas d'utilisation *Interface locale* comprend l'interrogation de l'API du Wikidata reconciliation for OpenRefine, à l'aide d'un logiciel téléchargé préalablement en local.

Le Wikidata reconciliation for OpenRefine est une API, qui offre pour sa part, la possibilité de réaliser un alignement automatique des données de type textuelle ou numérique. Cet alignement peut être effectué en fonction de paramètres préalablement définis comme par

exemple une propriété spécifique à Wikidata. L'API permet de faire des propositions sur les données disponibles sur la plateforme et sémantiquement proche avec les données locales et/ou présentes sur d'autres plateformes distantes. Cette API peut être utilisée indépendamment d'un logiciel pour répondre à un besoin spécifique, ou alors directement intégrée à l'interface d'un logiciel, comme par exemple c'est le cas dans les logiciels OpenRefine (Google sans date) ou OntoRefine, une extension de la base de graphe GraphDB.

L'interface Wikidata Query Service offre la possibilité d'une recherche dans le Wikibase Repository et sur d'autres bases de données externes à la plateforme, grâce au Linked Open Data (LOD). Les recherches sont effectuées en langage SPARQL. Il est également possible de simplifier les adresses url des requêtes SPARQL, à l'aide de préfix. Ces préfix concernent principalement l'ontologie Wikibase et d'autres vocabulaires d'ontologie externe.

Les préfix (Wikibase/Indexing/RDF Dump Format 2019) disponibles sur l'interface du Query Service sont répartis en quatre catégories sur l'interface : *Wikidata*, *W3C*, *Social/Other* et *Blazegraph*. La première catégorie rassemble un total de 18 préfix comprenant le vocabulaire d'ontologie Wikibase et permet l'utilisation de la valeur d'une propriété ou d'une entité (*item*).

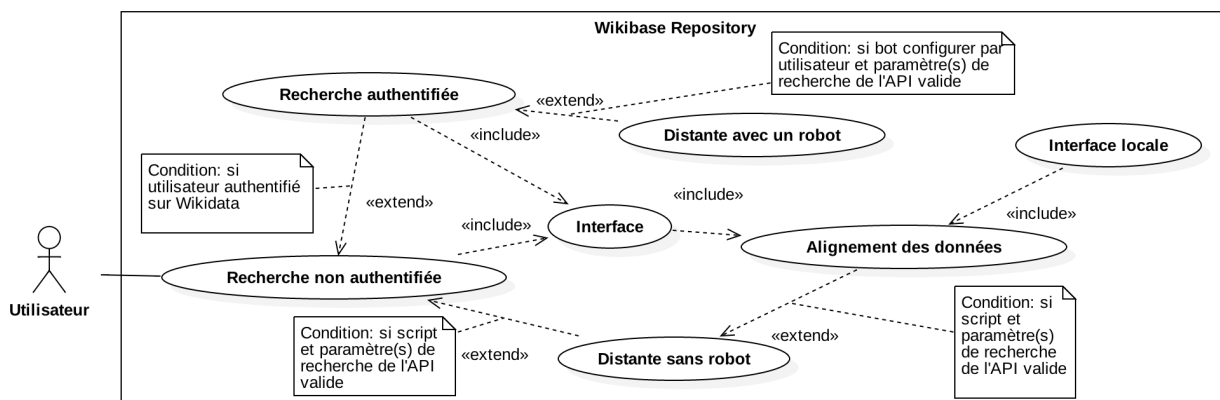
La catégorie *W3C* comprend les vocabulaires d'ontologie maintenu par le World Wide Web Consortium (W3C) et se compose de 6 préfix. Ces préfix contiennent notamment les vocabulaires d'ontologie Resource Description Framework Schema (RDFS), Simple Knowledge Organization System (SKOS), Web Ontology Language (OWL), XML Schema Definition (XSD) ou encore PROV Ontology (PROV-O). La troisième catégorie concerne des vocabulaires d'ontologie multi-domaines comme ceux proposés par Schema.org. La dernière catégorie contient des préfix optimisés pour une utilisation, à partir de la base de données de graph Blazegraph. Cette base de données est le moteur de base de données de graphe de l'interface du Wikidata Query Service.

Les résultats obtenus, à la suite de l'exécution d'une requête dans le Wikidata Query Service peuvent être directement affichés dans le navigateur et prendre diverses formes (liste, table, graphique, cartographie). Il est également possible d'exporter et de sauvegarder les résultats obtenus dans divers formats de fichiers interopérables.

Les diverses API proposées par la plateforme Wikidata permettent d'accomplir les quatre opérations de base sur les données, c'est-à-dire la création, la lecture, la mise à jour et la suppression des données (CRUD). Ces API proposent également, l'exécution à distance de procédures d'extraction, de transformation et de chargement des données (ETL). L'API de Wikidata (Wikibase/API 2019) a été adaptée de l'API MediaWiki Action (API:Main page 2018) employée pour l'interrogation distante d'autres projets collaboratifs de la fondation Wikimedia, comme Wikipédia ou encore Wikimedia Commons. L'accès à cette API est rendu possible, grâce à une adresse url permanente (Wikimedia sans date c) (endpoint). Cet url permet d'interagir avec la plateforme grâce à un script pouvant être écrit dans une grande variété de langage de programmation différents (Python, PHP, Java, JavaScript, Perl, etc..). Le Query Service est également interrogeable à distance via l'utilisation d'une API (Wikidata sans date b).

