

Alignement et enrichissement des données de l'inventaire d'un fonds d'archives en Linked Open Data: le cas du Montreux Jazz Digital Project

Alain Chardonnnens,
chardonnnens.alain@gmail.com

Résumé

L'évaluation du fonds des archives audiovisuelles du Montreux Jazz Digital Project (MJDP) a permis d'établir une liste de critères quantifiables et non quantifiables, en vue du transfert d'un jeu de données vers un collaborative knowledge graph. 5 critères différents sur les 27 initialement envisagés ont été appliqués avec succès sur ce fonds. Le volume des jeux de données transférables représente 5% des tables du modèle de données de la base de données relationnelle du MJDP.

La preuve de concept présentée a été adaptée, à partir d'un exemple similaire décrit dans la littérature. Les résultats obtenus ont souligné l'importance de la qualité des données et du facteur humain dans l'alignement vers Wikidata en Linked Open Data (LOD). Les identifiants de propriétés Wikidata ont permis d'effectuer une recherche fédérée vers d'autres bases de données musicales. Les limitations introduites par les outils client/serveur employés pour l'alignement ont montré l'incidence d'une recherche approximative et exacte sur les résultats de recherche.

The evaluation of the Montreux Jazz Digital Project (MJDP) audiovisual archive establishes a list of quantifiable and non-quantifiable criteria for select and transfer data sets to a collaborative knowledge graph. 5 different criteria among the 27 initially planned were successfully applied to the archive. The volume of transferable data sets concerns 5% of the data model of the MJDP relational database tables.

The proof of concept has been adapted from an example previously described in the literature. The results show the impact of data quality and human factor in a successful alignment to Wikidata with Linked Open Data (LOD). Wikidata's identifiers of properties has been used in a federated search to external music databases. The limitations introduced by client/server tools during the alignment show the impact of an approximate search in comparison with accurate search on results.

Mots-clés

Montreux Jazz Digital Project, Alignement, Linked Open Data, Wikidata, Collaborative Knowledge Graph

1. Introduction

La rédaction du présent article fait suite à la réalisation d'un travail de bachelor effectué dans le cadre de la formation en Information documentaire de la Haute Ecole de gestion de Genève (HEG-GE). Il est question dans ce travail de présenter brièvement la synthèse de quelques résultats obtenus lors de l'alignement et de l'enrichissement de jeux de données extraits de la base de données relationnelle du Montreux Digital Project (MJDP) avec la plateforme Wikidata. Cet article comporte également une revue de la littérature et un compte rendu des diverses méthodes expérimentées au cours du travail de bachelor. Il tente d'apporter une solution pratique à la problématique de l'alignement d'un jeu de données vers un collaborative knowledge graph.

Le cadre de cette étude est le MJDP du MetaMedia Center (MMC) de l'Ecole polytechnique fédérale de Lausanne (EPFL). Depuis 2010, ce centre multidisciplinaire a été mandaté par la fondation Claude Nobs, propriétaire des archives audiovisuelles du Montreux Jazz Festival (MJF) et l'EPFL pour la numérisation, la gestion et la valorisation du fonds d'archives. Depuis l'ouverture du centre, plus de 35 projets externes au MMC ont été réalisés à partir des données et des documents audiovisuels du fonds du MJDP. Ces différents projets ont participé activement à la valorisation des archives du MJF auprès d'un large public (Dufaux & Amsallem, 2017 ; 2019).

Le choix de réaliser un alignement vers Wikidata a principalement été motivé par la grande diversité des concepts décrits et par les nombreux outils disponibles pour le transfert, le traitement et la valorisation des données sur la plateforme. L'impact mondial de Wikidata et la réactivité de la communauté d'utilisateurs concernant la curation des contenus ont été des arguments supplémentaires en faveur de ce choix.

