

Exploitation et réutilisation des données : mise en perspective avec le patrimoine numérique de la BCU Lausanne

Le cas de la musique écrite

Noémie Payot

Cet article propose une réflexion sur la mise à disposition des collections de partitions avec pour objectif de faciliter leur réutilisation dans le cadre de projets ayant recours aux humanités numériques. En partie une et deux, les principes FAIR ainsi que les notions d'Open Data et de Linked Open Data sont mises en lien avec les plateformes utilisées par la BCU Lausanne pour mettre à disposition ses collections patrimoniales, dont en particulier le fonds des Archives musicales. La troisième et dernière partie compare les méta-catalogues de partitions disponibles en ligne afin d'évaluer dans quelle mesure ces méta-catalogues permettent la réutilisation de leurs données. Enfin, quelques exemples de projets exploitant des corpus de partitions sont présentés.

Dieser Artikel stellt Überlegungen zur Bereitstellung von Partitursammlungen an mit dem Ziel, die Nutzung im Rahmen von Digital Humanities-Projekten zu erleichtern. Im ersten und zweiten Teil werden die FAIR-Prinzipien sowie die Begriffe Open Data und Linked Open Data in Verbindung mit den Plattformen und ihrer Anwendung in der BCU Lausanne untersucht, dies mit Bezug zur Bereitstellung der patrimonialem Sammlungen, insbesondere der Bestände des Musikarchivs. Im dritten und letzten Teil werden die online verfügbaren Metakataloge für Partituren miteinander verglichen, um zu bestimmen, inwieweit diese Metakataloge zur Wiederverwendung ihrer Daten beitragen. Schliesslich werden exemplarisch einige Projekte vorgestellt, die Notencorpora auswerten.

This article looks at how score collections can be made available, with the aim of facilitating their re-use as part of projects involving the digital humanities. In parts one and two, the FAIR principles and the concepts of Open Data and Linked Open Data are examined with respect to the platforms used by the BCU Lausanne to make its heritage collections available, including the Music Archives, in partic-

ular. The third and final part compare the meta-catalogues of score corpora available online in order to assess the extent to which these meta-catalogues allow their data to be re-used. Finally, a few examples of projects using score corpora are presented.

1 Derrière le concept de mise à disposition

La Bibliothèque cantonale et universitaire de Lausanne (BCUL) possède des collections patrimoniales importantes et variées qui sont accessibles via diverses plateformes.¹ La numérisation, la préservation et la mise en valeur des collections patrimoniales font partie des missions de la BCUL² qui poursuit depuis longtemps l'objectif d'être une institution novatrice dans le domaine numérique.³ Toutefois, la notion-même de mise à disposition peut prendre différentes formes et doit faire face à des problématiques multiples qui dépendent du type de collection, de la licence ainsi que des outils et des moyens disponibles. Aujourd'hui, la mise à disposition peut ne concerner que la notice du document (soit des métadonnées qui lui sont associées), mais de plus en plus, il s'agit souvent de la notice et du document lui-même (document numérisé ou né numérique). Or, la mise à disposition de documents numériques sous un format exploitable pour la machine⁴ est une nouvelle étape importante, car permettant aux utilisateurs d'accéder directement à l'information elle-même en se passant du médium physique.

Cette thématique s'inscrit dans plusieurs domaines dont ceux de l'Open Data et de l'Open Science qui s'englobent de manière mutuelle mais partielle, puisque l'Open Science s'applique spécifiquement aux données de la recherche. L'Open Data a notamment pour objectif de garantir que les institutions publiques partagent le savoir dont elles sont dépositaires comme un bien commun. Plus précisément, ces principes impliquent un partage sous une licence ouverte, de manière gratuite, sous une forme lisible par la machine, et dans un format ouvert (par exemple PDF, CSV, TXT, etc.).⁵ Ces éléments recourent

1 Lien vers les collections de la BCUL, <<https://www.bcu-lausanne.ch/collections-et-patrimoine>>, consulté le 26.6.2023.

2 *La BCUL en 2025. Mission, vision et visite virtuelle*, 2020, p. 3, <<https://www.bcu-lausanne.ch/wp-content/uploads/2022/02/plan-directeur-abrege-2021-25.pdf>>, consulté le 26.6.2023.

3 « Qui sommes-nous? », *BCUL*, sect. Historique, <<https://www.bcu-lausanne.ch/qui-sommes-nous/>>, consulté le 26.06.2023.

4 Pour les documents numérisés, les technologies OCR (Optical Character Recognition) et HTR (Handwritten Text Recognition) permettent de rechercher l'information dans le texte et non plus uniquement dans les métadonnées.

en partie les principes FAIR (Findable Accessible Interoperable Reusable)⁶ qui, bien qu'élaborés pour mettre en œuvre l'Open Science, sont également pertinents pour les institutions culturelles.⁷ En particulier, les points concernant l'interopérabilité et la réutilisation des données revêtent une grande importance pour le partage du savoir.

Une des solutions techniques permettant de mettre en œuvre les principes FAIR réside dans le concept de linked data (ou données liées, également connu sous les noms de web de données ou de web sémantique). Antérieur de 10 ans aux principes FAIR, ce concept a principalement été promu par l'inventeur du World Wide Web, Tim Berners-Lee.⁸ Le linked data se focalise sur l'aspect de l'interopérabilité entre les données, ce concept est donc inclus dans le « i » de FAIR.⁹ Les principes linked data appliqués à l'Open Data sont à l'origine du mouvement Linked Open Data (LOD) dont l'objectif est de publier des données à la fois ouvertes et interopérables. Il convient néanmoins de distinguer les principes FAIR de ceux du LOD. En effet, des données ouvertes sont requises pour le LOD alors que FAIR implique uniquement de déclarer la licence.¹⁰ En outre, les métadonnées jouent un rôle clef pour le FAIR tandis que le LOD considère les métadonnées elles-mêmes comme des données interopérables.¹¹ Enfin, le LOD utilise les URIs comme identifiants et propose le modèle RDF ainsi que le protocole SPARQL pour accéder aux données. FAIR, de son côté, ne précise aucune technologie spécifique.

On constate depuis quelques années que les technologies LOD sont de plus en plus utilisées par les grands méta-catalogues comme DBpedia, Europeana, VIAF, GeoNames ou encore Getty vocabularies.¹² En effet, SPARQL permet d'interroger des ensembles importants de données et d'en extraire des in-

5 « Open Definition 2.1. Defining Open in Open Data, Open Content and Open Knowledge », *Open Knowledge Foundation*, <<https://opendefinition.org/od/2.1/>>, consulté le 26.06.2023.