Le Pywikibot (Manuel:Pywikibot 2019) est une librairie de script écrite en langage de programmation python. Cette librairie propose l'exécution d'un robot de navigation, et un accès distant (API) aux différents services des plateformes de la fondation Wikimedia.

Figure 1 : Cas d'utilisations de recherches disponibles sur Wikidata



Bibliographie

API:Main page. MediaWiki [en ligne]. Dernière modification de la page le 29 novembre 2018 à 22 h 35. [Consulté le 8 avril 2019]. Disponible à l'adresse : https://www.mediawiki.org/wiki/API:Main_page

GOOGLE, [sans date]. OpenRefine [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <http://openrefine.org/>

WIKIDATA, [sans date a]. Wikidata Query Service [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://query.wikidata.org/>

WIKIDATA, [sans date b]. Wikidata Query Service [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://query.wikidata.org/sparql>

WIKIMEDIA TOOLFORGE, [sans date a]. Mix'n'match [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/mix-n-match/#/>

WIKIMEDIA TOOLFORGE, [sans date b]. OpenRefine-Wikidata interface [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/openrefine-wikidata>

Wikibase/Indexing/RDF Dump Format. MediaWiki [en ligne]. Dernière modification de la page le 2 mai 2019 à 21:49. [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format

Wikibase/API. MediaWiki [en ligne]. Dernière modification de la page le 27 mai 2019 à 11:43. [en ligne]. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.mediawiki.org/wiki/Wikibase/API>

Manuel:Pywikibot. MediaWiki [en ligne]. Dernière modification de la page le 13 juin 2019 à 11:21. [Consulté le 8 juillet 2019]. Disponible à l'adresse : <https://www.mediawiki.org/wiki/Manuel:Pywikibot/fr>

Annexe 6 : Table de conversion de la largeur des coffrets

Support	Largeur [m]
Quadruplex; 2"	0,10160
1-inch Type B; 1"	0,03175
1-inch Type C; 1"	0,03175
1-inch HD video; 1"	0,03175
Bande magnétique; 1/2"	0,03175
Bande magnétique; 1/4"	0,01270
U-matic; 3/4"	0,03050
Digital D-2; 3/4"	0,03378
Betacam; 1/2"	0,03175
Betacam SP; 1/2"	0,03175
Digital Betacam; 1/2"	0,02540
Betamax; 1/2"	0,02540
HDCAM; 1/2"	0,02540
Digital D-5; 1/2"	0,02540
LTO-4; 1/2"	0,02100
LTO-6; 1/2"	0,02100
DVCAM; 1/4"	0,01270
MiniDV; 1/4"	0,00635
Digital Audio Tape (DAT); 3/20"	0,01050
Professional Disc (PFD)	0,01050

(Wolframalpha 2019)

Annexe 7 : Matrice d'adjacence des entités RiC

	RiC-E3 Record Set	RiC-E1 Record	RiC-E4 Agent (of type person)	RiC-E4 Agent (of type group)	RiC-E5 Occupation	RiC-E9 Activity	RiC-E10 Mandate	RiC-E12 Date	RiC-E13 place	has holding location (RiC-R141)	has coverage date (RiC-R141)	has holding location (RiC-R192)	has associated location (RiC-R65)	is subject of (RiC-R63)	is member of (RiC-R570)	is associated with (RiC-R266)	performs (RiC-R309)	is associated with (RiC-R135)	had creation date (RiC-R62)	
Archives	x																			
Audio file		x																		
Audio tape		x																		
Concert									x											
Concert Event		x																		
Concert files		x																		
Contract																				
Date Archives																				
Date Concert																				
Date Contract																				
Date Festival																				
Date Files																				
Date LTO tape																				
Date Photo file																				
Festival		x																		
Function																				
Group																				
LTO tape																				
Person																				
Photo file																				
Place archives																				
Place Concert																				
Place Festival																				
Place LTO tape																				
Playlist																				
Recording																				
Role																				
Song																				
Video file																				
Video tape																				

Annexe 8 : Requête CONSTRUCT en SPARQL

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ric: <http://purl.org/ica/ric#> # Introduit l'adresse IRI de
l'ontologie RiC
PREFIX mjf: <http://mjf-database.epfl.ch/mjf/>
PREFIX spif: <http://spinrdf.org/spif#>
```