2. Revue de la littérature

Wikidata est une plateforme de crowdsourcing ou collaborative knowledge graph, 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 partagées facilitant la diffusion et réutilisation sous licence Creative Commons CC0.
- Une édition libre et participative autogérée par les membres de la communauté d'utilisateurs de la plateforme.
- La création, l'édition, la recherche et la réutilisation de données multilingues.
- La possibilité de conserver une ambiguïté de sens dans la description des concepts sur la plateforme.
- La réutilisation des données grâce à l'utilisation de données structurées.
- La centralisation et l'homogénéisation des concepts décrits dans d'autres projets de la fondation Wikimedia.

Un workflow édité par la communauté GLAM (galleries, libraries, archives and museums) de Wikimedia documente les principales étapes nécessaires à un transfert de données vers les plateformes Wikimedia Commons et Wikidata (GLAM/Resources/Data and media partnerships workflow, 2019). Ce processus est découpé en quatre lots de tâches. Il s'agit des étapes intitulées pré-chargement, chargement, après-chargement et impact.

WikiProject Music (Wikidata:WikiProject Music, 2019) est un projet collaboratif et thématique de Wikidata. Le projet référence les bases de données documentaires et commerciales relatives au domaine de la musique. Une liste des identifiants des propriétés alignés avec la plateforme est régulièrement mise à jour par les membres du projet.

L'API Wikidata reconciliation (Wikimedia toolforge, sans date a) rend possible, du côté client, la réalisation de deux scénarios sur les cinq décrits par Delpeuch (Delpeuch, 2019 : p. 11). Le premier scénario s'apparente au Current OpenRefine reconciliation API proposé par Delpeuch (2019). Dans ce scénario, aucun transfert préalable de données vers une plateforme de crowdsourcing n'est nécessaire afin d'accomplir un alignement vers Wikidata. Le client réalise l'alignement des données directement depuis un poste de travail en local. Les logiciels open source OpenRefine ou l'interface OntoRefine (Ontotext, sans date) du logiciel de base de données de graphe GraphDB peuvent être utilisés afin d'accomplir cette opération.

Le second scénario (Server-side dataset matching) est applicable une fois les données transférées vers Wikidata. L'outil Mix'n'match offre la perspective d'un alignement automatique et/ou manuel par le crowdsourcing des membres de la communauté de Wikidata. Selon Zeng, cet outil constitue une source importante de données culturelles authentifiées (Zeng, 2019 : p. 12). Piscopo et al. (2017 : p. 544) définissent trois principaux types d'autorité sur Wikidata : il s'agit des types individuels (une ou plusieurs personnes directement identifiables), organisations (une entité directement identifiable) et collective (un groupe d'utilisateurs identifié uniquement par son nom d'utilisateur). Le dernier type n'est pas considéré à proprement parler comme une source d'autorité selon Piscopo et al. (2017), puisque les auteurs ne sont pas directement identifiables. Les résultats obtenus par Piscopo et al. montrent que les institutions GLAM représente 15 % des contributions au référencement de sources dans les publications de Wikidata. La répartition de cette contribution est assurée par des humains (7,7 %) ou à l'aide de robot indexeur (11,9 %). Les agences gouvernementales arrivent en tête des contributions sur la plateforme avec 37,7 %.

Le transfert des jeux de données vers Wikidata peut s'effectuer de manière asynchrone ou synchrone. L'outil QuickStatements (Wikimedia toolforge, sans date b) permet directement le transfert asynchrone d'un lot de données depuis un navigateur web. Le logiciel PAWS: A Web Shell (PAWS) (PAWS, 2019) comprend un éditeur de texte enrichi en langage Markdown et offre la possibilité d'éditer des codes sources en langage python. Cet outil a été développé à partir du modèle de notebook Jupyter et offre la possibilité d'effectuer une synchronisation directe des données depuis une base de données relationnelle en local. Cependant, ce cas d'utilisation est réalisable suite à l'ajout des plusieurs librairies comme décrit sur la page du projet : <https://wikitech.wikimedia.org/wiki/PAWS>

