



*BCUL, Service des Archives musicales, Suisse*

# Le patrimoine numérique de la BCU Lausanne :

## EXPLOITATION ET REUTILISATION DES DONNEES

Noémie Payot

Direction : Pablo Iriarte, coordinateur du pôle informatique documentaire à la division de l'information scientifique de l'Université de Genève

Travail de Master : MAS ALIS 2020-2022

Légende de l'image de couverture : Juon, Paul, éd. Kakteen: 7 Klavierstücke, op. 76/1-7. Berlin: Schlesinger, 1923, Bibliothèque cantonale et universitaire - Lausanne, Service des Archives musicales, Fonds Paul Juon, FPJ 64, <https://patrinum.ch/record/268962>.

### Remerciements

*Je tiens à exprimer toute ma reconnaissance à Pablo Iriarte, directeur de ce travail, pour sa disponibilité, son aide et ses judicieux conseils, qui ont largement contribué à alimenter ma réflexion.*

*Je désire également sincèrement remercier les responsables et professionnels de la BCUL Christophe Bezançon, Christian Bürki, Elisa Dell'Ambrogio, Sophie Donche Gay, Ramona Fritschi, Eloïse Haling, Michaël Hertig, Myriam Jouhar, Verena Monier, Théophile Naïto, Floriane Pochon Levit et Marc Tiefenhauer, pour leur temps, leur intérêt et les précieuses informations qu'ils m'ont fournies.*

*Je remercie aussi en particulier Alexis Rivier, conservateur à la Bibliothèque de Genève, ainsi que Pascal Repond et Nicolas Prongué, respectivement expert et Directeur adjoint à RERO+, pour leurs réponses rapides et efficaces à mes nombreuses questions.*

*Pour terminer, un grand merci à Lauriane Grosjean, Camille Noverraz, Sylvie Payot et Yannick Payot pour leur relecture, leurs corrections et leur soutien inestimable.*

### Déclaration d'authenticité

« Je déclare par la présente avoir rédigé ce travail personnellement et n'avoir pas utilisé d'autres sources que celles indiquées. J'ai signalé en tant que tels tous les passages repris de sources, de façon littérale ou du point de vue du sens. Je sais que, dans le cas contraire, le Sénat de l'Université de Berne et la Direction de l'Université de Lausanne sont légitimés à retirer le titre décerné sur la base de ce travail. »

Le 15 août 2022



## Résumé

Ce travail propose une réflexion sur les améliorations possibles concernant la mise à disposition des collections patrimoniales numériques de la Bibliothèque cantonale et universitaire – Lausanne (BCUL), afin de permettre la réutilisation de ces données dans le cadre de projets ayant recours aux humanités numériques. Une introduction contextuelle présente les notions clés de ce domaine, telles que l’open data, les principes FAIR, les linked open data ainsi que les humanités numériques. Un état des collections patrimoniales numériques de l’institution est ensuite réalisé, en utilisant les principes FAIR comme indicateurs de l’adéquation des plateformes numériques internes et externes par rapport à la réutilisation des données. Une comparaison avec la Bibliothèque de Genève est effectuée à la fin de ce chapitre, afin de donner une vision plus générale de la situation. La troisième partie de ce travail se focalise sur le cas des collections de partitions musicales et offre une analyse détaillée des technologies et outils exploitables dans ce domaine. Enfin, la partie quatre opère un retour au cadre général des collections patrimoniales numériques de la BCUL et propose différentes solutions qui permettraient de faciliter l’exploitation des données.

## Table des matières

1	Introduction .....	7
1.1	Contexte.....	7
1.1.1	L'Open data et l'open science.....	7
1.1.2	Restrictions juridiques pour les bibliothèques .....	9
1.1.3	Les Principes FAIR.....	9
1.1.4	Linked Open Data.....	10
1.1.5	Les Humanités numériques.....	12
1.1.6	La visualisation de l'information .....	12
1.2	Problématique .....	13
1.3	Objectifs .....	15
1.4	Méthodologie utilisée et regard critique.....	15
2	Etat des bibliothèques numériques BCUL.....	17
2.1	Les plateformes BCUL .....	17
2.1.1	Scriptorium.....	17
2.1.2	Patrinum.....	18
2.1.3	Mise en perspective des deux plateformes .....	18
2.2	Etat des fonds et plateformes externes.....	21
2.2.1	Les Archives musicales .....	21
2.2.2	L'Iconopôle.....	23
2.2.3	Les Manuscrits et archives .....	24
2.2.4	Le Dépôt légal numérique.....	25
2.2.5	La Documentation vaudoise .....	26
2.2.6	La Réserve précieuse.....	27
2.2.7	La Phonothèque .....	28
2.2.8	Synthèse relative à l'état des fonds et plateformes externes .....	28
2.3	État des plateformes relatives à la BGE .....	29
2.3.1	Plateforme des manuscrits et archives.....	30
2.3.2	Plateforme iconographique .....	31
2.3.3	Plateformes externes.....	31
2.4	Plateformes en liens avec les réseaux suisses de bibliothèques .....	32
2.4.1	Swisscollections.....	32
2.4.2	Plateformes en lien avec RERO+ .....	33

2.5	Mise en perspective entre la BCUL et la BGE .....	34
3	Le domaine de la musique écrite .....	36
3.1	Domaine choisi .....	36
3.2	L'IA dans le domaine musical .....	36
3.3	Normes et technologiques .....	39
3.3.1	MEI.....	39
3.3.2	MIDI .....	40
3.3.3	MusicXML .....	40
3.3.4	Incipits musicaux .....	41
3.3.5	APIs IIIF audio et vidéo .....	42
3.3.6	OMR.....	43
3.3.7	Le suivi de partition .....	45
3.4	Projets innovateurs .....	47
3.4.1	PerformScore par le MACLab .....	47
3.4.2	Prototype IIIF-AV-player par le DDMAL.....	48
3.4.3	Musiclibs.....	49
3.4.4	Distant listening.....	51
3.4.5	Le projet Neuma .....	52
3.4.6	ELVIS project.....	53
3.5	Méta-catalogues de partitions .....	54
3.6	Mise en perspective avec les Archives musicales de la BCUL .....	56
3.6.1	Normes et formats .....	56
3.6.2	Interopérabilité et méta-catalogues .....	58
3.6.3	Projet de suivi de partition ? .....	58
4	Comment améliorer la mise à disposition ?.....	59
4.1	Réutilisation depuis les plateformes internes.....	59
4.2	Les méta-catalogues Suisses .....	60
4.2.1	Le projet Connectome .....	60
4.3	Les méta-catalogues internationaux .....	63
4.3.1	Gallica .....	64
4.3.2	Google Arts and Culture .....	65
4.3.3	Europeana .....	65
4.3.4	Hathi Trust .....	66

4.3.5	Internet Archive .....	67
4.3.6	Wikimedia Commons et Wikisource .....	68
4.4	Synthèse .....	69
5	Conclusion .....	71
5.1	Discussion .....	72
6	Bibliographie .....	74
6.1	Contexte général .....	74
6.2	Domaine musical .....	75
6.3	Intelligence Artificielle .....	78
6.4	Humanités numériques .....	78
6.5	Réutilisation des données .....	80
7	Table des abréviations .....	82
8	Table des figures .....	84
9	Annexes .....	85
9.1	Tableaux récapitulatifs .....	85
9.1.1	Scriptorium, Patrinum et Alma digital .....	85
9.1.2	Tableau récapitulatif des collections patrimoniales numériques BCUL .....	91
9.1.3	Tableau récapitulatif des plateformes externes BCUL et BGE .....	99
9.1.4	Tableau récapitulatif des méta-catalogues internationaux .....	103
9.2	Questions posées aux responsables de collections lors des entretiens .....	105
9.3	Politique des bibliothèques numériques de la BCUL .....	106

# 1 Introduction

## 1.1 Contexte

Ce travail porte sur les collections patrimoniales numériques mises à disposition par la Bibliothèque cantonale et universitaire – Lausanne (BCUL). Cette institution possède en effet des collections patrimoniales importantes et variées, qu'elle rend accessibles via diverses plateformes. La numérisation, la préservation et la mise en valeur des collections patrimoniales font partie des missions de la BCUL<sup>1</sup>, qui poursuit depuis longtemps l'objectif d'être une institution novatrice dans le domaine numérique (premier système de catalogue informatisé de Suisse, première bibliothèque francophone à participer à Google Books).<sup>2</sup>

La notion de mise à disposition comprend aujourd'hui un large éventail de possibilités. Le simple fait de rendre accessibles physiquement les documents constitue déjà une mise à disposition. Le fait de rendre également disponible un inventaire des documents est l'étape suivante. Dans les années 1980, les bibliothèques ont commencé à se munir de catalogues en ligne (ou Online Public Access Catalog : OPAC), soit de catalogues au format numérique (et non papier) accessibles sur le web (et non uniquement sur un poste de consultation à l'intérieur de la bibliothèque)<sup>3</sup>. Dès la fin des années 2000, les OPAC ont fait place aux outils de découverte (ou Discovery Tools) qui se caractérisent par des possibilités de recherche plus développées et l'inclusion dans les résultats de documents externes (notices issues d'autres catalogues, documents numériques, etc.)<sup>4</sup>.