```
CONSTRUCT {
  ?myRowId a mjf:Row ; #Nœud principale du graphe
  ric:RiC-E1-Record ?audio_file ; # Introduit une relation
  ric:RiC-E3-RecordSet ?concert ;
  ric:RiC-E3-RecordSet ?concert_files ;
  ric:RiC-E12-Date ?date_concert ;
  ric:RiC-E12-Date ?date_files ;
  ric:RiC-E12-Date ?date_lto_tape ;
  ric:RiC-E9-Activity ?function ;
  ric:RiC-E4-Agent-Group ?group ;
  ric:RiC-E1-Record ?lto_tape ;
  ric:RiC-E4-Agent-Person ?person ;
  ric:RiC-E1-Record ?photo_file ;
  ric:RiC-E13-Place ?place_concert ;
  ric:RiC-E13-Place ?place_lto_tape ;
  ric:RiC-E5-Occupation ?role ;
  ric:RiC-E3-RecordSet ?recording ;
  ric:RiC-E1-Record ?video_file ;
  ric:RiC-E1-Record ?video_tape ;
  ric:RiC-R-isAssociatedWith ?ric_is_associated_with ;
  ric:RiC-R-hasSubject ?ric_has_subject ;
  ric:RiC-R-hasMember ?ric_has_member_ric ;
  ric:RiC-R-hasHoldingLocation ?ric_had_holding_location ;
  ric:RiC-R-performed ?ric_performs ;
  ric:RiC-R-hadCreationDate ?ric_had_creation_date ;
  rdf:type ?wd .
}
WHERE {
  SERVICE <ontorefine:1691076834100>{ # Indique l'adresse IRI du
répertoire dans la base de données de graphe
  ?row a mjf:Row ; # Sélectionne les colonnes de la matrice
d'adjacence
  mjf:audio_file ?audio_file ;
  mjf:concert ?concert ;
  mjf:concert_files ?concert_files ;
  mjf:date_concert ?date_concert ;
  mjf:date_files ?date_files ;
  mjf:date_lto_tape ?date_lto_tape ;
  mjf:function ?function ;
  mjf:group ?group ;
  mjf:lto_tape ?lto_tape ;
  mjf:person ?person ;
  mjf:photo_file ?photo_file ;
  mjf:place_concert ?place_concert ;
  mjf:place_lto_tape ?place_lto_tape ;
  mjf:role ?role ;
  mjf:recording ?recording ;
  mjf:video_file ?video_file ;
  mjf:video_tape ?video_tape ;
  mjf:ric_is_associated_with ?ric_is_associated_with ;
```

```

        mjf:ric_has_subject ?ric_has_subject ;
        mjf:ric_has_member_ric ?ric_has_member_ric ;
        mjf:ric_had_holding_location
?ric_had_holding_location ;
        mjf:ric_performs ?ric_performs ;
        mjf:ric_had_creation_date ?ric_had_creation_date ;
        mjf:wd ?wd .