L'alignement des jeux de données d'un catalogue en local vers Wikidata est décrit dans une preuve de concept réalisée par Allison-Cassin et Scott (2018). Les fonctions proposées par cette preuve de concept sont : l'interrogation distante, l'enrichissement et l'affichage des données relatives à un individu. L'exemple proposé permet d'afficher, à partir du nom et du

rôle d'un musicien, différents champs qui incluent le prénom, le nom, la photo, le rôle, l'instrument, le lieu d'origine, les adresses de contacts (site web, réseaux sociaux) et l'identifiant de l'entité Wikidata. De plus, un court extrait résume le contenu de la page anglaise de Wikipédia relative au musicien. Le code source de cette preuve de concept est téléchargeable à partir d'un répertoire Github : <https://gitlab.com/denials/wikidata-music-infocard>

3. Alignement et enrichissement des données

La première étape de la méthodologie présentée dans cet article comprend un état des lieux exhaustif de la typologie des documents et des données du MJDP. Cette étape propose la création de trois inventaires distincts comprenant les supports physiques, les formats des enregistrements numériques et les types de données de la base de données. L'analyse des documents et des données permet la création d'une liste de critères quantifiables et non quantifiables du fonds vers Wikidata.

Les risques liés à une fausse manipulation sur la base de données de production sont réduits grâce un export des jeux de données en langage SQL. Le fichier exporté est analysé sur un poste de travail en local à l'aide du logiciel open source MAMP. Ce logiciel permet l'émulation d'un serveur web Apache exploitant le système de gestion de base de données relationnelle (SGBDR) MySQL. Les opérations d'extraction, de transformation et de chargement (ETL) sur les données sont effectuées à partir de l'interface phpMyAdmin.

La seconde étape de ce travail concerne la recherche, le choix et la création d'une liste d'identifiants d'API de bases de données musicales alignés avec Wikidata. Le projet WikiProject Music répertorie un large choix d'identifiants de propriétés concernant cette thématique sur Wikidata. Cependant, il existe une liste complète des identifiants des propriétés de Wikidata répertoriant les différents concepts décrits sur la plateforme : https://www.wikidata.org/wiki/Wikidata:Database_reports/List_of_properties/all

La troisième étape propose l'application pratique des critères quantifiables et non quantifiables au schéma de la base de données du MJDP. La sélection des tables et des colonnes du MJDP est effectuée, à la suite de trois tours de choix. Ensuite, les colonnes sélectionnées sont reportées dans un tableau en fonction des propriétés sélectionnées manuellement dans l'étape précédente.

La quatrième étape propose l'export en format CSV d'un échantillon des colonnes sélectionnées. Les échantillons exportés sont importés dans la base de données de graphe à partir de l'interface OntoRefine du logiciel GraphDB.

La cinquième étape concerne l'alignement des colonnes importées avec les données disponibles sur la plateforme Wikidata et/ou en provenance d'autres bases de données externes. L'alignement des données est effectué dans l'interface OntoRefine à l'aide de l'API Wikidata reconciliation.

L'interface OntoRefine propose deux cas d'utilisations pour l'alignement vers Wikidata. Le premier cas d'utilisation prévoit l'interrogation directe de l'API Wikidata reconciliation afin de choisir automatiquement et/ou manuellement l'identifiant d'une entité (Qxxx) Wikidata. Il s'agit ensuite de valider l'identifiant se rapprochant le plus sémantiquement des valeurs comprises dans la colonne proposée par l'API. Le second cas d'utilisation est accessible une fois que la

colonne est alignée avec Wikidata. Ce deuxième cas d'utilisation offre la possibilité d'ajouter un ou plusieurs identifiants de propriétés (Pxxx) Wikidata aux valeurs comprises dans la colonne.