6 Jansen, Mascha, « FAIR Principles », *GO FAIR*, <<https://www.go-fair.org/fair-principles/>>, consulté le 26.06.2023.

7 Le webinaire *Publishing GLAM Data as FAIR Data*, organisé par la Research Data Alliance (RDA) et le Digital Repository of Ireland (DRI), est un bon exemple de cette mouvance. Voir : Harrower, Natalie ; Stanley, Roy ; Scheltjens, Saskia et al., *Publishing GLAM data as FAIR data*, Webinaire en ligne, 2020, <<https://www.rd-alliance.org/publishing-glam-data-fair-data>>, consulté le 26.06.2023.

8 Berners-Lee, Tim, « Linked Data », *W3C*, 2006, <<https://www.w3.org/DesignIssues/LinkedData.html>>, consulté le 26.06.2023.

9 Avanço, Karla, « FAIR Principles and Linked Open Data », Billet, *The road to FAIR*, part. LOD and FAIR, <<https://roadtofair.hypotheses.org/288>>, consulté le 26.06.2023.

10 Avanço, *The road to FAIR*, part. LOD and FAIR.

11 Avanço, *The road to FAIR*, part. LOD and FAIR.

12 Sugimoto, Go, *Open Data Empowerment of Digital Humanities by Wikipedia/DBpedia Gamification and Crowd Curation – WiQiZi's Challenges with APIs and SPARQL*, Mexico 2019, p. 1, <<https://doi.org/10.5281/zenodo.3465654>>, consulté le 26.06.2023.

formations selon des critères précis. La démocratisation de ces technologies facilite la recherche dans le cadre des humanités numériques en permettant d'extraire de la connaissance à partir de corpus de données textuelles, iconographiques, et, de plus en plus, à partir de ressources sonores ou vidéo.

Le plan directeur de la BCUL abonde dans le sens de l'Open Data en termes de mise à disposition gratuite et de partage du savoir en précisant que l'institution « continuera la numérisation, la préservation et la mise en valeur de ses collections patrimoniales et les mettra à disposition au travers de plateformes numériques ». ¹³ Ces éléments sont explicités par la politique des bibliothèques numériques de la BCUL qui mentionne que les plateformes de dépôt sont sélectionnées en fonction de leur pertinence pour la collection concernée, de leurs fonctionnalités, de leur durabilité, de leur coût ainsi que de leur apport par rapport aux plateformes internes de la BCUL. ¹⁴ Quant à l'interopérabilité, le protocole OAI-PMH ¹⁵ et les APIs définies par les standards IIIF ¹⁶ (International Image Interoperability Framework) sont les technologies utilisées ou envisagées. ¹⁷

2 Les plateformes numériques de la BCUL face aux principes FAIR

La BCUL possède des collections patrimoniales de types très variés, à savoir les Archives musicales, le Dépôt légal numérique, la Documentation vaudoise, l'Iconopôle, les Manuscrits et archives, le Patrimoine sonore et la Réserve précieuse. Ces collections incluent toutes en partie des documents numérisés ou nés numériques. Les documents numériques eux-mêmes sont généralement accessibles sur l'une des deux plateformes propres à la BCUL, Patrinum (PATRIMOINE NUMÉRIQUE) ou Scriptorium.

En outre, la BCUL participe à de nombreux méta-catalogues. En 2011, elle utilise une base interne dédiée aux autorités vaudoises pour créer 800 articles dans l'encyclopédie Wikipédia. ¹⁸ En 2012, près de 100'000 imprimés antérieurs à 1871 et libres de droits sont mis à disposition sur Google Books. ¹⁹ Puis, en 2020,

13 *La BCUL en 2025*, p. 3.

14 *Politique des bibliothèques numériques*. BCU Lausanne, 2019, Document non publié interne à l'institution.

15 *Open Archives Initiative Protocol for Metadata Harvesting*, <<https://www.openarchives.org/pmh/>>, consulté le 26.06.2023.

16 *International Image Interoperability Framework*, <<https://iiif.io/>>, consulté le 26.06.2023.

17 *Politique des bibliothèques numériques*, p. 5.

18 « Projet Valdensia », *Wikipédia*, 2019, <<https://fr.wikipedia.org/w/index.php?title=Projet:Valdensia&oldid=160466893>>, consulté le 26.06.2023.

l'institution verse 15 manuscrits sur la plateforme e-codices.²⁰ Enfin, quelques 120 imprimés anciens sont ajoutés sur e-rara²¹ en 2021.

Aujourd'hui, la BCUL participe également au méta-catalogue de partitions RISM²² (Répertoire International des Sources Musicales), où environ 1'300 pièces ont été versées, ainsi qu'à CCSA²³ (Catalogue Collectif des Affiches Suisses), qui présente plus de 800 affiches numérisées issues de l'Iconopôle. Il est aussi prévu que la BCUL dépose des archives audiovisuelles sur le portail suisse Memobase.²⁴ Or, à l'exception de Google books et Wikipédia, tous ces méta-catalogues respectent ou sont proches de respecter les principes FAIR.

2.1 Scriptorium et Patrinum

Scriptorium a pour but de donner un accès public à la presse vaudoise numérisée ainsi qu'aux annuaires, almanachs, et Bulletins du Grand Conseil. La plateforme inclut également quelques périodiques suisses publiés hors du canton (documents issus de la Réserve précieuse). Scriptorium a été lancé en 2012 et repose sur la solution MediaINFO.²⁵

Environ 400 utilisateurs consultent chaque jour ce catalogue qui regroupe environ 180 périodiques et plus de 560'000 fascicules.²⁶ Les documents consultables sur Scriptorium sont sous droit d'auteur et la convention établie entre la BCUL et les éditeurs permet une utilisation dans le cadre privé et non commercial.²⁷

Patrinum est un projet lancé en 2018 avec un double objectif : servir de système de gestion pour les collections de type archive gérées par la BCUL (remplacement de nombreuses bases de données internes) et permettre le stockage

19 « 100'000 livres de la BCU de Lausanne en ligne », RTS, 2009, <<https://www.rts.ch/info/sciences-tech/1043679-100000-livres-de-la-bcu-de-lausanne-en-ligne.html>>, consulté le 26.06.2023.