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, les documents numériques, dont le texte est lisible pour la machine<sup>5</sup>, permettent aux utilisateurs d'accéder directement à l'information elle-même en se passant du médium physique.

### 1.1.1 L'Open data et l'open science

L'open data, ou données ouvertes, consiste à mettre à disposition les données de manière libre sur le web. Lorsqu'il s'agit d'institutions publiques, l'open data a pour objectif, d'une part de garantir la transparence des activités, et, d'autre, part de permettre le partage du savoir comme un bien commun.<sup>6</sup>

---

<sup>1</sup> Service de Communication, « La BCUL en 2025 : Misson, vision et visite virtuelle » (Bibliothèque cantonale et universitaire – Lausanne, 2020), 3, <https://www.bcu-lausanne.ch/wp-content/uploads/2022/02/plan-directeur-abrege-2021-25.pdf>.

<sup>2</sup> « Qui sommes-nous ? », BCUL, sect. Historique, consulté le 10 juin 2022, <https://www.bcu-lausanne.ch/qui-sommes-nous/>.

<sup>3</sup> « Online Public Access Catalog », in *Wikipedia*, 4 décembre 2021, [https://en.wikipedia.org/w/index.php?title=Online\\_public\\_access\\_catalog&oldid=1058596163](https://en.wikipedia.org/w/index.php?title=Online_public_access_catalog&oldid=1058596163).

<sup>4</sup> « Discovery System (Bibliographic Search) », in *Wikipedia*, 3 mai 2022, [https://en.wikipedia.org/w/index.php?title=Discovery\\_system\\_\(bibliographic\\_search\)&oldid=1085918226](https://en.wikipedia.org/w/index.php?title=Discovery_system_(bibliographic_search)&oldid=1085918226).

<sup>5</sup> 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.

<sup>6</sup> La rédaction de Futura, « Définition | Open data - Donnée ouverte », Futura Tech, consulté le 22 juillet 2022, <https://www.futura-sciences.com/tech/definitions/internet-open-data-15502/>.

Les données ouvertes doivent respecter des impératifs juridiques, financiers et techniques afin d'être véritablement mises à disposition de manière libre.<sup>7</sup> Dans ce cadre, le terme libre est défini de la manière suivante : « free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness. ».<sup>8</sup> Plus précisément, cela implique : 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.).<sup>9</sup>

L'open science est un concept très vaste qui vise à « une ouverture complète du processus scientifique ».<sup>10</sup> Ce faisant, il touche de nombreux domaines, à savoir : l'open notebooks, l'open data, l'open peer review, l'open access, l'open source, la citizen science, le scientific social networks et enfin l'open educational ressources. Parmi les nombreux bénéfices de l'open science, on peut mentionner : la garantie de « recherche intègre, responsable, transparente, reproductible, valide et de qualité » ;<sup>11</sup> l'impact accru des nouvelles découvertes et la circulation du savoir ; la démocratisation du savoir par la réduction de l'écart entre État, institutions et citoyens ; et enfin le retour au citoyen des recherches financées par les fonds publics.<sup>12</sup> L'open data et l'open science sont deux concepts qui s'englobent de manière mutuelle mais partielle, puisque l'open science n'inclut que les données de la recherche dans le cadre de l'open data.<sup>13</sup>

Les domaines de l'open data et de l'open science sont mentionnés dans le plan directeur de la BCUL de la manière suivante :

« Open Access, Open Science, Open Data, Big Data, intelligence artificielle sont les nouveaux outils de la science, non seulement dans le cadre universitaire, mais de plus en plus pour l'ensemble de notre communauté. Face à ces enjeux multiples, la BCUL doit continuer à avancer avec les moyens qui sont les siens et faire les meilleurs choix possibles – tout n'étant pas une option réaliste. »<sup>14</sup>

Du côté des données patrimoniales, ce même document précise que la BCUL « continuera la numérisation, la préservation et la mise en valeur de ses collections patrimoniales et les mettra à disposition au travers de plateformes numériques ».<sup>15</sup> Ces éléments suivent les principes de l'open data en termes de mise à disposition gratuite et de partage du savoir. Quant à la forme et au format, qui doivent être lisibles par la machine et ouverts, les bibliothèques fournissent un effort de longue date dans ces domaines en conformité avec les principes de l'open data. La question de la licence est par contre plus complexe (1.1.2), car différentes restrictions légales s'appliquent.

---

<sup>7</sup> Ibid.

<sup>8</sup> « Open Definition 2.1 - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge », Open Knowledge Foundation, consulté le 22 juillet 2022, <https://opendefinition.org/od/2.1/>.

<sup>9</sup> Ibid.

<sup>10</sup> « Open Science », SAGW ASSH, consulté le 25 juillet 2022, <https://www.sagw.ch/fr/assh/themes/cultures-scientifiques/open-science>.

<sup>11</sup> Gérard Bagnoud et al., « Université de Lausanne : Stratégie Open Science et plan d'actions 2019-2021 », 17 décembre 2019, 6.

<sup>12</sup> Ibid., 6.

<sup>13</sup> « Open Science », part. Libre accès.

<sup>14</sup> Service de Communication, « Abrégé du Plan directeur 2021-2025 de la BCUL », 2.

<sup>15</sup> Ibid., 3.

## 1.1.2 Restrictions juridiques pour les bibliothèques

Les bibliothèques sont impactées par le droit d'auteur, et de manière plus restreinte, par la protection des données. Ainsi, elles ne sont pas autorisées à mettre à disposition dans leur intégralité des reproductions numériques de documents sous droit disponibles sur le marché<sup>16</sup>. Cette loi ne concerne pas les documents anciens (auteur décédé depuis au moins 70 ans),<sup>17</sup> qui peuvent être diffusés numériquement de manière libre. Toutefois, les ressources patrimoniales collectées par les bibliothèques peuvent également être acquises dans le cadre de conventions passées avec les éditeurs, auteurs, etc., ce qui peut limiter la marge de manœuvre de l'institution, y compris au-delà des 70 ans précédemment mentionnés. Une autre problématique concerne les ressources dont l'auteur ne peut être retrouvé et pour lesquelles on ignore donc le délai post mortem à respecter. Enfin, il est utile de mentionner les licences Creative Commons (CC)<sup>18</sup>, conçues dans le but de permettre aux auteurs de rendre leurs œuvres accessibles en limitant complètement ou partiellement leur droit. Certaines publications modernes peuvent donc être mises à disposition par ce biais.

## 1.1.3 Les Principes FAIR

Les principes FAIR<sup>19</sup> (Findable, Accessible, Interoperable et Reusable), publiés en 2016 dans Scientific Data, ont pour objectif de donner une ligne de conduite à la publication des données de la recherche, et plus largement aux données ouvertes, afin d'optimiser leur réutilisation. Plus précisément, les principes FAIR impliquent :

1. *Findable* : Les (méta)données doivent être lisibles par les usagers et par les machines afin de pouvoir être découvertes automatiquement. Un identifiant pérenne leur est attribué. Elles sont richement décrites (en conformité avec le point 4) et incluent l'identifiant de la ressource qu'elles décrivent. Enfin, elles sont indexées et accessibles via une interface permettant la recherche.<sup>20</sup>
2. *Accessible* : Un identifiant permet d'accéder aux (méta)données via un protocole de communication standardisé, ouvert, gratuit et universellement implémentable. Le protocole permet si nécessaire une procédure d'authentification et d'autorisation. Les métadonnées sont accessibles, même lorsque les données ne sont plus disponibles.<sup>21</sup>
3. *Interoperable* : Les données sont intégrées à d'autres données et interagissent avec des applications ou des flux de travail permettant l'analyse, le stockage et le traitement. Les (méta)données utilisent un langage formel, accessible, partagé et permettant la représentation des connaissances (par exemple RDF/XML, JSON-LD, etc.). Elles utilisent des vocabulaires qui suivent les

---

<sup>16</sup> Voir: RS 231.1 - LDA, Art. 19.

<sup>17</sup> Ce délai s'applique à partir du 1<sup>er</sup> juillet 1993 et ne concerne pas les productions photographiques dépourvues de caractère individuel. Pour ces dernières un délai de 50 ans est appliqué (RS 231.1 – LDA, Art. 29).

<sup>18</sup> Voir : <https://creativecommons.org>

<sup>19</sup> Mascha Jansen, « FAIR Principles », GO FAIR, consulté le 23 juillet 2022, <https://www.go-fair.org/fair-principles/>.

<sup>20</sup> Ibid., part. Findable.