Le dernier cas d'utilisation concerne l'alignement à partir d'une ou plusieurs colonnes comportant des jeux de données disponibles en local. Le but de cette opération est de contextualiser la valeur du terme recherché. Cette colonne est introduite à l'aide de l'identifiant d'une propriété Wikidata. Ce cas d'alignement est appliqué aux données concernant le nom d'un musicien joint avec les valeurs des instruments de musique décrits dans la base du MJDP. Une fois les données authentifiées et validées par l'utilisateur, les colonnes sont exportées et sauvegardées dans des fichiers CSV séparés. Les fichiers CSV alignés peuvent être ensuite transférés directement vers Wikidata avec l'outil QuickStatements.

La dernière étape de cette méthodologie concerne l'enrichissement des données du fonds MJDP en local comme décrit par la preuve de concept d'Allison-Cassin et Scott (2018). Cet enrichissement est obtenu grâce aux données collectées avec l'API du Wikidata Query Service : <https://query.wikidata.org>. La preuve de concept est modifiée au niveau des balises html auxquelles sont ajoutés les jeux de données exportés de la base de données du MJDP. Le code source de la requête SPARQL compris dans le fichier javascript est également adapté en fonction de notre besoin.

4. Résultats

Les inventaires réalisés ont permis d'établir une liste de 27 critères différents correspondant à la répartition suivante : 18 critères quantifiables et 9 critères non quantifiables (tableau 1). 19 identifiants de propriétés de bases de données spécifiques au domaine musical ont été sélectionnés à partir des sources consultées. Des 27 critères définis initialement pour l'évaluation du fonds, 5 critères ont été retenus pour le regroupement pratique des tables de la base de données du MJDP.

Tableau 1 : Liste des critères non quantifiables

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

Le résultat de l'évaluation du fonds du MJDP a permis de faire émerger un total de 9 colonnes de type chaîne de caractère distribué dans 8 tables, soit 5% des 160 tables de la base de données du MJDP. La taille de l'échantillon comprenant les noms des musiciens exportés de la base de données du MJDP est composé d'un total de 29'554 lignes.

Dans l'interface OntoRefine, l'alignement automatique de la colonne concernant les données d'un musicien (nom, prénom, pseudonyme, initiales) a été identifié par l'API Wikidata reconciliation comme appartenant à l'identifiant de l'entité human (Q5) de Wikidata. La colonne incluant des intitulés d'instruments de musique a été jointe à la colonne contenant les données de musicien avec l'identifiant de la propriété P1303 (instrument de musique).

A la suite du premier alignement de la colonne avec Wikidata, le second alignement montre que l'identifiant de la propriété MusicBrainz artist ID (P434) possède une occurrence avec 90% des valeurs des résultats. Ce résultat s'élève à 30% pour les propriétés Discogs artist ID (P1953) et AllMusic artist ID (P1728).

La figure 1 illustre un extrait du résultat de l'enrichissement des données obtenues à partir du nom de deux musiciens de la base de données du MJDP. On observe dans ce résultat que les identifiants des propriétés MusicBrainz artist ID, Discogs artist ID et AllMusic artist ID ont été ajoutés à la requête SPARQL du code source du fichier javascript de la preuve de concept d'Allison-Cassin et Scott (2018).

Figure 1 : Résultats de l'enrichissement des données du MJDP avec Wikidata

	Carlos Santana Mexican and American rock musician Instruments: voice; guitar Birth place: Jalisco Web site: http://www.santana.com MusicBrainz artist ID AllMusic artist ID Discogs artist ID Edit on Wikidata: Q819016 Carlos Santana (born July 20, 1947) is a Mexican and American guitarist who rose to fame in the late 1960s and early 1970s with his band Santana, which pioneered a fusion of rock and roll and Latin American jazz. ^W
	Bill Evans American jazz pianist Instruments: piano Birth place: Plainfield MusicBrainz artist ID AllMusic artist ID Discogs artist ID Edit on Wikidata: Q208205 William John Evans (August 16, 1929 – September 15, 1980) was an American jazz pianist and composer who mostly played in trios. His use of impressionist harmony, inventive interpretation of traditional jazz repertoire, block chords, and trademark rhythmically independent, "singing" melodic lines continue to influence jazz pianists today. ^W