        BIND(IRI(spif:buildString("http://mjf-
database.epfl.ch/mjf/{?1}", ENCODE_FOR_URI(?id))) AS ?myRowId) #
Initialise l'adresse IRI du nœud principale du graphe
    }
}

```

Annexe 9 : Sélection manuelle des propriétés Wikidata

wikidata_property	wikibase_data_type	property_id	property_uri
image	Commons media file	P18	http://www.wikidata.org/prop/P18
image of interior	Commons media file	P5775	http://www.wikidata.org/prop/P5775
media legend	Commons media file	P2096	http://www.wikidata.org/prop/P2096
nighttime view	Commons media file	P3451	http://www.wikidata.org/prop/P3451
scanned file on Wikimedia Commons	Commons media file	P996	http://www.wikidata.org/prop/P996
All About Jazz musician ID	external-id	P5121	http://www.wikidata.org/prop/P5121
AllMusic artist ID	external-id	P1728	http://www.wikidata.org/prop/P1728
AllMusic song ID	external-id	P1730	http://www.wikidata.org/prop/P1730
Amazon Music artist ID	external-id	P6276	http://www.wikidata.org/prop/P6276
Amazon Standard Identification Number	external-id	P5749	http://www.wikidata.org/prop/P5749
B.R.A.H.M.S. artist ID	external-id	P5226	http://www.wikidata.org/prop/P5226
Carnegie Hall agent ID	external-id	P4104	http://www.wikidata.org/prop/P4104
CDMC musician ID	external-id	P5272	http://www.wikidata.org/prop/P5272
Discogs artist ID	external-id	P1953	http://www.wikidata.org/prop/P1953
Discogs composition ID	external-id	P6080	http://www.wikidata.org/prop/P6080
Discogs label ID	external-id	P1955	http://www.wikidata.org/prop/P1955
Discogs track ID	external-id	P6079	http://www.wikidata.org/prop/P6079
Europeana ID	external-id	P727	http://www.wikidata.org/prop/P727
Genius album ID	external-id	P6217	http://www.wikidata.org/prop/P6217
Genius artist numeric ID	external-id	P6351	http://www.wikidata.org/prop/P6351
Genius song ID	external-id	P6218	http://www.wikidata.org/prop/P6218
Google Play Music album ID	external-id	P4199	http://www.wikidata.org/prop/P4199
Google Play Music artist ID	external-id	P4198	http://www.wikidata.org/prop/P4198
International Standard Musical Work Code	external-id	P1827	http://www.wikidata.org/prop/P1827
ISO 639-1 code	external-id	P218	http://www.wikidata.org/prop/P218
International Standard Recording Code	external-id	P1243	http://www.wikidata.org/prop/P1243
iTunes album ID	external-id	P2281	http://www.wikidata.org/prop/P2281
iTunes artist ID	external-id	P2850	http://www.wikidata.org/prop/P2850
iTunes music video ID	external-id	P5655	http://www.wikidata.org/prop/P5655
Last.fm ID	external-id	P3192	http://www.wikidata.org/prop/P3192
Library of Congress authority ID	external-id	P244	http://www.wikidata.org/prop/P244
Library of Congress JukeBox ID	external-id	P2089	http://www.wikidata.org/prop/P2089
LocFDD ID	external-id	P3266	http://www.wikidata.org/prop/P3266
MusicBrainz area ID	external-id	P982	http://www.wikidata.org/prop/P982
MusicBrainz artist ID	external-id	P434	http://www.wikidata.org/prop/P434
MusicBrainz event ID	external-id	P6423	http://www.wikidata.org/prop/P6423
MusicBrainz instrument ID	external-id	P1330	http://www.wikidata.org/prop/P1330
MusicBrainz label ID	external-id	P966	http://www.wikidata.org/prop/P966
MusicBrainz place ID	external-id	P1004	http://www.wikidata.org/prop/P1004
MusicBrainz recording ID	external-id	P4404	http://www.wikidata.org/prop/P4404
MusicBrainz release group ID	external-id	P436	http://www.wikidata.org/prop/P436
MusicBrainz release ID	external-id	P5813	http://www.wikidata.org/prop/P5813
MusicBrainz series ID	external-id	P1407	http://www.wikidata.org/prop/P1407
MusicBrainz work ID	external-id	P435	http://www.wikidata.org/prop/P435
PRONOM file format identifier	external-id	P2748	http://www.wikidata.org/prop/P2748
RollDaBeats artist ID	external-id	P5240	http://www.wikidata.org/prop/P5240
RYM artist ID	external-id	P5404	http://www.wikidata.org/prop/P5404
Shazam artist ID	external-id	P5917	http://www.wikidata.org/prop/P5917
Shazam track ID	external-id	P5915	http://www.wikidata.org/prop/P5915
Songfacts song ID	external-id	P5241	http://www.wikidata.org/prop/P5241
Spotify album ID	external-id	P2205	http://www.wikidata.org/prop/P2205
Spotify track ID	external-id	P2207	http://www.wikidata.org/prop/P2207
Swiss Federal Archives ID	external-id	P3889	http://www.wikidata.org/prop/P3889
Swiss municipality code	external-id	P771	http://www.wikidata.org/prop/P771
Tidal artist ID	external-id	P4576	http://www.wikidata.org/prop/P4576
VIAF ID	external-id	P214	http://www.wikidata.org/prop/P214
WhoSampled artist ID	external-id	P6517	http://www.wikidata.org/prop/P6517
coordinate location	Geographic coordinates	P625	http://www.wikidata.org/prop/P625

wikidata_property	wikibase_data_type	property_id	property_uri
aspect ratio	item	P2061	http://www.wikidata.org/prop/P2061
attributed to	item	P1773	http://www.wikidata.org/prop/P1773
author	item	P50	http://www.wikidata.org/prop/P50
composer	item	P86	http://www.wikidata.org/prop/P86
continent	item	P30	http://www.wikidata.org/prop/P30
country	item	P17	http://www.wikidata.org/prop/P17
depicts	item	P180	http://www.wikidata.org/prop/P180
discography	Item	P358	http://www.wikidata.org/prop/P358
family name	item	P734	http://www.wikidata.org/prop/P734
genre	item	P136	http://www.wikidata.org/prop/P136
given name	item	P735	http://www.wikidata.org/prop/P735
instrument	item	P1303	http://www.wikidata.org/prop/P1303
located in the administrative territorial entity	item	P131	http://www.wikidata.org/prop/P131
manufacturer	item	P176	http://www.wikidata.org/prop/P176
mapping relation type	item	P4390	http://www.wikidata.org/prop/P4390
producer	item	P162	http://www.wikidata.org/prop/P162
published in	item	P1433	http://www.wikidata.org/prop/P1433
record label	item	P264	http://www.wikidata.org/prop/P264
nickname	Monolingual text	P1449	http://www.wikidata.org/prop/P1449
official name	Monolingual text	P1448	http://www.wikidata.org/prop/P1448
quote	Monolingual text	P1683	http://www.wikidata.org/prop/P1683
title	Monolingual text	P1476	http://www.wikidata.org/prop/P1476
data size	Quantity	P3575	http://www.wikidata.org/prop/P3575
duration	Quantity	P2047	http://www.wikidata.org/prop/P2047
frequency	quantity	P2144	http://www.wikidata.org/prop/P2144
height	Quantity	P2048	http://www.wikidata.org/prop/P2048
width	Quantity	P2049	http://www.wikidata.org/prop/P2049
file extension	String	P1195	http://www.wikidata.org/prop/P1195
inventory number	string	P217	http://www.wikidata.org/prop/P217
media type	String	P1163	http://www.wikidata.org/prop/P1163
pseudonym	String	P742	http://www.wikidata.org/prop/P742
dissolved, abolished or demolished	time	P576	http://www.wikidata.org/prop/P576
end time	time	P582	http://www.wikidata.org/prop/P582
inception	time	P571	http://www.wikidata.org/prop/P571
production date	time	P2754	http://www.wikidata.org/prop/P2754
publication date	time	P577	http://www.wikidata.org/prop/P577
start time	time	P580	http://www.wikidata.org/prop/P580
official website	url	P856	http://www.wikidata.org/prop/P856
date of official opening	time	P1619	http://www.wikidata.org/prop/P1619
recorded at	item	P483	http://www.wikidata.org/prop/P483
numeric value	quantity	P1181	http://www.wikidata.org/prop/P1181

(Wikidata:Database reports/List of properties/all 2019)

Annexe 10 : Requêtes SPARQL avec identifiants de type external-id