<sup>21</sup> Ibid., part. Accessible.

principes FAIR (par exemple Dublin Core)<sup>22</sup> et elles incluent des liens vers d'autres (méta)données (l'objectif est de créer des liens porteurs de sens entre les ressources).<sup>23</sup>

4. *Reusable* : La description des (méta)données est riche d'une pluralité d'attributs précis et pertinents, dont une licence d'utilisation claire et des informations détaillées concernant la provenance. De plus, elles répondent aux normes et aux standards applicables dans leur domaine.<sup>24</sup>

Les principes FAIR étant généralement associés au domaine de la recherche, ils sont rarement mis en lien avec les bibliothèques. Ainsi, l'Université de Lausanne s'est munie d'une stratégie Open Science,<sup>25</sup> tandis que la BCUL ne les mentionne nulle part. Il s'agit néanmoins de principes qui commencent à être de plus en plus associés au patrimoine culturel,<sup>26</sup> bien que certains auteurs préconisent une adaptation de certains points.<sup>27</sup>

### 1.1.4 Linked Open Data

Une des solutions techniques qui permet de mettre en œuvre les principes FAIR réside dans le concept du linked data (ou données liées, également connu sous les noms de web de données ou 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.<sup>28</sup> Le linked data se focalise sur l'aspect de l'interopérabilité entre les données, en ce sens, il est donc inclus dans le « i » de FAIR.<sup>29</sup> Pour répondre à cet objectif, Berners-Lee propose les solutions suivantes :

1. "Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs. so that they can discover more things."<sup>30</sup>

Les principes linked data appliqués à l'open data sont à l'origine du mouvement Linked Open Data (LOD)<sup>31</sup> dont l'objectif est de publier des données à la fois ouvertes et interopérables.

---

<sup>22</sup> Dublin Core est un vocabulaire utilisé pour décrire les documents de manière simple. Voir : <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>.

<sup>23</sup> « FAIR Principles », part. Interoperable.

<sup>24</sup> Ibid. Reusable.

<sup>25</sup> Bagnoud et al., « Université de Lausanne : Stratégie Open Science et plan d'actions 2019-2021 », 10-11.

<sup>26</sup> 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 (29 octobre 2020), <https://www.rd-alliance.org/publishing-glam-data-fair-data>.

<sup>27</sup> Lukas Koster et Saskia Woutersen-Windhouwer mentionnent notamment, dans leur article « FAIR Principles for Library, Archive and Museum Collections: A proposal for standards for reusable collections », que les principes FAIR sont lacunaires par rapport à la préservation à long terme des objets numériques (*The Code4Lib Journal*, 2018).

<sup>28</sup> Tim Berners-Lee, « Linked Data », W3C, 27 juillet 2006, <https://www.w3.org/DesignIssues/LinkedData.html>.

<sup>29</sup> Karla Avanço, « FAIR Principles and Linked Open Data », Billet, *The Road to FAIR* (blog), part. LOD and FAIR, consulté le 23 juillet 2022, <https://roadtofair.hypotheses.org/288>.

<sup>30</sup> Berners-Lee, « Linked Data ».

<sup>31</sup> Avanço, « FAIR Principles and Linked Open Data ».

Berners-Lee propose également un système de classement 5-Star afin d'évaluer la qualité des données mises en ligne :

- ★ make your stuff available on the Web (whatever format) under an open license
- ★★ make it available as structured data (e.g., Excel instead of image scan of a table)
- ★★★ make it available in a non-proprietary open format (e.g., CSV instead of Excel)
- ★★★★ use URIs to denote things, so that people can point at your stuff
- ★★★★★ link your data to other data to provide context<sup>32</sup>

Ce classement se retrouve partiellement dans les principes FAIR, comme le montre le tableau suivant.<sup>33</sup>

	★	★★	★★★	★★★★	★★★★★
Findable	X			X	X
Accessible	X	X	X	X	
Interoperable	X	X	X	X	X
Reusable	X	X	X	X	X

Une comparaison entre le LOD et les principes FAIR fait néanmoins ressortir quelques différences majeures :

- Des données ouvertes sont requises pour le LOD alors que FAIR implique uniquement de déclarer la licence.<sup>34</sup>
- Les métadonnées jouent un rôle clef pour le FAIR alors que le LOD considère les métadonnées elles-mêmes comme des données interopérables.<sup>35</sup>
- Le LOD utilise les URIs comme identifiants et propose le modèle RDF ainsi que le protocole SPARQL pour accéder aux données. FAIR 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.<sup>36</sup> En effet, SPARQL permet d'interroger des ensembles importants de données et d'en extraire des informations selon des critères précis. La démocratisation de ces technologies facilite la recherche dans le cadre des humanités numériques (ou digital humanities) en permettant d'extraire de la connaissance à partir de lots de données textuelles, iconographiques, et de plus en plus, également de ressources sonores ou vidéos.

<sup>32</sup> Berners-Lee, « Linked Data », part. Is your Linked Open Data 5 Star?

<sup>33</sup> Ali Hasnain et Dietrich Rebholz-Schuhmann, « Assessing FAIR Data Principles against the 5-Star Open Data Principles » (Springer Verlag, 2018), chap. 5.1, [https://doi.org/10.1007/978-3-319-98192-5\\_60](https://doi.org/10.1007/978-3-319-98192-5_60). Le tableau est présenté dans ce travail sous une forme simplifiée.

<sup>34</sup> Avanço, « FAIR Principles and Linked Open Data », part. LOD and FAIR.

<sup>35</sup> Ibid.

<sup>36</sup> Sugimoto Go, « Open Data Empowerment of Digital Humanities by Wikipedia/DBpedia Gamification and Crowd Curation –WiQiZi's Challenges with APIs and SPARQL » (Mexico City, Zenodo, 30 septembre 2019), 1, <https://doi.org/10.5281/zenodo.3465654>.

### 1.1.5 Les Humanités numériques

Les humanités numériques consistent en l'application « des méthodes, des dispositifs et des perspectives heuristiques liés au numérique dans le domaine des Sciences humaines et sociales ». <sup>37</sup> Cette nouvelle discipline est étroitement liée à l'open data ainsi qu'aux principes du FAIR et au LOD, car elle requiert la mise à disposition de (méta)données ouvertes et interopérables. <sup>38</sup> Elle est également favorable « à la diffusion, à la circulation et au libre enrichissement des méthodes, du code, des formats et des résultats de la recherche ». <sup>39</sup> De ce fait, elle promeut l'open source et l'open science.

### 1.1.6 La visualisation de l'information

Les opportunités ouvertes par le domaine des humanités numériques révolutionnent l'accès à l'information. D'une part, elles permettent l'extraction d'informations à partir de lots de données pour la recherche, et d'autre part, elles permettent de représenter ces données d'une manière attractive et compréhensible pour le grand public.

Comme l'explique Frédéric Giuliano, <sup>40</sup> le fait de présenter les données de différentes manières permet aux utilisateurs « de participer au processus d'analyse, de sonder et d'explorer plus avant, de remixer et de jouer avec les données pour parvenir à une nouvelle compréhension ». <sup>41</sup> La présentation des données sous différentes formes, qu'il s'agisse de graphes, de lignes chronologiques, de regroupement via des méthodes de clustering, etc., permet une meilleure appréhension des grandes collections et de leur contenu. <sup>42</sup> Utilisés pour répondre à une question de recherche, ces outils aident à extraire les données pertinentes et à comprendre les liens qui unissent les éléments. A ce sujet, Klaus Keil et Laurent Pugin précisent même que la visualisation de données doit devenir une fonctionnalité systématiquement intégrée aux interfaces de recherche. <sup>43</sup> Le portail visualize.admin, <sup>44</sup> qui permet de visualiser les données ouvertes de l'administration publique suisse sous forme de graphiques, est un exemple récent qui va dans ce sens. De manière générale, on trouve aussi de plus en plus d'outils qui permettent aux chercheurs de charger leur propre corpus de données en proposant différents modes de visualisation. Si les interfaces de recherche ne permettent pas directement de visualiser les données, il est donc utile, voire nécessaire de permettre au chercheur d'extraire les données par lot, notamment l'ensemble des résultats

---

<sup>37</sup> Pierre Mounier, « Manifeste des Digital Humanities », *Journal des anthropologues. Association française des anthropologues*, n° 122-123 (1 décembre 2010): paragr. 3, <https://doi.org/10.4000/jda.3652>.

<sup>38</sup> Ibid.

<sup>39</sup> Ibid.

<sup>40</sup> Frédéric Giuliano, « Humanités numériques et archives : la longue émergence d'un nouveau paradigme », *Documentation et bibliothèques* 65, n° 2 (2019): 37-46, <https://doi.org/10.7202/1063788ar>.

<sup>41</sup> Anne Bahde, « Conceptual Data Visualization in Archival Finding Aids: Preliminary User Responses », *Portal: Libraries and the Academy* 17, n° 3 (2017): 485-506, <https://doi.org/10.1353/pla.2017.0031>, cité par Giuliano dans « Humanités numériques et archives : la longue émergence d'un nouveau paradigme », paragraphe 26.

<sup>42</sup> Giuliano, « Humanités numériques et archives », paragr. 25.

<sup>43</sup> Klaus Keil et Laurent Pugin, « Das Internationale Quellenlexikon der Musik, RISM: Ein Gemeinschaftsprojekt zum Nutzen und als Aufgabe für Forschung und Bibliotheken », *Bibliothek Forschung und Praxis* 42, n° 2 (26 juillet 2018): part. 5, <https://doi.org/10.1515/bfp-2018-0042>.