5. Discussion des résultats

L'utilisation d'un nombre restreint de critères quantifiables et non quantifiables semble être une stratégie à privilégier afin de gagner en efficacité dans l'évaluation. Une classification avec une granularité élevée ne semble pas être optimale dans ce contexte précis. En effet, une majorité des critères précédemment définis a été jugée trop spécifique pour les besoins de l'évaluation pratique du fonds. Ces critères ont été écartés, à la suite d'une première tentative de regroupement. Le résultat de l'évaluation montre avant tout que l'ordre de grandeur des tables partageables pour la base de données du MJDP se situe aux alentours de 5%.

L'alignement manuel d'un échantillon de données du MJDP avec Wikidata a montré une très bonne occurrence concernant les concepts relatifs à des données biographiques de musicien. Les termes désignant des instruments ou des lieux géographiques ont également offerts de bons résultats. Cela montre qu'il existe un intérêt marqué de la part de la communauté pour ces thématiques. Les jeux de données testés concernant les noms des musiciens provenaient de données précédemment authentifiées par des opérateurs humains. En revanche, les noms d'instruments testés étaient lacunaires dans de nombreux cas.

Le paramétrage manuel d'un identifiant de propriété par l'utilisateur dans OntoRefine augmente sensiblement la précision de l'algorithme pour la mesure de la distance de Levenshtein d'un terme aligné avec Wikidata. Le résultat de cette mesure s'affiche dans

OntoRefine sous la forme d'un score. Plus le score de l'alignement est élevé, plus le terme aligné possède une proximité sémantique avec l'identifiant de l'entité disponible sur Wikidata.

En revanche, l'API est peut performante lorsque divers concepts sont mélangés dans une même colonne. Dans ce cas de figure, un choix doit être opéré par l'opérateur en faveur d'un identifiant d'entité générique. Ce choix a pour conséquence de favoriser un concept en particulier, au détriment d'autres concepts secondaires compris dans la colonne. Il en résulte une perte de précision importante et l'affichage de résultats lacunaires ou ambigus.

Dans certains cas, on obtient un résultat ambigu pour un même terme aligné avec la valeur d'un ou plusieurs identifiants de propriétés externe à Wikidata. Ce cas de figure a fréquemment été observé à partir de la valeur d'un terme précédemment aligné avec une entité de Wikidata. Ces termes concernaient les identifiants de propriétés externes en provenance d'autres bases de données comme MusicBrainz artist ou Discogs. Ces résultats ambigus peuvent s'expliquer en partie par des erreurs de référencements manuel et/ou automatique des URL de ces plateformes distantes sur Wikidata. La valeur d'un identifiant de propriété externe peut également être très fortement influencée par l'ordre du classement des valeurs référencées pour un même concept. En effet, plusieurs valeurs sémantiquement proches et/ou éloignées d'un même concept peuvent être référencées sur la page Wikidata d'une entité. Dans ce cas, seul l'ordre du classement valide le sens du terme le plus communément accepté par la communauté d'utilisateurs (Vrandečić, 2013 : p. 90).

L'API Wikidata reconciliation utilise un mode de recherche fondé sur la recherche approximative ou fuzz search. Chaque terme recherché est dépendant d'une orthographe définie par la forme retenue du terme correspondant sur Wikidata. Par conséquent, chaque terme ou groupe de termes alignés avec un identifiant d'une entité ou d'une propriété externe nécessite une vérification et une authentification manuelle par un opérateur humain avant validation. En effet, l'API limite significativement l'automatisation de cette tâche, puisqu'elle ne permet pas d'assurer la fiabilité des données alignées à partir d'un score élevé de la mesure de la distance de Levenshtein. Cependant, les récents travaux entrepris par Delpeuch (2019 : p. 12) proposent une amélioration de la précision de l'API Wikidata reconciliation grâce à l'enregistrement et la réutilisation des scores de Levenshtein obtenus par les utilisateurs suite à l'alignement des termes. Cette nouvelle perspective offre la possibilité d'augmenter sensiblement la précision de l'alignement automatique du service et permet une potentielle intégration du machine learning.