```
SELECT (COUNT (?o) AS ?objet)
WHERE{
    {?All_About_Jazz_musician_id wdt:P5121 ?o .}
    UNION {?AllMusic_artist_id wdt:P1728 ?o .}
    UNION {?AllMusic_song_id wdt:P1730 ?o .}
    UNION {?Amazon_Music_artist_id wdt:P6276 ?o .}
    UNION {?Discogs_artist_id wdt:P1953 ?o .}
    UNION {?Discogs_composition_id wdt:P6080 ?o .}
    UNION {?Discogs_label_id wdt:P1955 ?o .}
    UNION {?Discogs_track_id wdt:P6079 ?o .}
    UNION {?Genius_album_id wdt:P6217 ?o .}
    UNION {?Genius_artist_numeric_id wdt:P6351 ?o .}
    UNION {?Genius_song_id wdt:P6218 ?o .}
    UNION {?Google_Play_Music_album_id wdt:P4199 ?o .}
    UNION {?Google_Play_Music_artist_id wdt:P4198 ?o .}
    UNION {?iTunes_album_id wdt:P2281 ?o .}
    UNION {?iTunes_artist_id wdt:P2850 ?o .}
    UNION {?iTunes_music_video_id wdt:P5655 ?o .}
    UNION {?Last_fm_id wdt:P3192 ?o .}
    UNION {?Library_of_Congress_JukeBox_id wdt:P2089 ?o .}
    UNION {?MusicBrainz_artist_id wdt:P434 ?o .}
    UNION {?MusicBrainz_event_id wdt:P6423 ?o .}
    UNION {?MusicBrainz_instrument_id wdt:P1330 ?o .}
    UNION {?MusicBrainz_label_id wdt:P966 ?o .}
    UNION {?MusicBrainz_recording_id wdt:P4404 ?o .}
    UNION {?MusicBrainz_work_id wdt:P435 ?o .}
    UNION {?RollDaBeats_artist_id wdt:P5240 ?o .}
    UNION {?RYM_artist_id wdt:P5404 ?o .}
    UNION {?Shazam_artist_id wdt:P5917 ?o .}
    UNION {?Shazam_track_id wdt:P5915 ?o .}
    UNION {?Songfacts_song_id wdt:P5241 ?o .}
    UNION {?Spotify_album_id wdt:P2205 ?o .}
    UNION {?Spotify_track_id wdt:P2207 ?o .}
    UNION {?Tidal_artist_id wdt:P4576 ?o .}
    UNION {?WhoSampled_artist_id wdt:P6517 ?o .}
    UNION {?International_Standard_Musical_Work_Code wdt:P1827 ?o .}
    UNION {?International_Standard_Recording_Code wdt:P1243 ?o .}
    FILTER(?o)
}
```

```
SELECT (COUNT (?o) AS ?objet)
WHERE{
    {?MusicBrainz_area_ID wdt:P982 ?o .}
    UNION {?MusicBrainz_artist_ID wdt:P434 ?o .}
    UNION {?MusicBrainz_event_ID wdt:P6423 ?o .}
    UNION {?MusicBrainz_instrument_ID wdt:P1330 ?o .}
    UNION {?MusicBrainz_label_ID wdt:P966 ?o .}
    UNION {?MusicBrainz_place_ID wdt:P1004 ?o .}
    UNION {?MusicBrainz_recording_ID wdt:P4404 ?o .}
    UNION {?MusicBrainz_release_group_ID wdt:P436 ?o .}
    UNION {?MusicBrainz_release_ID wdt:P5813 ?o .}
    UNION {?MusicBrainz_series_ID wdt:P1407 ?o .}
    UNION {?MusicBrainz_work_ID wdt:P435 ?o .}
    FILTER(?o)
}
```

```

SELECT (COUNT (?o) AS ?objet)
WHERE{
    {?Discogs_label_id wdt:P1955 ?o .}
    UNION {?MusicBrainz_label_id wdt:P966 ?o .}
    FILTER(?o)
}