20 Méta-catalogue suisse de manuscrits médiévaux et modernes. Voir les collections de la BCUL sur e-codices : <<https://www.e-codices.unifr.ch/fr/search/?iCurrent-Page=1&s>>, consulté le 26.06.2023.

21 Méta-catalogue suisse pour les imprimés anciens. Voir les collections de la BCUL sur e-rara : <https://www.e-rara.ch/nav/index?&facets=domain-Group%3D%22lac1_g%22>, consulté le 26.06.2023.

22 Voir les collections de la BCUL sur RISM : <<https://rism.online/institutions/30000526/sources>>, consulté le 26.06.2023.

23 Voir les collections de la BCUL sur CCSA, <<https://www.posters.nb.admin.ch/discovery/search?query=any,contains,%3F>>, consulté le 26.06.2023.

24 MEMORIAV MEMOBASE, <<https://memobase.ch>>, consulté le 26.06.2023.

25 MediaINFO Digital Library, <<https://www.mediainfo.com/>>, consulté le 26.06.2023.

26 Chiffres de juillet 2022.

27 Scriptorium, <<https://scriptorium.bcu-lausanne.ch/page/bienvenue>>, consulté le 26.06.2023.

pérenne des documents numériques. Un troisième objectif s'y est greffé par la suite, à savoir : mettre à disposition du public les métadonnées et les documents numériques. Cette plateforme est basée sur le framework *INVENIO*,²⁸ également utilisé pour le SIGB (Système Intégré de Gestion de Bibliothèques) *RERO ILS*.²⁹

Patrinum est essentiellement utilisé pour l'ensemble des collections patrimoniales de la BCUL, à l'exception de la Réserve précieuse. On y trouve environ 250'000 notices, dont quelques 75'000 notices d'autorité.³⁰ Près de 28'000 notices disposent d'un document numérique attaché. Le nombre de notices visibles pour les lecteurs est par contre bien inférieur, à savoir 62'000 notices dont 45'000 notices d'autorité.³¹ Chaque document numérique est associé à une licence qui précise le cadre légal de son utilisation.³²

Au niveau des formats de description, *Scriptorium* utilise le Dublin Core. Les champs de métadonnées sont par contre entièrement personnalisables, le système ne respecte donc pas véritablement ce format. De son côté, *Patrinum* est conçu comme un SIGB et est donc basé sur le format MARC. Néanmoins, le logiciel a été adapté pour permettre de décrire des fonds d'archives, ce qui a nécessité l'ajout de la notion d'arborescence. Les collections sur *Patrinum* sont donc divisées en deux groupes :

- Bibliothèques : format MARCXML. Cela concerne principalement les collections du Dépôt légal numérique et de la Documentation vaudoise.
- Archives : format MARCXML adapté selon ISAD(G). Les collections des Archives musicales, de l'Iconopôle, des Manuscrits et archives ainsi que du Patrimoine sonore y sont regroupées.

Comme on le constate, *Patrinum* et *Scriptorium* sont deux plateformes qui ont vu le jour avec des objectifs initiaux très différents. En particulier *Patrinum* fait le grand écart autant au niveau de ses objectifs que des collections qui y sont accueillies.

Concernant les principes FAIR, *Scriptorium* n'y répond que de manière très partielle puisqu'il n'utilise pas d'identifiants pérennes, ne précise pas la licence et utilise un format standard modifié. Du côté de l'interopérabilité, il ex-

28 *INVENIO*, <<https://inveniosoftware.org/>>, consulté le 26.06.2023.

29 Le code source de ce SIGB est disponible sur Github, <<https://github.com/rero/rero-ils>>, consulté le 26.06.2023.

30 Les notices d'autorité de *Patrinum* décrivent notamment des personnes, des collectivités et des lieux.

31 Tous les chiffres mentionnés dans ce paragraphe sont de juillet 2022.

32 Les conditions d'utilisation sont détaillées sur le site de *Patrinum*, <<https://patrinum.ch/pages/?ln=fr&page=ConditionsUtilisation>>, consulté le 26.06.2023.

iste des possibilités pour partager les données : APIs dédiées, APIs IIIF, protocole OAI-PMH. Par contre, ces outils sont actuellement inactifs ou non documentés.

Du côté de Patrinum, les principes FAIR sont plus proches d'être atteints : la plateforme utilise des liens pérennes et précise la licence d'utilisation. Le format est basé sur un standard, mais qui est également transformé pour une grande partie des collections. Concernant l'interopérabilité, une publication OAI-PMH a été mise en place pour les métadonnées de la partie bibliothèques, mais cela n'est pas documenté publiquement. Quant aux APIs, un développement est en cours pour implémenter IIIF.

Patrinum doit également relever un autre défi qui se situe dans le nombre restreint de documents disponibles sous licence libre et sur le chiffre très important de notices non visibles pour le public, proche de 160'000.³³ Cet état de fait est assez symptomatique des bibliothèques dont le travail de conservation des fonds, de traitement, de numérisation et de catalogage est nécessairement préalable à toute diffusion des données.

Enfin, beaucoup de documents numérisés sur Patrinum étant des manuscrits, des partitions, des images ainsi que des pistes sonores ou audiovisuelles, la recherche plein texte est peu exploitable. À l'inverse, Scriptorium met déjà à disposition une très grande quantité de documents dont le texte est rechercher et analysable, ce qui en fait une source d'informations qui pourrait être facilement exploitable dans le cadre du data mining. Un projet a d'ailleurs été initié pour renégocier les conventions afin de permettre la mutualisation des contenus (par exemple le versement sur un autre méta-catalogue) et l'usage des documents numériques (PDF et données OCR associées) en tant que corpus pour la recherche.

3 Le cas des Archives musicales

Fondées en 1971, les Archives musicales de la BCUL regroupent environ 140 fonds qui illustrent la création musicale et l'activité musicographique dans le canton de Vaud et la Suisse romande du milieu du 19^e siècle à nos jours. Environ 70 inventaires papier ont été numérisés complètement tandis que quelques 2200 partitions et 1500 photographies sont en cours de catalogage sur Patrinum.³⁴ Les documents numérisés sont chargés au format TIFF pour la sauvegarde à long terme et des fichiers JPEG et PDF sont générés automatiquement

33 Chiffres de juillet 2022.

34 Chiffres de juillet 2022.

pour les lecteurs. Concernant la licence, seulement 30% des documents numériques sont libres de droits.³⁵ Pour les autres documents, une convention est négociée par la BCUL, mais cette dernière n'inclut que très rarement le droit de diffusion. Les usagers de Patrinum peuvent consulter un peu moins de 600 documents numériques dans le domaine public issus de ces fonds et il s'agit exclusivement de partitions.³⁶

Outre les partitions numérisées, les Archives musicales conservent aussi des partitions récentes nées numériques. Ces dernières sont souvent versées à la BCUL dans un format propriétaire. Comme il s'agit de l'original, ces fichiers sont sauvegardés sur Patrinum tels quels. Toutefois, des copies transformées aux formats PDF et MusicXML³⁷ sont également ajoutées afin de permettre la visualisation des documents ainsi que leur réutilisation.