La preuve de concept proposée par Allison-Cassin et Scott (2018) a été adaptée avec plus ou moins de succès à notre cas d'utilisation. L'enrichissement des données réalisé avec la preuve de concept a montré la faisabilité d'une utilisation de Wikidata comme un répertoire central d'identifiant et un outil de recherche fédéré vers plusieurs bases de données. Cependant, la requête SPARQL adapté dans la preuve de concept ne semble pas autoriser un mode de recherche approximatif comme cela est le cas pour l'API Wikidata reconciliation. En effet, seul le premier résultat d'une requête SPARQL est retourné sur les dix pouvant être potentiellement affiché dans le fichier html par l'API du Wikidata Query Service. Le nombre de résultat retourné pour une requête est limité pour des raisons de performance.

Si la requête SPARQL ne retourne aucun résultat, rien ne s'affiche dans le fichier html. En revanche, si plusieurs résultats sont disponibles à la suite d'une même requête, seul le premier résultat s'affiche dans le fichier html. Ce mode de fonctionnement limite drastiquement les

possibilités d'afficher des résultats ambigus comme, par exemple, plusieurs musiciens possédant un même patronyme. Par conséquent, les noms recherchés sont dépendants de la forme orthographique choisie par l'intitulé anglais de l'identifiant de l'entité Wikidata. Cependant, l'intitulé anglais de l'identifiant de l'entité d'un instrument de musique permet de contextualiser un terme de recherche et de retourner un résultat correct, nul ou erroné.

Une piste pourrait être envisagée avec l'alignement du vocabulaire de l'actuelle taxonomie utilisée pour la description des instruments de musique du MJDP avec des concepts équivalents décrits sur Wikidata. Dans ce cas, une nouvelle modélisation doit être envisagée à partir du diagramme de classes du MJDP concernant les instruments de musique vers le vocabulaire d'ontologie Simple Knowledge Organization System (SKOS). A la suite de la transcription du vocabulaire de la taxonomie en SKOS, un alignement vers le Wikidata Query Service peut être facilement envisageable. La similarité des termes obtenue par cet alignement permettrait d'améliorer significativement la fiabilité de la mesure et la réduction des biais de recherche.

6. Conclusion

Premièrement, la qualité des données partagées est un élément déterminant pour la réussite d'un alignement vers la plateforme Wikidata. Cela implique un effort important de vérification manuelle des données de la part du producteur, afin d'en garantir l'authenticité et la similarité orthographique avec l'entité décrite sur Wikidata. Il a été démontré qu'il était néanmoins envisageable de procéder à un alignement du côté client et/ou server avec des données peu ou partiellement vérifiées avec les outils OntoRefine ou Mix'n'match.

Deuxièmement, l'opérateur humain est toujours actuellement déterminant lors de l'étape de l'alignement dans le choix, la vérification, l'authentification et la validation de la référence de l'identifiant de l'entité proposée approximativement par l'API Wikidata reconciliation. Nous avons constaté qu'un score élevé de la mesure de la distance de Levenshtein n'était pas une garantie suffisante pour valider de manière fiable une valeur alignée. Cependant, les travaux de Delpeuch (2019) ont montré que l'amélioration de la précision de l'API Wikidata reconciliation pourrait être envisagée grâce à l'utilisation du machine learning.

Nous avons également abordé dans cet article la possibilité d'enrichir les données d'un inventaire local en LOD sans pour autant procéder à un transfert de données vers Wikidata. Dans ce cas de figure, nous avons expérimenté une utilisation de Wikidata comme un répertoire central de données offrant la possibilité d'effectuer des recherches fédérées vers d'autres bases de données.