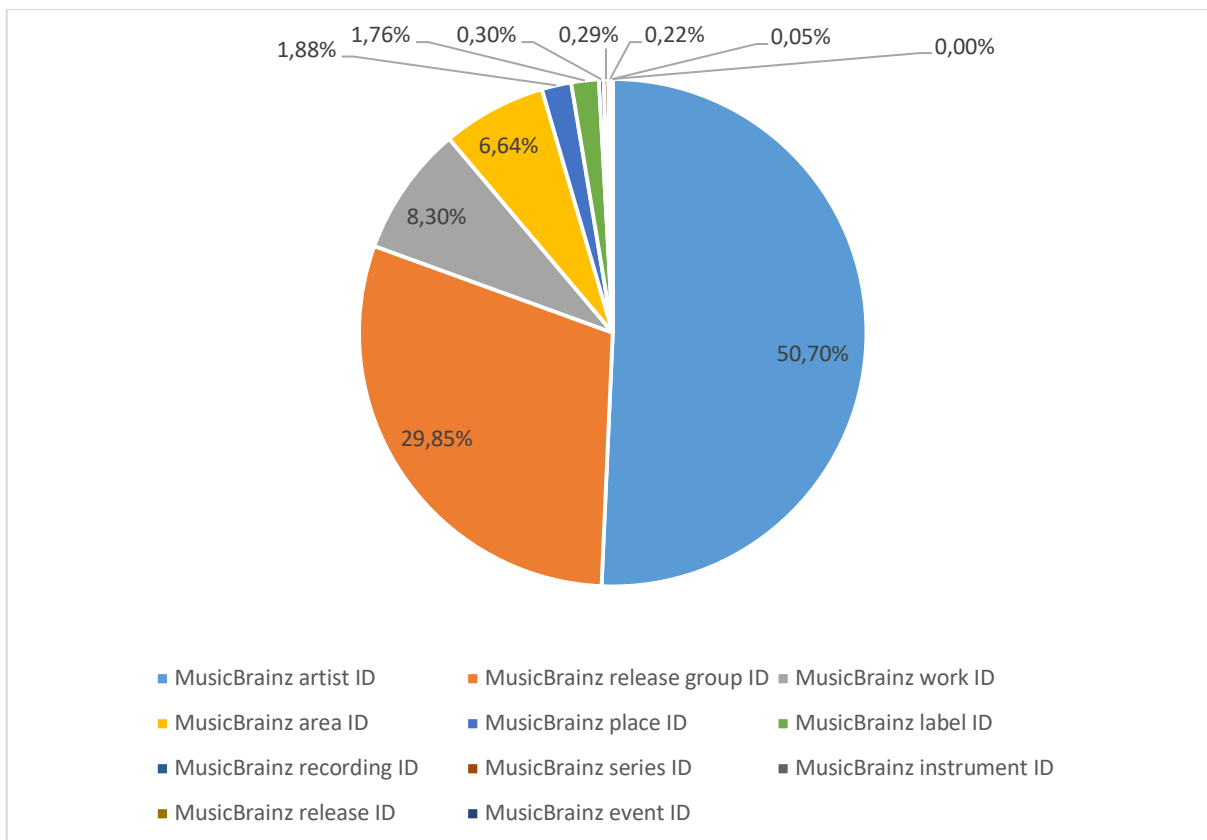
SELECT (COUNT (?o) AS ?objet)
WHERE{
    {?Genius_album_id wdt:P6217 ?o .}
    UNION {?MusicBrainz_recording_id wdt:P4404 ?o .}
    UNION {?Spotify_album_id wdt:P2205 ?o .}
    UNION {?Google_Play_Music_album_id wdt:P4199 ?o .}
    UNION {?iTunes_album_id wdt:P2281 ?o .}
    FILTER(?o)
}

SELECT (COUNT (?o) AS ?objet)
WHERE{
    {?AllMusic_song_id wdt:P1730 ?o .}
    UNION {?Discogs_composition_id wdt:P6080 ?o .}
    UNION {?Discogs_track_id wdt:P6079 ?o .}
    UNION {?Genius_song_id wdt:P6218 ?o .}
    UNION {?Library_of_Congress_JukeBox_id wdt:P2089 ?o .}
    UNION {?MusicBrainz_recording_id wdt:P4404 ?o .}
    UNION {?MusicBrainz_work_id wdt:P435 ?o .}
    UNION {?Shazam_track_id wdt:P5915 ?o .}
    UNION {?Songfacts_song_id wdt:P5241 ?o .}
    UNION {?Spotify_track_id wdt:P2207 ?o .}
    UNION {?International_Standard_Musical_Work_Code wdt:P1827 ?o .}
    UNION {?International_Standard_Recording_Code wdt:P1243 ?o .}
    FILTER(?o)
}