<sup>44</sup> Voir : <https://visualize.admin.ch/fr>

d'une recherche. Parmi les outils de ce type, on peut mentionner Voyant,<sup>45</sup> adapté aux données textuelles, et Palladio,<sup>46</sup> conçu pour les fichiers délimités (par exemple csv).

Concernant la présentation des données au grand public, les humanités numériques ouvrent également un grand champ de possibilités pour la conception d'applications interactives et ludiques. Il est néanmoins important d'accompagner les données présentées avec des explications qui leur donnent du sens. En effet, un défaut récurrent dans ce type de projet est de laisser le soin à l'utilisateur d'identifier les données intéressantes ainsi que de les interpréter, ce qui n'est évidemment pas pertinent pour la majorité d'entre eux. Parfois, le parti est pris de simplement présenter les données visuellement attractives sans ajouter d'explications, ce qui peut permettre d'attirer le public mais laisse un manque à combler au niveau de l'information. De manière générale, il est plus complexe de s'adresser au grand public, car l'attractivité de l'interface joue alors un rôle fondamental. Or, les institutions culturelles ont souvent trop peu de moyens financiers et moins de compétence dans ce domaine, qui est finalement assez éloigné de leurs mandats initiaux. Dans ce cadre, des partenariats publics privés peuvent donc faciliter la réalisation de tels projets. À cet égard, on peut citer LIFE Tags,<sup>47</sup> un projet de Google Arts & Culture où 4 millions d'images provenant des archives de LIFE magazine ont été organisées en une encyclopédie interactive.<sup>48</sup> Le projet Europe Time Machine,<sup>49</sup> coordonné par l'École Polytechnique Fédérale de Lausanne (EPFL), est un autre exemple. Cet immense projet en cours de réalisation se décline pour différentes villes d'Europe. L'objectif est de retracer leur développement historique au niveau économique, social, culturel et géographique en proposant notamment des reconstitutions spatio-temporelles.<sup>50</sup>

## 1.2 Problématique

Il n'est plus possible pour les bibliothèques et autres institutions GLAM (Galleries, Libraries, Archives, Museums) de se contenter de mettre à disposition uniquement les métadonnées des documents. Il est aussi nécessaire de numériser le document lui-même et de le mettre à disposition selon les principes FAIR afin de permettre une réutilisation dans le cadre des humanités numériques. Les bibliothèques ont conscience de ces impératifs mais doivent faire face à plusieurs problématiques, comme le manque de ressources humaines et financières ainsi que les limitations dues au droit d'auteur et, pour certaines collections, à la protection des données personnelles.

La BCUL s'est munie en janvier 2019 d'une politique des bibliothèques numériques qui précise ses objectifs dans ce domaine :

« Chaque document numérique est sauvegardé dans une unique bibliothèque numérique par souci de bonne gestion des ressources disponibles.

Lorsqu'il existe une bibliothèque numérique externe qui :

---

<sup>45</sup> Voir : <https://voyant.lincspjproject.ca>

<sup>46</sup> Voir : <http://hdlab.stanford.edu/palladio>

<sup>47</sup> Voir : <https://experiments.withgoogle.com/life-tags>

<sup>48</sup> Gaël Hugo, « LIFE Tags », Experiments with Google, mai 2018, <https://experiments.withgoogle.com/life-tags>.

<sup>49</sup> Voir : <https://www.timemachine.eu>

<sup>50</sup> « About Us », Time Machine Europe, consulté le 27 juillet 2022, <https://www.timemachine.eu/about-us/>.

- est pertinente pour la collection concernée,
- est une bibliothèque numérique faisant partie des systèmes incontournables dans le domaine,
- est satisfaisante du point de vue des fonctionnalités offertes,
- est prometteuse en termes de durabilité et de développements futurs,
- offre des conditions financières supportables par la BCUL,
- ne fait pas double-emploi avec une bibliothèque numérique gérée par la BCUL,

alors elle est utilisée pour la collection en question. »<sup>51</sup>

Cette même politique mentionne également les principes suivis par la BCUL pour permettre la réutilisation de ses documents :

« La BCUL exploite diverses techniques permettant de diffuser ses ressources à travers les bibliothèques numériques et outils de découverte pertinents. Les techniques utilisées et envisagées comprennent le protocole OAI-PMH et les APIs définis par le standard IIIF. »<sup>52</sup>

Toutefois, si un travail très important est fourni par la BCUL pour diffuser ses documents numériques sur ses propres plateformes<sup>53</sup> ainsi que sur différents méta-catalogues, aucune information n'est fournie concernant l'accès OAI-PMH<sup>54</sup> (Open Archives Initiative Protocol for Metadata Harvesting) ou les APIs IIIF<sup>55</sup> (International Image Interoperability Framework). Cet état de fait est assez symptomatique des bibliothèques qui concentrent leur énergie sur le traitement, la numérisation, la préservation et la diffusion des fonds, sans véritablement proposer de solution technique pour réutiliser et exploiter l'ensemble de leurs données.

Une enquête effectuée par SWITCH dans le cadre du projet Connectome (4.2.1) auprès des chercheurs dans les domaines des sciences sociales et humaines (260 réponses reçues) a montré que les documents patrimoniaux détenus par les institutions GLAM sont une source qui n'est pas considérée comme suffisamment accessible. Selon les répondants, le problème peut résulter d'une couverture insuffisante des documents numérisés ainsi que de la non mise à disposition de ces documents sur le catalogue de la bibliothèque ou sur des méta-catalogues comme Europeana.<sup>56</sup> Les autres barrières à la réutilisation sont également : un format de numérisation de qualité insuffisante, la non prise en charge de certains protocoles techniques (notamment IIIF), et même une insuffisance au niveau de la structuration ou des métadonnées.<sup>57</sup> Parmi les critères principaux à la sélection des données dans le cadre de leur travail de recherche, on trouve (par ordre décroissant) : la confiance dans les sources, la présence de matériel additionnel (notamment la documentation), l'accès aux données brutes, la facilité d'accès, la qualité des données, l'usage d'outils familiers, la compréhension des

<sup>51</sup> BCU Lausanne, « Politique des bibliothèques numériques » (Document non publié interne à l'institution, 28 janvier 2019), 4.

<sup>52</sup> BCU Lausanne, 5.

<sup>53</sup> Plateforme des collections de la BCUL : <https://www.bcu-lausanne.ch/collections-et-patrimoine>

<sup>54</sup> « Open Archives Initiative Protocol for Metadata Harvesting », consulté le 14 juin 2022, <https://www.openarchives.org/pmh/>.

<sup>55</sup> « Home », International Image Interoperability Framework, consulté le 14 juin 2022, <https://iiif.io/>.

<sup>56</sup> Nicolai Hauf et al., « Data Reuse in the Social Sciences and Humanities : Project Report of the SWITCH Innovation Lab "Repositories & Data Quality" », 32, application/pdf, 2021, 19-20, <https://doi.org/10.21256/ZHAW-2404>.

<sup>57</sup> Ibid., 20.

métadonnées et finalement la transparence de la licence.<sup>58</sup> Comme cette enquête le montre, il y a donc un écart important entre les attentes des chercheurs et la mise à disposition des données patrimoniales.

### 1.3 Objectifs

L'enjeu de ce travail est de proposer des améliorations techniques concrètes tout en tenant compte de la réalité du terrain. Il est donc nécessaire de trouver un équilibre entre les améliorations qu'il serait souhaitable de fournir pour une réutilisation optimale des données, et les contraintes juridiques, financières, humaines et politiques.

La première partie de ce travail présente l'état de la situation actuelle afin de donner une vision d'ensemble : des données disponibles, des solutions utilisées pour les partager, et de l'adéquation de ces solutions avec les principes FAIR.

La partie suivante creuse la thématique de la mise à disposition en lien avec une collection choisie, ceci avec l'objectif de déterminer si les choix techniques de la BCUL concernant les normes, les formats et les méta-catalogues utilisés permettent l'exploitation de ces données à des fins de recherche et d'innovation.

La dernière partie de ce travail opère un retour aux collections patrimoniales numériques de la BCUL dans leur ensemble et a pour objectif de proposer des solutions pour améliorer leur interopérabilité.

Ce travail n'inclut pas la mise en pratique des améliorations qui sont proposées, car ces dernières vont au-delà de la politique actuelle de la BCUL.<sup>59</sup> L'objectif de ce travail consiste donc plus à démontrer :

- qu'il est possible de mettre à disposition les données de manière à permettre et même encourager leur réutilisation ;
- qu'il est utile de faire cet effort, car cela répond à un besoin direct des chercheurs ;
- qu'il est nécessaire de permettre la réutilisation de ces données à l'échelle internationale.

### 1.4 Méthodologie utilisée et regard critique

La méthodologie utilisée dans ce travail se divise en trois étapes :

1. Effectuer un état des lieux des collections numériques patrimoniales de la BCUL ainsi que des plateformes utilisées pour la gestion et la diffusion de ces documents. Cet état des lieux repose sur des entretiens avec : les différents responsables de collections, le responsable des bibliothèques numériques, et les responsables hiérarchiques de ces secteurs. Les tableaux récapitulatifs des catalogues, des fonds, et des projets se trouvent en annexe (9.1).