Les perspectives soulevées par ce travail sont multiples et prévoient une étude statistique, afin de quantifier le taux de données actuelles du MJDP pouvant être directement aligné avec Wikidata et/ou d'autres bases de données musicales ; une implémentation de la preuve de concept à l'actuel framework play du MJDP ; la modélisation de la taxonomie des instruments de musique vers le vocabulaire d'ontologie SKOS ; l'alignement des données concernant les instruments de musique et œuvres musicales des concerts du MJF vers Wikidata ; enfin, le développement d'une interface inspirée du modèle proposé par l'interface visuelle de l'outil Mix'n'match, afin de faciliter la vérification et l'authentification manuelle des données alignées en local par des opérateurs humains.

BIBLIOGRAPHIE

- ALLISON-CASSIN, Stacy et SCOTT, Dan, 2018. Wikidata : a platform for your library's linked open data [en ligne]. 4 mai 2018. Code4lib Journal. Vol. 40. [Consulté le 20 octobre 2019]. Disponible à l'adresse : <https://journal.code4lib.org/articles/13424>
- DELPEUCH, Antonin, 2019. A survey of OpenRefine reconciliation services [en ligne]. 21 août 2019. [Consulté le 8 octobre 2019]. Disponible à l'adresse : <https://arxiv.org/pdf/1906.08092>
- 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 20 octobre 2019]. Disponible à l'adresse : http://www.ressi.ch/num18/article_138
- DUFAUX, Alain et AMSALLEM, Thierry, 2019. The Montreux Jazz Digital Project: From preserving heritage to a platform for innovation. Journal of Digital Media Management. 1 janvier 2019. Vol. 7, n° 4. pp. 315-329. [Consulté le 20 octobre 2019].
- 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 20 octobre 2019]. Disponible à l'adresse : https://outreach.wikimedia.org/wiki/GLAM/Resources/Data_and_media_partnerships_workflow
- ONTOTEXT, [sans date]. Loading data using OntoRefine [en ligne]. 19 juin 2019. [Consulté le 20 octobre 2019]. Disponible à l'adresse : <http://graphdb.ontotext.com/documentation/free/loading-data-using-ontorefine.html>
- PAWS. MediaWiki [en ligne]. Dernière modification de la page le 14 octobre 2019 à 17:17. [Consulté le 8 octobre 2019]. Disponible à l'adresse : <https://www.mediawiki.org/wiki/PAWS>
- PISCOPO, Alessandro et al., 2017. Provenance Information in a Collaborative Knowledge Graph: An Evaluation of Wikidata External References. The Semantic Web – ISWC 2017. C. d'Amato et al. (Eds.). 4 octobre 2017. pp. 542–558. ISBN 978-3-319-68288-4. Disponible à l'adresse : https://doi.org/10.1007/978-3-319-68288-4_32
- VRANDEČIĆ, Denny, 2013. The Rise of Wikidata. 2013. IEEE Intelligent Systems. Vol. 28, n°4. pp. 90-95. ISSN 1541-1672. Disponible à l'adresse : <https://doi.org/10.1109/MIS.2013.119>
- WIKIMEDIA TOOLFORGE, [sans date a]. OpenRefine-Wikidata interface [en ligne]. [Consulté le 20 octobre 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/openrefine-wikidata>
- WIKIMEDIA TOOLFORGE, [sans date b]. QuickStatements [en ligne]. [Consulté le 20 octobre 2019]. Disponible à l'adresse : <https://tools.wmflabs.org/quickstatements/#/>
- Wikidata:WikiProject Music. Wikidata [en ligne]. Dernière modification de la page le 17 septembre 2019 à 02:20. [Consulté le 20 octobre 2019]. Disponible à l'adresse : https://www.wikidata.org/wiki/Wikidata:WikiProject_Music
- 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 20 octobre 2019]. Disponible à l'adresse : <https://doi.org/10.3145/epi.2019.ene.03>