SELECT (COUNT (?o) AS ?objet)
WHERE{
    {?All_About_Jazz_musician_id wdt:P5121 ?o .}
    UNION {?AllMusic_artist_id wdt:P1728 ?o .}
    UNION {?Amazon_Music_artist_id wdt:P6276 ?o .}
    UNION {?Discogs_artist_id wdt:P1953 ?o .}
    UNION {?Genius_artist_numeric_id wdt:P6351 ?o .}
    UNION {?iTunes_artist_id wdt:P2850 ?o .}
    UNION {?MusicBrainz_artist_id wdt:P434 ?o .}
    UNION {?RollDaBeats_artist_id wdt:P5240 ?o .}
    UNION {?RYM_artist_id wdt:P5404 ?o .}
    UNION {?Shazam_artist_id wdt:P5917 ?o .}
    UNION {?Tidal_artist_id wdt:P4576 ?o .}
    UNION {?WhoSampled_artist_id wdt:P6517 ?o .}
    UNION {?Last_fm_id wdt:P3192 ?o .}
    UNION {?Library_of_Congress_JukeBox_id wdt:P2089 ?o .}
    FILTER(?o)
}

```

Annexe 11 : Distribution des propriétés MusicBrainz



(Wikidata sans date a)

Annexe 12 : Modèle de versement opendata.swiss

<!-- Modèle de versement pour la plateforme opendata.swiss adapté au contexte du Montreux Jazz Digital Project (MJDP) et construit, à partir de l'exemple de swisstopo disponible à l'adresse : https://handbook.opendata.swiss/samples/ogdch_dcatap_import.rdf -->

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcat="http://www.w3.org/ns/dcat#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:vcard="http://www.w3.org/2006/vcard/ns#"
  xmlns:odrs="http://schema.theodi.org/odrs#"
  xmlns:schema="http://schema.org/">
  <!-- Initialise la classe Catalog du modèle Data Catalog Vocabulary (DCAT) -->
  <dcat:Catalog>
    <!-- Initialise la propriété dataset -->
    <dcat:dataset>
      <dcat:Dataset
rdf:about="https://metamedia.epfl.ch/montreux_jazz_digital_project-about">
<!-- Initialise la classe Dataset -->
      <dct:title xml:lang="fr">Montreux Jazz Digital Project</dct:title>
<!-- Titre principal du jeu de données -->
      <!-- Description du jeu de données -->
      <dct:description xml:lang="fr">Ce document comprend un jeu de données extrait des enregistrements de la base de données du Montreux Jazz Digital Project. Ces enregistrements concernent principalement des données factuelles liées aux concerts du Montreux Jazz Festival de 1966 à 2018. Les informations présente sur ce jeu de données concerne des thématiques liées aux concerts (nom, date, durée), musiciens (nom) et œuvres musicales (titre, ordre, durée) du fonds.</dct:description>
      <dct:issued
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2019-09-04T01:00:00Z</dct:issued> <!-- Date de création du jeu de données -->
      <!-- Informations sur le producteur principal du jeu de données -->
      <dct:publisher>
        <rdf:Description rdf:about="https://metamedia.epfl.ch/">
          <rdfs:label>Montreux Jazz Digital Project</rdfs:label>
        </rdf:Description>
      </dct:publisher>
      <!-- Informations sur l'entité dont dépend le producteur -->
      <dct:publisher>
        <rdf:Description rdf:about="https://www.epfl.ch/">
          <rdfs:label>Ecole Polytechnique Fédérale de Lausanne (EPFL)</rdfs:label>
        </rdf:Description>
      </dct:publisher>
      <!-- Moyens de contact du producteur -->
      <dcat:contactPoint>
        <vcard:Organization>
          <vcard:fn>Metamedia Center (MMC)</vcard:fn>
          <vcard:hasEmail rdf:resource="mailto:metamedia@epfl.ch"/>
        </vcard:Organization>
      </dcat:contactPoint>
      <!-- Thématiques opendata.swiss traités par le jeu de données -->
      <dcat:theme rdf:resource="http://opendata.swiss/themes/culture"/>
      <dct:language>fr</dct:language>
```

```

<!-- Mots clés pour aider au référencement du jeu de données -->
<dc:keyword xml:lang="fr">Montreux Jazz Festival</dc:keyword>
<dc:keyword xml:lang="fr">Metamedia Center (MMC)</dc:keyword>
<dc:landingPage>https://metamedia.epfl.ch/</dc:landingPage>
<dc:temporal>
  <!-- Dates extrêmes du jeu de données -->
  <dc:PeriodOfTime>
    <schema:startDate
rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1966-01-
01</schema:startDate>
    <schema:endDate
rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2018-14-
07</schema:endDate>
  </dc:PeriodOfTime>
</dc:temporal>
  <!-- Mise à jour du jeu de données -->
  <dc:accrualPeriodicity
rdf:resource="http://purl.org/cld/freq/daily"/>
  <dc:distribution>
    <dc:Distribution
rdf:about="https://metamedia.epfl.ch/montreux_jazz_digital_project-about">
      <!-- Date de téléchargement du jeu de données -->
      <dc:issued
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2019-09-
04T01:00:00Z</dc:issued>
      <!-- URL de téléchargement du jeu de données -->
      <dc:accessURL
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">https://metamedia.ep
fl.ch/mjf_dump.json</dc:accessURL>
      <!-- Information sur les droits d'utilisation du jeux de
données -->
      <dc:rights>NonCommercialAllowed-CommercialAllowed-
ReferenceNotRequired</dc:rights>
      <!-- Informations sur le type d'encodage du jeu de données -->
      <dc:mediaType>application/json</dc:mediaType>
    </dc:Distribution>
  </dc:distribution>
</dc:Dataset>
</dc:dataset>
</dc:Catalog>
</rdf:RDF>