Conformément à la demande de la direction des études, cette partie inclut également l'état des lieux d'une deuxième bibliothèque, afin de mettre en perspective la situation de la

---

<sup>58</sup> Ibid., 16.

<sup>59</sup> Service de Communication, « Abrégé du Plan directeur 2021-2025 de la BCUL ».

BCUL par rapport à celle d'une autre institution suisse. La Bibliothèque de Genève (BGE) a été choisie pour cette analyse comparative, car sa situation rejoint celle de la BCUL sur plusieurs points. C'est une institution dont l'ampleur est comparable à celle de la BCUL. Comme cette dernière, elle accueille le dépôt légal de son canton. Il s'agit également d'une ancienne bibliothèque du Réseau romand (RERO) qui utilise aujourd'hui le logiciel Alma pour ses collections standards. Finalement, elle abrite également des collections patrimoniales importantes incluant des manuscrits, des archives et des documents iconographiques.

Afin de compléter cette comparaison, les solutions proposées aux bibliothèques affiliées à SLSP et RERO+ sont sommairement étudiées. L'analyse comparative avec la BGE se base sur les données récoltées dans le cadre d'un entretien mené avec Alexis Rivier, conservateur dans le secteur Projets et développement à la Bibliothèque de Genève. Les professionnels de RERO+ Pascal Repond, expert en bibliothèques, et Nicolas Prongué, Directeur adjoint, ont été consultés concernant les systèmes de gestion Sonar et RERO ILS.

2. Rechercher les projets innovants et méta-catalogues particulièrement adaptés à un type de collection choisi suite aux entretiens. La restriction de la recherche à un domaine précis est le cadre initialement donné à ce travail, cette limitation étant indispensable face au paysage extrêmement dense des projets d'humanités numériques dans le domaine patrimonial.

Cette étape se concentre donc sur Les Archives musicales<sup>60</sup> de la BCUL, dont les collections ont été sélectionnées pour être celles qui comptent le plus important nombre de documents numérisés librement accessibles et téléchargeables sur Patrimum. Il n'existe malheureusement pas encore de statistiques fiables concernant le nombre de consultations ou le nombre de téléchargements effectués depuis Patrimum. L'entretien réalisé avec Verena Monier, responsable des Archives musicales, a toutefois permis de mettre en évidence un intérêt majoritaire des usagers pour les partitions.

La recherche de projets innovants a été effectuée au niveau international. Dans les cas où plusieurs projets similaires ont été identifiés, ceux ayant une proximité géographique avec la BCUL ont été privilégiés dans le but de favoriser une collaboration locale.

3. Identifier les actions à entreprendre afin de favoriser la réutilisation des données de la BCUL au niveau des plateformes internes. Cette partie inclut également une présentation des méta-catalogues dont les technologies et formats de données optimiseraient l'interopérabilité des ressources de l'institution, ainsi qu'une analyse des avantages que ces méta-catalogues lui apporteraient.

Selon le concept, cette partie aurait également dû inclure l'état des points forts et des points faibles des bibliothèques numériques utilisées par la BCUL. Pour des raisons de structuration, cette partie a finalement été déplacée au chapitre 2.2.8. Le concept mentionne également l'inclusion dans ce chapitre d'une sélection de projets auxquels la BCUL pourrait participer. Or, également pour des raisons de structure, cette analyse est présentée à la fin de la partie précédente, soit dans le chapitre 3.6.

---

<sup>60</sup> « Archives musicales », BCUL, consulté le 14 juin 2022, <https://www.bcu-lausanne.ch/collections-patrimoine/archives-musicales/>.

## 2 Etat des bibliothèques numériques BCUL

Les services suivants de la BCUL possèdent des collections de type patrimonial : 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 de ces collections sont généralement accessibles sur l'une des deux plateformes propres à la BCUL, Patrinum (PATRImoine NUMérique) ou Scriptorium. De plus, la BCUL met à disposition certains lots de documents ou de notices sur plusieurs plateformes externes, dont notamment e-rara, e-codices ou encore Google books et Wikipédia.<sup>61</sup>

### 2.1 Les plateformes BCUL

Les deux plateformes spécifiques à la BCUL utilisées pour les collections patrimoniales sont Scriptorium et Patrinum. Ces plateformes ont été conçues avec des objectifs différents et ne sont donc pas utilisées pour les mêmes types de collections.

On peut également mentionner d'autres plateformes en lien avec la BCUL, à savoir :

- Alma digital. Cette plateforme est gérée par Renouvaud. Destinée à la littérature grise, elle est utilisée par plusieurs bibliothèques du réseau. Elle est actuellement testée par la BCUL. Etant donné sa potentielle utilisation future, elle a été ajoutée au tableau comparatif (9.1.1).
- Serval. Cette plateforme est gérée par l'UNIL. Elle est destinée aux fichiers numériques de l'UNIL et du CHUV.

#### 2.1.1 Scriptorium

Scriptorium a été mis en place en 2012. La solution utilisée est MediaINFO<sup>62</sup> et le prestataire la société Geneza<sup>63</sup>. Le projet a été lancé dans le but spécifique de donner un accès à la presse numérisée vaudoise ainsi qu'aux annuaires, almanachs, et Bulletins du grand conseil. Elle inclut également quelques périodiques suisses publiés hors du canton (documents issus de la Réserve précieuse).

C'est une plateforme orientée utilisateurs, très conviviale au niveau de l'interface et de la navigation. Elle ne dispose par contre pas de toutes les fonctionnalités nécessaires pour garantir la pérennité sur le long terme des documents numériques qui y sont stockés.

Scriptorium est consulté par environ 400 utilisateurs par jour. Dans sa globalité, la plateforme met à disposition environ 180 périodiques, plus de 560'000 fascicules et plus de 7 millions de pages. Les documents consultables sur Scriptorium sont sous droit d'auteur (signature d'une

---

<sup>61</sup> « Projet:Valdensia », in *Wikipédia*, 27 juin 2019, <https://fr.wikipedia.org/w/index.php?title=Projet:Valdensia&oldid=160466893>.

<sup>62</sup> « MediaINFO Digital Library », MediaINFO Digital Library, consulté le 16 juin 2022, <https://www.mediainfo.com/>.

<sup>63</sup> « About MediaINFO, Geneza Company », *MediaINFO Digital Library* (blog), consulté le 16 juin 2022, <https://www.mediainfo.com/about/>.

convention entre la BCUL et les éditeurs), leur utilisation est donc réservée à un usage privé et non commercial<sup>64</sup>.

## 2.1.2 Patrinum

Patrinum a été lancé en 2018. La plateforme est basée sur le framework Invenio<sup>65</sup>, également utilisé pour le Système Intégré de Gestion de Bibliothèques (SIGB) RERO ILS<sup>66</sup>. La société TIND<sup>67</sup> a été mandatée pour le développement et la maintenance de l'outil. Ce projet a été lancé avec un double objectif, à savoir : servir de système de gestion pour les collections de type archive gérées par la BCUL (remplacement de nombreuses bases de données maison) et permettre le stockage 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 utilisée pour les collections des Archives musicales, du Dépôt légal numérique, de la Documentation vaudoise, de l'Iconopôle, du Patrimoine sonore ainsi que des Manuscrits et archives. Patrinum est également utilisé comme dépôt institutionnel par certains membres du réseau Renouvaud, dont les HES (Haute Ecole Spécialisée) vaudoises, la HEP (Haute Ecole Pédagogique) et l'UNIL, qui y déposent leurs travaux d'étudiants.

On trouve environ 250'000 notices dans Patrinum, dont quelques 75'000 notices d'autorités. Près de 28'000 notices ont 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'autorités.

Chaque document numérique est associé à une licence qui précise le cadre légal de son utilisation<sup>68</sup>.

## 2.1.3 Mise en perspective des deux plateformes

Scriptorium a déjà 10 ans d'existence. C'est donc un outil stable et connu des utilisateurs. A l'inverse, Patrinum n'est sorti du mode projet qu'en 2019 et les collections de la BCUL y sont migrées petit à petit. La plateforme est donc en constante amélioration.

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 également 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. Les documents à cataloguer comme des notices bibliographiques sont donc dans cette catégorie qui regroupe les collections

---

<sup>64</sup> « Bienvenue sur Scriptorium ! », Scriptorium, s. d., <https://scriptorium.bcu-lausanne.ch/page/bienvenue>.

<sup>65</sup> « Welcome », INVENIO, consulté le 16 juin 2022, <https://inveniosoftware.org/>.

<sup>66</sup> « rero-ils », Python, GitHub (2017; repr., RERO+, 26 mai 2022), <https://github.com/rero/rero-ils>.

<sup>67</sup> « THERE'S A NEW ILS IN TOWN », TIND : Cern spin-off, consulté le 16 juin 2022, <https://www.tind.io/ils?hsLang=en>.