3.1 Normes et technologies

Du côté des formats d'encodage et de stockage relatifs aux partitions, on peut d'abord mentionner le standard MEI (Music Encoding Initiative).³⁸ Il s'agit d'un système d'encodage qui permet de rendre la structure des partitions lisible par la machine.³⁹ Ce format s'applique autant pour les métadonnées que pour l'encodage des fichiers eux-mêmes. MEI est open source et repose sur le langage XML. Ce format a été créé dans le but de devenir un standard international. Concernant le stockage des partitions numériques, c'est toutefois le format MusicXML, un format ouvert destiné spécifiquement à l'échange de fichiers qui est le plus pérenne, car MEI peine à s'imposer au niveau international.⁴⁰ MusicXML a donc l'avantage de garantir la compatibilité du fichier stocké avec la plupart des logiciels d'édition et de lecture de partitions.

Pour les partitions numérisées, la reconnaissance OCR est remplacée par la reconnaissance OMR (Optical Music Recognition). Il existe différentes conventions d'écriture musicale, et la majorité des systèmes OMR sont conçus pour identifier la notation la plus commune CWMN (Common Western Music Notation), tandis que quelques travaux spécifiques portent sur les notations his-

35 Chiffre de juillet 2022.

36 Chiffres de juin 2023.

37 « musicXML », *makemusic*, <<https://www.musicxml.com/>>, consulté le 26.06.2023.

38 *MEI*, <<https://music-encoding.org/>>, consulté le 26.06.2023. Le projet est également disponible sur Github, <<https://github.com/music-encoding/>>, consulté le 26.06.2023.

39 « What is MEI? », *MEI*, 2022, part. An introduction to MEI, <<https://music-encoding.org/about/>>, consulté le 26.06.2023.

40 Lee, Lin Hao, *Musical Score Following and Audio Alignment*, Londres 2022, p. 29, <<http://arxiv.org/abs/2205.03247>>, consulté le 26.06.2023.

toriques comme la mensurale⁴¹ ou les Neumes.⁴² Il faut néanmoins souligner que les systèmes OMR sont encore loin d'égaliser l'OCR, une difficulté majeure étant le petit nombre de partitions numérisées.

Enfin, concernant l'accès aux partitions numériques en tant qu'image de haute qualité, les standards ouverts de la communauté IIF apportent une solution efficace, car ils permettent le partage en évitant la duplication de données. Le groupe de travail dédié aux spécifications techniques en lien avec les ressources audio/vidéo a créé en 2016 un ensemble de fonctionnalités pour ces ressources qui sont intégrées dans la version 3.0⁴³ des APIs IIF.⁴⁴ Ces APIs permettent notamment de lire un morceau de musique tout en utilisant la surbrillance pour suivre la partition correspondante.⁴⁵

3.2 Les méta-catalogues de partitions

Quatre des fonds des Archives musicales ont été versés sur le méta-catalogue RISM,⁴⁶ soit environ 1'300 pièces dont 9 sont accompagnées de numérisations.

La plateforme RISM est issue d'un projet international lancé en 1952 qui soutient des groupes de travail dans 36 pays, dont la Suisse. Ce catalogue agrège plus de 1,5 million de notices de partitions dont plus de 180'000 disposent d'une numérisation.⁴⁷ Des métadonnées très détaillées sont disponibles selon les principes LOD et les noms de personnes ainsi que de collectivités sont liés aux autorités GND et VIAF. De plus, il est possible d'effectuer des recherches par incipit en jouant les notes sur un clavier virtuel. Actuellement, les ressources du 19^e siècle sont en cours d'ajout par le groupe de travail suisse. La majorité des fonds BCUL datant du 20^e siècle, les Archives musicales prévoient un nouveau

41 Calvo-Zaragoza, Jorge ; Toselli, Alejandro H ; Vidal, Enrique, « Handwritten Music Recognition for Mensural notation with convolutional recurrent neural networks », *Pattern Recognition Letters* 1982, 2019, pp. 115-121, <<https://doi.org/10.1016/j.patrec.2019.08.021>>, consulté le 26.06.2023.

42 Barton, L.W.G., « The NEUMES Project: digital transcription of medieval chant manuscripts », in : *Second International Conference on Web Delivering of Music (WEDEL-MUSIC)*, Darmstadt 2002, pp. 211-218, <<https://doi.org/10.1109/WDM.2002.1176213>>, consulté le 26.06.2023.

43 IIF, <<https://iif.io/api/presentation/3.0>>, consulté le 26.06.2023.

44 « IIF A/V Technical Specification Group », IIF, <<https://iif.io/community/groups/av/av-tsg/>>, consulté le 26.06.2023.

45 Rabun, Sheila, « IIF Update: Interoperability for Audio & Video Resources », *CLIR*, 2017, <<https://rethinking.clir.org/blog/iif-update-interoperability-for-audio-video-resources/>>, consulté le 26.06.2023.

46 Pour retrouver les collections de la BCUL sur le méta-catalogue RISM, voir : <<https://rism.online/institutions/30000526/sources>>, consulté le 26.06.2023.

47 Chiffres de juin 2023.

versement lorsque la plateforme sera prête à accueillir des documents plus récents.

Outre RISM, il existe d'autres projets de catalogues en lien avec la musique écrite qui sont à différents stades d'achèvement, parmi eux se trouvent Neuma, Musiclibs, et la base de données ELVIS (Electronic Locator of Vertical Interval Successions).