```

Annexe 13 : Exemple de fichier R2RML

```
@prefix rr: <http://www.w3.org/ns/r2rml#>. # Introduit l'adresse IRI
de l'ontologie r2rml
@prefix mjf: <http://mjf-database.epfl.ch/mjf>.

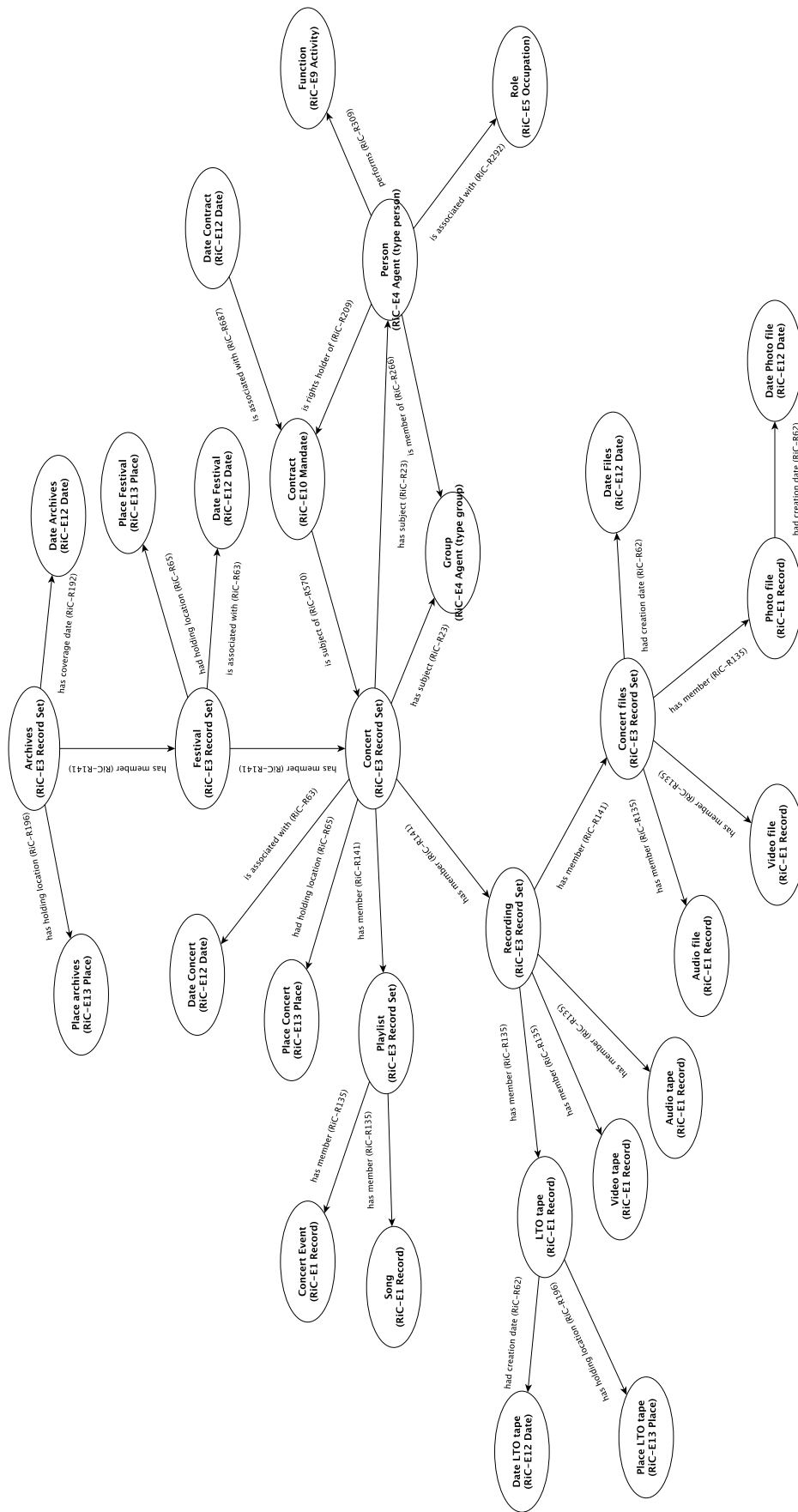
<#> #Nœud principale du graphe
  rr:logicalTable [ rr:tableName "#concert"]; # Sélectionne le nom
d'une table de la base de données relationnelle
  rr:subjectMap [ # Défini le Sujet du triplet
    rr:template "http://mjf-database.epfl.ch/mjf/concert_{id}";
# Introduit l'adresse IRI du triplet
    rr:class mjf:concert; # Introduit une nouvelle classe
  ];
  rr:predicateObjectMap [ # Défini le Prédicat du triplet
    rr:predicate mjf:name;
    rr:objectMap [ rr:column "name" ]; # Sélectionne une colonne
spécifique à partir de la table
  ];
  rr:predicateObjectMap [
    rr:predicate mjf:name;
    rr:objectMap [ rr:column "concert_date" ];
  ];
  rr:logicalTable [ rr:tableName "#composition"];
  rr:subjectMap [
    rr:template "http://mjf-
database.epfl.ch/mjf/composition_{id}";
    rr:class mjf:composition;
  ];
  rr:predicateObjectMap [
    rr:predicate mjf:name;
    rr:objectMap [ rr:column "title" ];
  ];
  rr:logicalTable [ rr:tableName "#instrument"];
  rr:subjectMap [
    rr:template "http://mjf-
database.epfl.ch/mjf/instrument_{id}";
    rr:class mjf:instrument;
  ];
  rr:predicateObjectMap [
    rr:predicate mjf:name;
    rr:objectMap [ rr:column "name" ];
  ];
  rr:logicalTable [ rr:tableName "#song"];
  rr:subjectMap [
    rr:template "http://mjf-database.epfl.ch/mjf/song_{id}";
    rr:class mjf:song;
  ];
  rr:predicateObjectMap [
    rr:predicate mjf:name;
    rr:objectMap [ rr:column "title" ];
  ];
  rr:logicalTable [ rr:tableName "#location"];
  rr:subjectMap [
    rr:template "http://mjf-database.epfl.ch/mjf/location_{id}";
    rr:class mjf:location;
  ];
  rr:predicateObjectMap [
```

```

        rr:predicate mjf:name;
        rr:objectMap [ rr:column "name" ];
];
rr:logicalTable [ rr:tableName "#phototag"];
rr:subjectMap [
    rr:template "http://mjf-database.epfl.ch/mjf/phototag_{id}";
    rr:class mjf:phototag;
];
rr:predicateObjectMap [
    rr:predicate mjf:name;
    rr:objectMap [ rr:column "name" ];
];
rr:logicalTable [ rr:tableName "#artist"];
rr:subjectMap [
    rr:template "http://mjf-database.epfl.ch/mjf/artist_{id}";
    rr:class mjf:artist;
];
rr:predicateObjectMap [
    rr:predicate mjf:name;
    rr:objectMap [ rr:column "computed_public_name" ];
];
rr:logicalTable [ rr:tableName "#band"];
rr:subjectMap [
    rr:template "http://mjf-database.epfl.ch/mjf/band_{id}";
    rr:class mjf:band;
];
rr:predicateObjectMap [
    rr:predicate mjf:name;
    rr:objectMap [ rr:column "name" ];
];
].

```


Annexe 14 : Diagramme du nouveau modèle conceptuel



Annexe 15 : Entités de l'exemple liées à la plateforme Wikidata