<sup>68</sup> Les conditions d'utilisation détaillées sur le site de Patrinum : <https://patrinum.ch/pages/?ln=fr&page=ConditionsUtilisation>.

suivantes : Dépôt légal numérique, Documentation vaudoise (ou Documents vaudois numériques), et quelques autres collections particulières.

- Archives : format MARCXML adapté selon ISAD(G). Les collections des Archives musicales, de l'Iconopôle et des Manuscrits et archives y sont regroupées.

Concernant les possibilités de partage des données et documents, les deux plateformes permettent la publication des métadonnées via le protocole OAI PMH, mais la seule publication active actuellement concerne les documents 'bibliothèque' de Patrinum. Quant à l'implémentation des APIs IIIF, il s'agit d'une amélioration en cours pour Patrinum, tandis que quelques problèmes de mise en conformité doivent encore être résolus pour Scriptorium. Des APIs sont également disponibles pour les deux plateformes, mais il n'y a pas de documentation publique à ce sujet. Les utilisateurs disposant d'un compte sur Patrinum<sup>69</sup> peuvent également télécharger les métadonnées à partir du résultat d'une recherche ou d'une sélection (fonction panier) dans différents formats dont MARC, MARCXML<sup>70</sup>, MODS (Metadata Object Description Schema)<sup>71</sup>, Dublin Core, HTML brief, BibTeX, EndNote, NLM, RefWorks<sup>72</sup>.

Les deux plateformes permettent de stocker des documents de type : texte, images, vidéos et son. Patrinum gère tous les formats de fichiers, mais n'offre d'interface de consultation que pour les JPEG et PDF. Il est par contre toujours possible d'offrir le téléchargement comme moyen de consultation. Scriptorium accepte les set JPEG2000/PDF/XML (documents texte/image) ainsi que les fichiers PDF (documents texte/image), MP4 (documents vidéo), AVI (vidéo), MOV (vidéo) et MTS (vidéo).

Patrinum et Scriptorium disposent d'une recherche simple et avancée intégrant notamment les recherches par expressions régulières. La recherche plein texte est activée par défaut pour Scriptorium mais pas dans Patrinum. Ce choix est dû au fait que Patrinum possède peu de documents numériques dont le texte est lisible pour les machines. A l'inverse, du côté de Scriptorium, la numérisation est prise en charge dans le cadre du projet Impresso<sup>73</sup> (DHLAB de l'EPFL) qui prend également en charge la reconnaissance textuelle, effectuée à l'aide d'IA.

Pour l'affichage des résultats, Scriptorium est la seule plateforme à mettre en évidence la portion du texte correspondante. De plus, cette plateforme permet aussi d'afficher les résultats sur une Timeline selon la date de publication.

Du côté de la consultation, les deux plateformes permettent le téléchargement (selon la licence pour Patrinum) et disposent d'une visionneuse pour les documents. Par contre, Patrinum ne dispose que de fonctions basiques (zoom, rotation) alors que Scriptorium permet aux usagés connectés de sélectionner des portions de texte et de le copier (les retours à la ligne dus au format PDF sont ignorés, ce qui optimise la réutilisation). La visionneuse dispose d'un mode de traduction automatique, mais la langue sélectionnée n'est pas mémorisée.

---

<sup>69</sup> Pour l'instant uniquement les professionnels disposent d'un compte, mais un développement est en cours pour permettre aux usagers d'utiliser leurs identifiants Renouvaud.

<sup>70</sup> MARCXML est un schéma XML basé sur le format MARC. Site officiel : <https://www.loc.gov/standards/marcxml/>.

<sup>71</sup> MODS est un vocabulaire de description bibliographique XML à la croisée du MARC et du Dublin Core. Site officiel : <http://www.loc.gov/standards/mods/>.

<sup>72</sup> BibTeX, EndNote, NLM et RefWorks sont des formats de citation.

<sup>73</sup> Site officiel : <https://www.epfl.ch/labs/dhlab/projects/impresso-2/>.

Pour les modalités de téléchargement, Patrinum a la particularité de permettre l'export des documents PDF sous format TXT (les retours à la ligne sont par contre conservés). Si plusieurs documents PDF sont rattachés à une même notice, ils peuvent tous être téléchargés dans un fichier ZIP. S'il s'agit d'images, elles peuvent être téléchargées ensemble dans un PDF. Pour Scriptorium, les utilisateurs connectés peuvent télécharger une sélection de pages d'un même document, une portion de page au format image ou une sélection textuelle directement utilisable comme citation.

Enfin, Scriptorium permet d'obtenir une URL pour chaque liste de résultats et chaque affichage particulier d'une page. Cette fonctionnalité peut notamment être utilisée pour faire des liens vers une collection. Patrinum permet de pointer de manière pérenne sur une notice (attribution de DOI) ou sur une page donnée d'un PDF, mais cette dernière fonction n'est pas documentée dans l'interface.

Pour aller plus loin, un tableau comparatif est disponible (9.1.1).

### 2.1.3.1 Synthèse

Comme on le constate, les deux plateformes sont très complémentaires au niveau des fonctionnalités :

D'un côté, Patrinum peut prendre en charge tous les formats de fichiers et les préserver à long terme. De plus, des options de téléchargement et d'export (y compris par lots) dans des formats standardisés permettent de facilement réutiliser les métadonnées sur d'autres plateformes ou outils. Quant aux documents, la licence est explicitement renseignée à côté de l'option de téléchargement, ce qui garantit que l'utilisateur soit informé de ses droits. Tout est donc déjà en place pour réutiliser les données, exception faite de l'absence de solution pour télécharger les ressources par lots.

De l'autre côté, Scriptorium est parfaitement optimisé pour que les utilisateurs aient de la facilité à trouver une information et puissent aisément en citer la source. La possibilité de se créer un compte utilisateur a récemment été mise en place afin de donner plus de droits aux usagers connectés. Les fonctionnalités actuellement en test sont la correction du résultat de l'OCR des documents, ainsi que l'ajouter des annotations qui peuvent être partagées.

Aucune des deux plateformes n'est vraiment adaptée pour mettre en valeur ses collections auprès d'un usager qui voudrait juste se promener et découvrir des documents ayant une valeur patrimoniale particulière. Toutefois, un effort dans cette direction est fourni via la page « Dans le rétro »<sup>74</sup> de Scriptorium, qui donne accès à plusieurs sélections de documents en lien avec différentes thématiques.

Concernant les principes FAIR, Scriptorium n'y répond que de manière très partielle. En effet, la plateforme ne fournit pas d'identifiant pérenne aux (méta)données. Elle permet en théorie le partage des données via des APIs dédiées, des APIs IIF et le protocole OAI-PMH, mais ces options sont inactives ou non documentées. Les données de Scriptorium ne peuvent donc pas être mises en lien avec d'autres ressources sur le web. Enfin, les métadonnées ne contiennent

---

<sup>74</sup> Voir : <https://scriptorium.bcu-lausanne.ch/page/58>.

pas d'informations relatives à la licence ou à la provenance (information renseignée de manière globale sur le site web), ce qui est une barrière à leur réutilisation.

Du côté de Patrinum, les principes FAIR sont plus proches d'être atteints. Toutefois, le point central qui consiste à partager des (méta)données dans un format lisible par la machine n'est pas rempli (situation identique à Scriptorium), ce qui empêche également de lier ces données à d'autres ressources. De plus, le format MARC/ISAD(G), spécifique à Patrinum, ne remplit pas la condition d'utiliser un format standard de données.

L'implémentation des APIs IIIF sur Patrinum et la mise en conformité de celles de Scriptorium ferait une grande différence concernant l'interopérabilité des données. En outre, des APIs accessibles aux utilisateurs et documentées seraient également très utiles pour permettre la réutilisation dans le cadre de la recherche.

Pour Patrinum, le défi se situe également dans le nombre restreint de documents sous licence libre ainsi que sur le nombre très important de notices non visibles, proche de 160'000. Ce nombre s'explique par la grande majorité des documents issus de migrations. Un laps de temps est donc nécessaire pour nettoyer les données (les anciennes bases restent d'ailleurs consultables dans l'intervalle).

De manière générale, la BCUL est en train d'ouvrir la réflexion pour permettre la réutilisation des données disponibles sur Scriptorium. En effet, un projet a été initié pour renégocier les conventions afin de permettre la mutualisation des contenus (par exemple le versement sur d'autres méta-catalogues) et l'usage des documents numériques (PDF et données OCR) dans le cadre du data mining. S'il était possible de mettre en place des conventions incluant systématiquement le droit de diffusion sur Patrinum, de réutilisation non commerciale et d'usage en tant que sets d'entraînement ou de test pour les IA, un nombre énorme de possibilités seraient ouvertes pour cette plateforme également.

La jeunesse de Patrinum a tendance à évincer la réutilisation des données des priorités. En effet, les préoccupations actuelles se portent plus sur l'amélioration des processus de travail dont en particulier : automatiser l'import des notices cataloguées sur Excel et améliorer le module de prêt (pas utilisé actuellement) pour que le lecteur puisse emprunter directement l'ensemble d'un dossier (ces deux points concernent uniquement les documents de la partie « Archives »). En parallèle, un travail continu est effectué pour augmenter le nombre de documents et de notices visibles pour les lecteurs ainsi que pour développer l'interface utilisateur (adaptation des facettes, affichage de l'arborescence, possibilité de se connecter avec ses identifiants Renouvaud). Il serait également utile de mettre en place des statistiques de consultation et de promouvoir la plateforme auprès des usagers, car elle semble encore peu connue.