Neuma⁴⁸ est un méta-catalogue qui agrège plusieurs sources dont RISM. La plateforme est issue d'un projet de recherche qui a débuté en 2009 et est actuellement maintenu par l'institut IReMus (Institut de recherche en Musicologie) qui se situe sous la tutelle de la BnF (Bibliothèque Nationale de France), du CNRS (Centre National de la Recherche Scientifique), du ministère de la Culture et de l'Université Paris-Sorbonne.⁴⁹ L'objectif de ce projet est triple, à savoir : donner accès à des partitions rares dans une édition soignée et homogène ; fournir des outils pour la recherche et l'analyse ; permettre d'extraire de larges corpus pour une utilisation externe, que ce soit à des fins d'analyse, de recherche, de diffusion ou d'éducation pédagogique.⁵⁰ Il est possible de télécharger les partitions au format PDF, MEI et MusicXML depuis l'interface. En outre, le lecteur de partitions Verovio⁵¹ est intégré à la plateforme. Toutefois, l'interface n'est actuellement pas aboutie (les fonds sont consultables uniquement via l'arborescence et non via la fonction de recherche). Concernant la réutilisation des données, un service API REST⁵² est disponible.

Musiclibs⁵³ est un projet de méta-catalogue de partitions lancé par SIMSSA (Single Interface for Music Score Searching and Analysis)⁵⁴ en 2017. SIMSSA travaille activement au développement d'une interface web permettant non seulement d'appliquer une reconnaissance OMR aux partitions scannées, mais aussi d'effectuer des analyses et des recherches sur ces dernières.⁵⁵ Musiclibs est un projet intéressant au niveau des choix technologiques, bien qu'il ne soit pas abouti. Ce projet consiste à récupérer les manifestes IIIF de partitions,

48 *Neuma*, <<http://neuma.huma-num.fr>>, consulté le 26.06.2023.

49 « Contact », *Neuma*, <<http://neuma.huma-num.fr/home/contact>>, consulté le 26.06.2023.

50 « Accueil », *Neuma*, <<http://neuma.huma-num.fr/>>, consulté le 26.06.2023.

51 Verovio, développé par RISM, est spécifiquement adapté aux fichiers MEI, voir : <<https://www.verovio.org/index.xhtml>>, consulté le 28.06.2023.

52 « Neuma REST API », *Neuma*, <<http://neuma.huma-num.fr/rest/swagger>>, consulté le 26.06.2023.

53 « About », *Musiclibs*, <<https://www.musiclibs.net/about/>>, consulté le 26.06.2023.

54 SIMSSA est un ensemble de projets de recherche dirigés par Ichiro Fujinaga basé à la Schulich School of Music de la McGill University (Montreal, Canada).

55 « About the SIMSSA project », *Simssa*, <<https://simssa.ca/about/>>, consulté le 26.06.2023.

puis à appliquer une reconnaissance OMR grâce à l'outil web Rodan⁵⁶ avec pour objectif de créer un fichier MEI permettant l'analyse et la recherche dans le contenu de la partition. Enfin, l'interface Musiclibs devrait à terme mettre à disposition le document scanné accompagné du fichier MEI,⁵⁷ le but étant d'en faire une interface de recherche *full-music*.⁵⁸

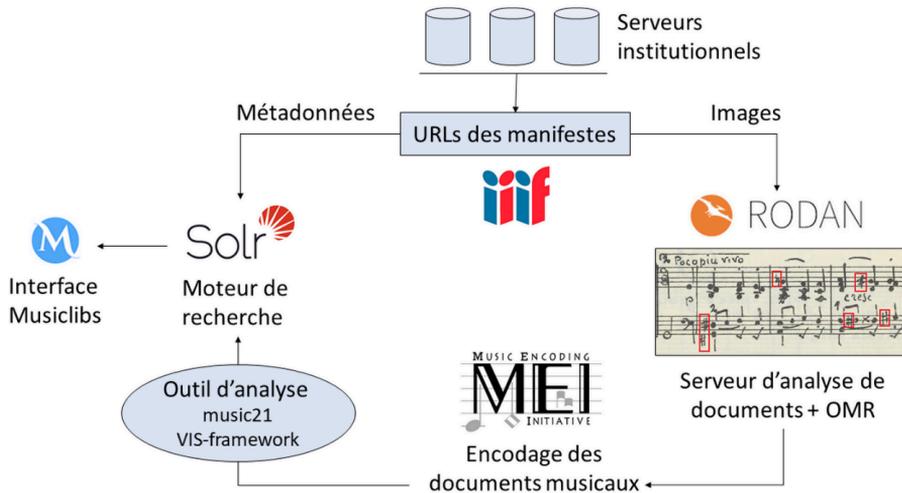


Figure 1. Musiclibs, processus de reconnaissance et d'indexation. Andrew Hankinson, *Optical Music Recognition*, Oxford, 2017, p. 52 [image modifiée et traduite].

La base de données ELVIS⁵⁹ a été lancée en janvier 2012 et est maintenue par SIMSSA.⁶⁰ Contrairement à Musiclibs ou Neuma, ce projet est abouti. La base réunit actuellement près de 3'000 pièces dans un format lisible par la machine.⁶¹ Après s'être créé un compte, les utilisateurs peuvent facilement télécharger et ajouter des partitions sur la plateforme. Les fichiers peuvent être téléchargés par lots et sont en général disponibles au format MusicXML et PDF, ainsi que

56 Le projet est disponible sur Github, <<https://github.com/ddmal/rodan>>, consulté le 26.06.2023.

57 Hankinson, Andrew, *Optical Music Recognition*, Oxford 2017, p. 52, <https://tm.web.ox.ac.uk/sites/default/files/tm/documents/media/hankinson_omr.pdf>, consulté le 26.06.2023.

58 La recherche *full-music* (en référence à la recherche plein texte) devrait permettre de renseigner des notes à l'aide d'un clavier virtuel et de les rechercher dans la partition. Il s'agit donc du même principe que la recherche par incipit, mais appliquée à la partition complète.

59 *ELVIS Database*, <<https://database.elvisproject.ca/>>, consulté le 26.06.2023.

60 « About the Elvis Database », *ELVIS Database*, <<https://database.elvisproject.ca/about/>>, consulté le 26.06.2023.

61 Chiffre de juin 2023.

parfois dans d'autres formats comme MEI ou MIDI. En outre, une application web⁶² permet de téléverser, rechercher et télécharger les ressources d'ELVIS. Pour cette plateforme, les éléments qui nuisent aux principes FAIR sont l'absence de la licence dans les métadonnées et l'usage d'un format non standardisé pour ces dernières.

On peut par contre noter qu'il est annoncé qu'ELVIS évolue en SIMSSA database. Ce futur projet devrait, comme Musiclibs, permettre une recherche *full-music* ainsi qu'offrir des fonctionnalités d'extraction pour les outils MIR⁶³ (Music Information Retrieval). De plus, les données devraient être mises à disposition au format LOD.⁶⁴ SIMSSA database promet donc une bonne compatibilité avec les principes FAIR et la réutilisation des données en général.

Parmi ces quatre catalogues, RISM se distingue par ses métadonnées très soignées, leur mise à disposition selon les principes du LOD et son objectif de réunir toutes les sources au niveau mondial. C'est avant tout un catalogue de métadonnées dont environ 8% des pièces sont rattachées à un document numérique.⁶⁵ Neuma et ELVIS ont en commun de permettre le téléchargement des documents dans des formats exploitables, tels que MEI ou MusicXML, mais ne proposent pas de version scannée (les PDF étant générés à partir des fichiers MusicXML ou MEI), à l'inverse de RISM. En outre, si Neuma et ELVIS permettent la réutilisation de leurs documents numériques à des fins de recherche et d'analyses, leurs métadonnées sont peu détaillées et ne respectent pas un format standard.

Hormis ces trois catalogues, il convient encore de mentionner IMSLP⁶⁶ (International Music Score Library Project) et MuseScore⁶⁷ qui sont devenues deux sources incontournables non seulement pour les musiciens, mais aussi pour les projets de recherche en humanités numériques, en raison de la taille de leur catalogue, qui surpasse largement Neuma, ELVIS et même RISM, si l'on s'attache uniquement aux partitions numérisées.

62 Le projet est disponible sur la Github, <<https://github.com/ELVIS-Project/elvis-database>>, consulté le 26.06.2023.

63 Pour plus de détails sur les outils MIR, voir : Assayag, Gérard, « Du signal temps réel à l'intelligence digitale », *Recherche musicale & intelligence digitale*, MOOC en ligne, entretien réalisé par Dominique Moulon, Paris 2017, <<https://moocdigital.paris/cours/recherche-musicale-intelligence-digitale/signal-temps-reel-lintelligence-digitale>>, consulté le 26.06.2023.

64 Ju, Yaolong ; McKay, Cory ; Hopkins, Emily et al., *SIMSSA DB. Symbolic Music Discovery and Search*, 2019, p. 4, <<https://www.semanticscholar.org/paper/SIMSSA-DB-Ju-McKay/b1cc3cadabe60c622bf3afe008ccfd505c74ce4d>>, consulté le 26.06.2023.

65 Chiffre de juin 2023.

66 *IMSLP*, <<https://imslp.org>>, consulté le 26.06.2023.

67 *musescore*, <<https://musescore.com/sheetmusic>>, consulté le 26.06.2023.

IMSLP, également connu sous le nom de Bibliothèque musicale Petrucci, est un projet lancé en 2006 par Edward W. Guo et actuellement maintenu par la société à but non lucratif Project Petrucci LLC (Wilmington, États-Unis).⁶⁸ Ce catalogue met à disposition près de 716'000 partitions et 79'000 enregistrements audio sous licence libre qui couvrent toutes les périodes jusqu'à nos jours, dont majoritairement celle qui s'étend du 5^e au 18^e siècle.⁶⁹ Par contre, seule une partie des partitions peut être téléchargée au format MusicXML.⁷⁰ Enfin, une API est disponible pour accéder aux pièces et aux compositeurs.⁷¹

MuseScore est avant tout un logiciel libre pour l'édition de partitions.⁷² Le catalogue de MuseScore était donc initialement limité aux utilisateurs du logiciel afin de leur permettre de partager leurs travaux.⁷³ Aujourd'hui, ce catalogue en ligne compte plus de 1,6 millions de partitions dont l'immense majorité est chargée par les utilisateurs. Toutes ces partitions peuvent être consultées en ligne, mais seulement les quelques 360'000 pièces dans le domaine public peuvent être téléchargées dans les formats MuseScore, MusicXML, MIDI et PDF (la création d'un compte gratuit requise).⁷⁴ Le logiciel MuseScore inclut une fonctionnalité web pour la reconnaissance OMR basée sur Audiveris,⁷⁵ ce qui permet aux utilisateurs d'importer des partitions scannées.

En 2017, MuseScore et IMSLP ont lancé l'initiative OpenScore.⁷⁶ Cette dernière visait à créer des fichiers MuseScore et MusicXML sous licence CC0⁷⁷ à partir de partitions dans le domaine public issues de IMSLP ; ceci en s'appuyant sur la communauté des deux plateformes pour améliorer les transcriptions automatiques.⁷⁸ Il s'agit d'un projet toujours en cours, les partitions étant

68 « About », *IMSLP*, <<https://imslp.org/wiki/IMSLP:About>>, consulté le 26.06.2023.

69 Chiffres de juin 2023.

70 Il n'est pas possible de faire une recherche par type de fichier, le nombre de partitions disponibles au format MusicXML n'est donc pas connu.

71 IMSLP met à disposition une API dont l'accès est simplifié grâce à un projet indépendant accessible sur Github, <<https://github.com/joseflevanton/imslp-api>>, consulté le 26.06.2023.

72 Le projet est accessible sur Github : <<https://github.com/musescore/MuseScore/>>, consulté le 27.06.2023.

73 « MuseScore », *Wikipedia*, 2022, <<https://en.wikipedia.org/w/index.php?title=MuseScore&oldid=1100370114>>, consulté le 27.06.2023.

74 Chiffres de juin 2023.

75 Audiveris est un logiciel open source de reconnaissance OMR qui a l'avantage de proposer une interface permettant à l'utilisateur de corriger directement les erreurs de détection. Le projet est accessible sur Github, <<https://github.com/Audiveris/audiveris>>, consulté le 26.06.2023.

76 « OpenScore », *musescore*, <<https://musescore.com/openscore>>, consulté le 26.06.2023.

77 Pour plus de détail sur cette licence, voir : <<https://creativecommons.org/share-your-work/public-domain/cc0>>, consulté le 26.06.2023.

78 « Liberating Sheet Music », *OpenScore*, <<https://openscore.cc>>, consulté le 26.06.2023.

disponibles sur le site de MuseScore. En plus du projet principal, deux groupes de transcription annexes se sont créés : OpenScore Lieder Corpus⁷⁹ (chansons françaises, allemandes et anglaises) et OpenScore Braille⁸⁰ (conversion de partitions MusicXML en braille).