## 2.2 Etat des fonds et plateformes externes

### 2.2.1 Les Archives musicales

Fondé en 1971, les Archives musicales 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<sup>e</sup> siècle à nos jours. De nombreux fonds portent sur la musique classique ou contemporaine, mais d'autres styles sont également représentés, comme le jazz (Sylvie

Courvoisier) et la musique folklorique (André Pernet, André-Olivier Clerc). Le service compte également des legs importants de chefs d'orchestre et interprètes au rayonnement international.<sup>75</sup>

Environ 70 inventaires papier ont été numérisés complètement. Quelques 2200 partitions et 1500 photographies ont été récemment numérisées et seront bientôt complètement cataloguées sur Patrinum. De manière générale, les fonds sont numérisés selon les besoins des usagers. Pour cette raison, un accent particulier a été mis sur les partitions. Les documents numérisés sur demande sont également chargés sur Patrinum.

Concernant la licence, seulement 30% des documents numériques sont libres de droit. Pour les documents sous droit, une convention est négociée par la BCUL, mais cette dernière n'inclut que très rarement un droit de diffusion. Sur Patrinum, le public peut consulter un peu plus de 300 documents numériques dans le domaine public.

Outre les documents numérisés, Les Archives musicales conservent aussi des documents récents nés numériques. Ces derniers sont souvent versés 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 MusicXML (3.3.2) et PDF sont également ajoutées, afin de permettre la visualisation des documents ainsi que leur réutilisation.

### 2.2.1.1 RISM Suisse (Répertoire International des Sources Musicales)

RISM Suisse<sup>76</sup> est une plateforme qui réunit des partitions manuscrites et imprimées issues des bibliothèques et archives suisses.<sup>77</sup> Cette plateforme s'inscrit dans le projet international RISM<sup>78</sup>, fondé en 1952 par l'organisation à but non lucratif du même nom, dont l'objectif est d'inventorier de manière exhaustive les sources musicales issues du monde entier.<sup>79</sup> Les initiateurs du projet sont l'Association internationale des bibliothèques, archives et centres de documentation musicaux (AIBM) et de la Société internationale de musicologie (SIM).<sup>80</sup> Ces deux grands acteurs ont également lancé trois autres entreprises à la suite de RISM, à savoir : le RILM (Répertoire international de littérature musicale) en 1967; le RIDIM (Répertoire international d'iconographie musicale) en 1971 ; et le RIPM (Répertoire international de la presse musicale) en 1980.<sup>81</sup> RISM se définit comme « la plus grande et la seule organisation mondiale qui documente les sources musicales écrites ».<sup>82</sup> Des groupes de travail nationaux œuvrent actuellement dans 36 pays.<sup>83</sup>

---

<sup>75</sup> « Archives musicales ».

<sup>76</sup> Voir : <https://rism.digital/rism-ch>

<sup>77</sup> « RISM Suisse », RISM Digital Center, consulté le 22 juillet 2022, <https://rism.digital/rism-ch.html>.

<sup>78</sup> Voir : <https://opac.rism.info>

<sup>79</sup> « About us », RISM Digital Center, consulté le 22 juillet 2022, <https://rism.digital/organization.html>.

<sup>80</sup> Laurence Languin, « Panorama des ressources documentaires en ligne », in *Musique en bibliothèque*, Bibliothèques (Paris: Éditions du Cercle de la Librairie, 2012), paragr. 5, <https://doi.org/10.3917/elec.alix.2012.02.0323>.

<sup>81</sup> Languin, paragr. 5.

<sup>82</sup> « RISM », Répertoire International des Sources Musicales, consulté le 5 août 2022, <https://rism.info/>.

<sup>83</sup> Keil et Pugin, « Das Internationale Quellenlexikon der Musik, RISM », part. 1.

Le catalogue international de RISM agrège plus de 1'601'000 notices dont 103'000 sont liées à des fichiers numériques. RISM Suisse agrège environ 101'000 notices issues de plus de 400 institutions suisses et environ 10'300 partitions numériques y sont disponibles.

RISM Suisse contacte activement les institutions et leur équipe catalogue directement certains fonds choisis. Les premières ressources ajoutées à la plateforme étaient antérieures à 1800. Actuellement, les ressources jusqu'à 1900 sont en cours d'ajout. La plateforme permet de visualiser les documents disposant d'un manifeste IIF et les métadonnées des notices sont très détaillées. De plus, il est possible d'effectuer des recherches par incipit (information renseignée dans près de 85'000 documents pour la Suisse), en jouant les notes sur un clavier virtuel. Enfin, les noms de personnes et de collectivités sont liés aux autorités GND<sup>84</sup> et VIAF<sup>85</sup>.

La BCUL a déjà versé 4 fonds<sup>86</sup>, soit plus de 1'300 pièces, dont 9 sont accompagnées de numérisations (qui peuvent pointer sur Patrinum ou des ressources numérisées par d'autres institutions comme Gallica, Internet Archives, etc.). La majorité des fonds BCUL datant du 20<sup>e</sup> siècle, les Archives musicales prévoient un nouveau versement lorsque la plateforme sera prête à accueillir ces documents.

### 2.2.2 L'Iconopôle

Créé en 2019, l'Iconopôle a pour mission de conserver, inventorier, valoriser et acquérir les documents iconographiques liés à l'histoire vaudoise. Ses collections comprennent notamment des négatifs, des tirages, des albums, des diapositives, des affiches ainsi que des fonds complets relatifs à différentes personnalités (dont Nitsche, Pizzotti, Jeker) constitués de boîtes d'archives, livres ou encore de catalogues d'exposition. Ces collections s'étendent de 1850 jusqu'à nos jours.<sup>87</sup>

Les notices d'un peu plus de mille documents sont publiquement accessibles sur Patrinum. Environ 250 de ces notices sont rattachées à un document numérique, dont environ 60 sont sous droit indéterminé, et très peu dans le domaine public.

L'Iconopôle étant un service récent, les spécificités relatives à certains types de documents, comme les affiches ou les négatifs, ne sont pas reflétées par les métadonnées disponibles sur Patrinum. Un format plus adéquat que MARC/ISAD(G) serait par exemple CDWA<sup>88</sup>.

L'Iconopôle a versé environ mille affiches sur le Catalogue Collectif Suisse des Affiches (CCSA)<sup>89</sup> géré par la Bibliothèque nationale suisse (BN). Certaines de ces affiches étant sous droit, le parti a été pris de toujours rendre disponible l'image dans une résolution de qualité suffisamment basse pour que le droit d'auteur ne s'applique pas (400 px max) ; un fichier en haute résolution non accessible au public est également versé.

---

<sup>84</sup> Voir : <https://explore.gnd.network>

<sup>85</sup> Voir : <https://viaf.org>

<sup>86</sup> Voir : <https://rism.online/institutions/30000526/sources>

<sup>87</sup> « Iconopôle », BCUL, consulté le 15 mai 2022, <https://www.bcu-lausanne.ch/collections-patrimoine/iconopole/>.

<sup>88</sup> « Related Standards », *ICOM CIDOC* (blog), consulté le 16 mai 2022, <https://cidoc.mini.icom.museum/working-groups/lido/lido-overview/about-lido/related-standards/>.

<sup>89</sup> Voir :

[https://www.posters.nb.admin.ch/discovery/collectionDiscovery?vid=41SNL\\_53\\_INST:posters&collectionId=8145718840003978&lang=fr](https://www.posters.nb.admin.ch/discovery/collectionDiscovery?vid=41SNL_53_INST:posters&collectionId=8145718840003978&lang=fr)

### 2.2.2.1 Memobase

Memobase (9.1.3) est la plateforme pour le patrimoine audiovisuel au niveau suisse. Elle inclut : des films, des photographies, des enregistrements audios, des diaporamas, des vidéos, ainsi que des enregistrements radio et télévision<sup>90</sup>. Fondée en 2001 par Memoriav (centre de compétence suisse dans le domaine audiovisuel), elle rend accessible aujourd'hui quelques 400'000 documents.<sup>91</sup>

La BCUL a rejoint récemment le réseau des institutions rattachées à Memoriav. Un projet pilote a été lancé pour l'import de fonds photographiques et de films fixes sur Memobase. Toutefois, aucun document de la BCUL n'est encore disponible sur cette plateforme<sup>92</sup>.

### 2.2.3 Les Manuscrits et archives

Le service des Manuscrits et archives, dont les premières collections remontent à 1932, conserve les documents relatifs à des personnalités ou institutions marquantes du canton de Vaud. Actuellement, y sont conservés plus de 600 fonds, de nombreuses pièces isolées, une trentaine de manuscrits médiévaux et des cartes géographiques anciennes.<sup>93</sup> Aujourd'hui le service conserve également des documents nés numériques tels que des fichiers texte (au format Word). Un nouveau défi concerne le stockage, la conservation et la consultation des boîtes mails qui représentent la correspondance actuelle des auteurs.