OpenScore, en mettant en place une solution de crowdsourcing, a permis de mettre en ligne plus de 1'000 partitions transcrites.⁸¹ On peut néanmoins remarquer que, même dans ce cadre, le recrutement d'une équipe rémunérée a été nécessaire afin de faire avancer le projet.⁸²

3.3 Mise en perspective

Que ce soit du côté canadien, avec le projet SIMSSA, ou du côté français, avec la plateforme Neuma, on constate une volonté forte de rendre librement accessibles des partitions dans des formats exploitables pour la recherche. Toutefois, aucune solution fonctionnelle à grande échelle n'est en place actuellement. Musiclubs et Neuma ne sont pas des projets aboutis et ils ne semblent pas évoluer. Pour ces plateformes, un problème vient probablement du fait que la conversion des documents numérisés au format MEI est assumée du côté de l'équipe qui maintient le projet et non par les contributeurs. Or, les fonds conservés dans les institutions patrimoniales sont de nature très variée : partitions manuscrites et imprimées, différences dans les styles d'écriture, systèmes de notation musicale divers. Chaque fonds de partitions aurait donc besoin d'un système de reconnaissance OMR optimisé selon ses spécificités. Les projets qui réussissent à aboutir sont donc plus ceux reposant sur le travail des utilisateurs comme ELVIS, IMSLP, MuseScore et leur projet commun OpenScore. Par contre, les métadonnées de ces projets ne respectent pas de standard, ce qui semble normal puisque le format doit être adapté à un public non professionnel pour le catalogue.

Du côté des institutions, on trouve des projets permettant d'accéder à des fichiers MEI ou MusicXML, à l'instar de la base Gesualdo Online,⁸³ un projet mené par l'Université de Tours avec la participation de la BnF,⁸⁴ qui propose un catalogue en ligne des œuvres complètes du compositeur Carlo Gesualdo (plus

79 « OpenScore Lieder », *musescore*, <<https://musescore.com/OpenScore-Lieder-Corpus>>, consulté le 26.06.2023.

80 « OpenScore Braille », *musescore*, <<https://musescore.com/openscore-braille>>, consulté le 26.06.2023.

81 « 1000 scores in the Lieder Corpus! », *OpenScore. Kickstarter*, 2021, <<https://www.kickstarter.com/projects/openscore/openscore-join-the-sheet-music-revolution>>, consulté le 23.06.2023.

82 « 1000 scores in the Lieder Corpus! », part. Commentaires.

de 200 pièces) au format MEI, Sibelius et PDF. Les fichiers de cette base ne semblent par contre pas être accessibles depuis un méta-catalogue, ce qui est symptomatique de ce type de projet, où un grand travail de fond est fourni, mais sans être mis à disposition sur une plateforme qui multiplierait ses chances d'utilisation.

3.4 Projets de recherche et exemples d'applications

Le chapitre trois a jusqu'ici abordé la question de la mise à disposition des partitions, non seulement en tant qu'image numérisée, mais surtout en tant que fichier exploitable par la machine. Or, il est légitime de se demander s'il est véritablement nécessaire de mettre à disposition des fichiers exploitables et quels pourraient en être les usages. Afin de répondre à cette question, ce sous-chapitre présente deux projets qui mettent en relief l'utilité des corpus de partitions ainsi que la plus-value qu'ils peuvent apporter aux professionnels de la musique.

Distant listening est un projet du DCML (Digital and Cognitive Musicology Lab) de l'EPFL financé par le FNS (Fonds National Suisse), qui a pour objectif l'étude du développement de l'harmonie entre le 18^e et le 20^e siècle.⁸⁵ Actuellement en cours (2019-2023), ce projet a pour but la modélisation des structures harmoniques de la tonalité courante, de la tonalité étendue (19^e siècle) et des nouvelles formes harmoniques utilisées dans les musiques pop, rock et jazz.⁸⁶ Cette recherche requiert dans un premier temps la création de corpus musicaux annotés ainsi que le développement d'un modèle d'analyse harmonique.⁸⁷ Pour atteindre son objectif, ce projet combine des hypothèses fondées sur la théorie musicale avec des mesures utilisant la théorie de l'information (qui permet de quantifier le contenu moyen en information des corpus) et des structures théoriques musicales inférées (ou déduites).⁸⁸ Parmi les billets de blog concernant les résultats de ces analyses, on peut mentionner « Tracing historical changes in the exploration of tonal space », qui, via l'application de méthodes

83 *Gesualdo Online*, <<https://ricercar.gesualdo-online.cesr.univ-tours.fr/the-project#>>, consulté le 23.06.2023.

84 Voir la page de la BnF dédiée : « Partitions numériques de la Renaissance », *BnF. API et jeux de données*, <https://api.bnf.fr/fr/partitions-numeriques-de-la-renais-sance#scroll-nav_5>, consulté le 23.06.2023.

85 « Distant Listening », *EPFL*, <<https://www.epfl.ch/labs/dcml/projects/distant-listening/>>, consulté le 26.06.2023.

86 « Distant Listening ».

87 « Distant Listening ».

88 « Distant Listening ».

statistiques et d'apprentissage automatique appliquées à un corpus de 2'000 pièces (75 compositeurs), a permis de faire ressortir l'usage prédominant des quintes ainsi qu'une exploration des gammes toujours plus larges sur cet axe au fil du temps.⁸⁹

Dans un contexte plus large, ce type d'analyse permet de mettre en évidence les modèles théoriques implicites utilisés par les compositeurs, soit la manière dont ils combinent les notes et selon quels principes généraux.⁹⁰ Appliqué à des corpus importants, cela offre une approche pour étudier à grande échelle les changements stylistiques.⁹¹ Quant aux usages pratiques, on peut par exemple mentionner les logiciels d'aide à la composition.⁹²

Dans une optique très différente, le MACLab⁹³ (Music and Audio Computing Lab) de la Graduate School of Culture Technology de Kaist (Corée du Sud) a réalisé en 2017 le projet PerformScore,⁹⁴ soit une interface web permettant de suivre une partition en fonction d'un enregistrement audio. Il est possible de choisir entre plusieurs enregistrements, chacun interprété par un pianiste différent, pour une même partition. Cette interface permet ainsi de comparer le jeu des musiciens en permutant entre plusieurs pistes audio.⁹⁵

L'approche AMT (Automatic Music Transcription) est utilisée dans ce projet pour permettre un alignement automatisé entre la partition et les enregistrements. Cette approche permet de transformer les enregistrements audio en fichiers MIDI⁹⁶ qui sont ensuite alignés avec les partitions, également converties en fichiers MIDI grâce à la méthode DTW⁹⁷ (Dynamic Time Warping).

-
- 89 Moss, Fabian C., « Tracing historical changes in the exploration of tonal space », *EPFL*, 2020, <<https://www.epfl.ch/labs/dcml/tracing-historical-changes/>>, consulté le 26.06.2023.
- 90 Moss, « Tracing historical changes in the exploration of tonal space ».
- 91 Moss, « Tracing historical changes in the exploration of tonal space ».
- 92 Assayag, Gérard, « Évolution musicale : entre art et technologies », MOOC en ligne, *Recherche musicale & intelligence digitale*, entretien réalisé par Dominique Moulon, Paris 2017, <<https://moocdigital.paris/cours/recherche-musicale-intelligence-digitale/evolution-musicale-entre-art-technologies>>, consulté le 26.06.2023.
- 93 *Music and Audio Computing Lab*, <<https://mac.kaist.ac.kr/>>, consulté le 26.06.2023.
- 94 *PerformScore*, <<https://jdasam.github.io/PerformScore>>, consulté le 26.06.2023.
- 95 « Score and Performance Alignment », *Music and Audio Computing Lab*, <https://mac.kaist.ac.kr/score_performance_alignment.html>, consulté le 26.06.2023.
- 96 Pour plus de détails voir : Lee, « Musical Score Following and Audio Alignment », p. 27.
- 97 Pour plus de détails sur cette méthode voir : Kwon, Taegyun ; Jeong, Dasaem ; Nam, Juhan, *Audio-to-Score Alignment Of Piano Music Using RNN-based Automatic Music Transcription*, Daejeon 2017, p. 1, <<https://mac.kaist.ac.kr/pubs/KwonJeongNam-smc2017.pdf>>, consulté le 27.06.2023.

Ce projet s'inscrit plus largement dans le domaine du suivi de partition (score following) qui s'est beaucoup développé dès le milieu des années 2'000⁹⁸ et peut être exploité dans divers outils destinés aux musiciens, de la simple application pour tourner automatiquement les pages,⁹⁹ à des logiciels permettant de générer un accompagnement complexe.¹⁰⁰ Concernant cette deuxième catégorie d'outils, le développement du deep learning a rendu possible l'arrivée d'applications capables de co-improviser avec les artistes en temps réel, sur la base d'une musique jouée en live.¹⁰¹

Pour les bibliothèques et archives, un usage intéressant pour valoriser leurs fonds serait la mise en relation d'enregistrements audio ou vidéo avec des partitions. Ce type de projet serait facilité s'il existait des bases FAIR et ouvertes de partitions MEI ou MusicXML d'un côté, et des bases de documents audio ou vidéo respectant les mêmes principes de l'autre. Ainsi, on pourrait imaginer la mise en place de plateformes capables de puiser dans ces bases pour mettre en relation diverses partitions et interprétations dans une interface similaire à PerformScore.

4 Conclusion

L'état des collections patrimoniales numériques de la BCUL et des plateformes utilisées pour les diffuser a permis de faire ressortir la diversité des fonds ainsi que la pluralité des catalogues et méta-catalogues adoptés par l'institution. Cette analyse a également mis en exergue plusieurs difficultés concernant la publication des ressources en ligne, à savoir, le nombre restreint de documents sous licence libre et la quantité importante de notices non diffusées pour cause de qualité insuffisante. Ces problématiques constituent un premier blocage pour l'exploitation des ressources patrimoniales. Un second blocage tient à la publication des données dans des systèmes institutionnels qui ne sont pas suffisam-

98 Assayag, « Du signal temps réel à l'intelligence digitale ».

99 Lee, « Musical Score Following and Audio Alignment », pp. 19-20.

100 L'application Antescofo, développée en France par l'Ircam (Institut de Recherche et de Coordination Acoustique/Musique), se base sur une partition combinée à une machine d'écoute pour suivre une performance musicale et jouer un accompagnement préprogrammé tout en s'adaptant au jeu musical tels que les ralentissements ou les accélérations. Voir : *Ircam*, <<https://forum.ircam.fr/projects/detail/antescofo>>, consulté le 23.06.2023; Une démonstration peut également être visualisée à l'adresse <<http://repmus.ircam.fr/antescofo>>, consulté le 23.06.2023.

101 L'application SoMax2, également conçue au sein de l'Ircam est capable d'analyser en temps réel une performance et de concevoir un accompagnement complexe. Voir : « Somax2 », *Sciences et Technologies de la Musique et du Son*, <<https://www.stms-lab.fr/projects/pages/somax2/>>, consulté le 23.06.2023.

ment interopérables. Toutefois, il existe des solutions pour parer à ce problème comme la publication des images sous forme de manifeste IIIF, le rapprochement des principes FAIR et la participation à des méta-catalogues.

Dans le domaine de la musique écrite, la transformation de partitions scannées en fichiers exploitables par la machine requiert un écosystème spécifique qui diffère des ressources textuelles et soulève des défis technologiques tels que la reconnaissance OMR. Pour cette raison, il n'existe pas, à l'heure actuelle, un seul méta-catalogue pour les partitions qui soit FAIR et qui garantisse à la fois une bonne qualité de métadonnées, l'accès à des partitions numérisées en PDF et l'accès à des fichiers MEI ou MusicXML. Par contre, la plateforme RISM répond à la première partie de ces éléments, tandis que d'autres méta-catalogues, comme IMSLP, MuseScore, OpenScore ainsi qu'ELVIS, permettent d'accéder à des fichiers exploitables. L'idéal serait que ces bases soient LOD compatibles, pour que leurs données puissent être croisées avec celles de RISM, ce qui n'est pas le cas. Néanmoins, la situation pourrait s'améliorer si le catalogue SIMSSA database est mis en place.

Du côté applicatif, l'accès à des corpus exploitables permet de faire avancer la recherche par l'entremise d'études telles que l'évolution musicale à travers les époques. Le suivi de partition est un autre domaine en constante amélioration qui peut être exploité à travers des applications concrètes comme l'accompagnement automatisé. Plus simplement, ce domaine apporte également une aide à l'apprentissage et à l'analyse grâce à des plateformes permettant de suivre une partition en écoutant différentes interprétations professionnelles (et non des sons de synthèse MIDI).

Enfin, le deep learning appliqué aux corpus de partitions ainsi qu'aux fichiers audio et vidéo ouvre encore de nombreuses possibilités. Toutefois, le manque de partitions exploitables est encore un frein pour les développements utilisant cette technologie.