Ces documents ont été initialement catalogués dans des bases FileMaker (toujours accessibles) et sont en cours de migration vers Patrimum. Près de 600 notices, cataloguées entre le niveau du dossier et de la pièce, ont déjà été migrées. Le nombre de numérisations dans le domaine public accessibles sur Patrimum est actuellement d'environ 60.

Certains manuscrits médiévaux sont quant à eux mis à disposition sur la plateforme suisse e-codices<sup>94</sup>. Concernant les quelques 900 cartes géographiques, elles ne sont pas cataloguées sur Patrimum mais sur Renouvaud, et une trentaine d'entre elles sont également disponibles sous format numérique via la plateforme réro doc (9.1.3).

Le traitement des fonds et le choix des numérisations est priorisé en fonction des projets de recherche. Il arrive que des documents soient numérisés par lot mais que le catalogage reste au niveau du dossier ou du sous-dossier, ce qui est problématique car un catalogage à la pièce est nécessaire pour le partage via IIF.

Le droit d'auteur constitue une problématique complexe pour ces fonds. La mise à disposition repose souvent sur des conventions anciennes qui ne clarifient pas tous les aspects. Concernant les documents natifs numériques, notamment les boîtes mail, le problème ne touche plus seulement le droit d'auteur mais également la protection des données, car ces documents comportent beaucoup plus d'informations personnelles qu'une correspondance papier.

---

<sup>90</sup> « Participer à Memobase », MEMORIAV MEMOBASE, consulté le 26 juin 2022, <https://memobase.ch/fr/info-memobase/rejoignez-memobase>.

<sup>91</sup> « À propos de Memobase », MEMORIAV MEMOBASE, consulté le 16 mai 2022, <https://memobase.ch/fr/info-memobase>.

<sup>92</sup> Page de la BCUL sur MEMOBASE :

<https://memobase.ch/fr/institution/bcl?term=lausanne&context=institutions&position=2>

<sup>93</sup> « Manuscrits et archives », BCUL, consulté le 12 mai 2022, <https://www.bcu-lausanne.ch/collections-patrimoine/manuscritsetarchives/>.

<sup>94</sup> Voir : <https://www.e-codices.unifr.ch>

Le service des manuscrits participe en outre à différents projets, on peut mentionner les plateformes Egodocuments.ch,<sup>95</sup> LumièresLausanne<sup>96</sup> et e-codices.<sup>97</sup>

Actuellement, un défi majeur pour le service des Manuscrits et archives est de mettre à disposition des transcriptions des documents manuscrits. En effet, si certains documents bénéficient déjà d'une transcription, disponible sur le site LumièresLausanne, il est également envisagé de participer à la plateforme de crowdsourcing From the Page,<sup>98</sup> afin d'obtenir des transcriptions manuelles pour un plus grand nombre d'entre eux. L'étape suivante serait d'utiliser ces transcriptions manuelles pour entraîner l'algorithme de machine learning de eScriptorium<sup>99</sup>, afin de bénéficier d'une transcription de tous les documents manuscrits.

### 2.2.3.1 E-Codices

Il s'agit d'une bibliothèque virtuelle suisse pour la mise à disposition des manuscrits médiévaux. Cette plateforme existe depuis 2008 et regroupe aujourd'hui environ 2'600 documents. Initialement lancé par l'Université de Fribourg, ce projet était soutenu depuis 2013 par swissuniversities<sup>100</sup>. Malheureusement, il existe actuellement des incertitudes concernant le financement futur de la plateforme.<sup>101</sup>

Depuis 2020, la BCUL participe à cette plateforme, avec, à ce jour, une sélection de 15 manuscrits médiévaux<sup>102</sup>. Les standards de numérisations nécessaires pour participer à e-codices sont exigeants, la BCUL a donc dû faire scanner les documents concernés par un prestataire externe. Les métadonnées doivent également être plus poussées que celles qui existent dans Patrimium<sup>103</sup>. Un nouveau projet de versement sur e-codices devrait commencer prochainement (l'identification des documents à numériser est en cours).

### 2.2.4 Le Dépôt légal numérique

Le Dépôt légal collecte tous les documents édités dans le canton de Vaud depuis 1938. Cela concerne les livres, les revues, les journaux, les bulletins, les brochures, les rapports annuels, les rapports scientifiques, les catalogues d'exposition et les programmes.<sup>104</sup> Les publications vaudoises antérieures à 1938 font partie de la Réserve précieuse.

Depuis 2014, la LPMI<sup>105</sup> précise que les publications au format numérique sont également incluses, soit notamment les eBooks. Patrimium a été développé entre autres pour répondre à ce nouveau besoin et permettre de conserver à long terme les collections nées numériques, les collections numérisées et les bases de données. À l'exception de la presse vaudoise numérisée et des documents numérisés antérieurs à 1938, mis à disposition sur Scriptorium, les autres documents numériques sont catalogués sur Patrimium et moissonnés par Renouvaud

---

<sup>95</sup> Voir : <https://wp.unil.ch/egodocuments>

<sup>96</sup> Voir : <https://lumières.unil.ch>

<sup>97</sup> Ces plateformes sont décrites dans le tableau comparatif en annexe : 9.1.3.

<sup>98</sup> Voir : <https://fromthepage.com>

<sup>99</sup> Voir : <https://escriptorium.fr>

<sup>100</sup> Voir : <https://www.swissuniversities.ch>

<sup>101</sup> « Brève histoire », e-codices, consulté le 22 juin 2022, <https://www.e-codices.unifr.ch/fr/about/history>.

<sup>102</sup> « e-codices – Bibliothèque virtuelle des manuscrits en Suisse », consulté le 18 février 2022, <https://www.e-codices.unifr.ch/fr>.

<sup>103</sup> Par exemple, la reliure est décrite de manière séparée.

<sup>104</sup> Heloise Hahling, « Dépôt légal », BCUL, consulté le 8 juillet 2022, <https://www.bcu-lausanne.ch/collections-patrimoine/depot-legal/>.

<sup>105</sup> 446.12 – Loi sur le patrimoine mobilier et immatériel (LPMI) du 8 avril 2014, Article 32.

via le protocole OAI-PMH. Tous les documents imprimés du Dépôt légal sont directement catalogués dans Renouvaud.

De manière générale, les documents numériques catalogués sur Patrinum ne sont pas accessibles en ligne mais en consultation sur place sur le site Riponne, car le droit d'auteur s'applique (y compris pour des documents accessibles librement en ligne sur le site des éditeurs). La récolte des documents est effectuée manuellement (pour l'instant, il n'y a pas de dépôt direct des éditeurs sur Patrinum), le défi actuel est donc de mettre en place un processus de récolte qui soit le plus systématique possible.

## 2.2.5 La Documentation vaudoise

La Documentation vaudoise est le service chargé de préserver et mettre à disposition toutes les publications qui concernent le canton de Vaud. La plus grande partie de la collection est constituée d'ouvrages imprimés tels que des livres, des guides ou des publications évoquant des artistes, associations, communes, institutions ou événements relatifs au canton. Toutefois, on y trouve également des cartes, des plaquettes évoquant la vie de sociétés locales, des enregistrements vidéo, des DVD documentaires, des portraits filmés, de la littérature enregistrée, des revues et des articles.<sup>106</sup> Cela représente environ 13'000 documents qui sont intégralement catalogués et empruntables via Renouvaud. Le degré de préservation requis pour ces documents est moins élevé que celui du Dépôt légal, car c'est uniquement le contenu des documents qui doit être préservé et non le support.

Du côté numérique, près de 500 documents sont disponibles sur Patrinum, dont environ 260 documents sonores (CD importés au format mp3) et 170 vidéos (VHS importées au format mpg). Pour ces documents, le changement de support rend caduc le droit de prêt, seules les métadonnées sont donc accessibles au public sur Patrinum. Un travail est en cours pour retrouver les producteurs afin de renégocier les conventions. Si ce travail de négociation aboutit, il serait également envisageable de rendre accessibles ces documents via Memobase. Du côté de Patrinum, il serait nécessaire d'ajouter un lecteur audio et vidéo afin de permettre la mise à disposition des documents dont seul le streaming serait autorisé.

Les ressources numériques de la Documentation vaudoise sont complétées par 3 bases de données, dont : une base de liens<sup>107</sup> vers des sites web, des blogs et des bases de données ; une base des communes vaudoises<sup>108</sup> disponible sur Patrinum ; et une base des personnalités vaudoises<sup>109</sup> également accessible sur Patrinum.<sup>110</sup>

La base des personnalités regroupe plus de 7000 notices d'autorités. Initialement, il s'agissait d'une base développée en interne, dont les notices ont été migrées sur Patrinum. Ces dernières sont encore en cours de nettoyage (un travail important est notamment nécessaire concernant la protection des données, car de nombreuses personnalités sont encore vivantes). Les notices issues de cette base ont été utilisées en 2011 dans le cadre du projet

---

<sup>106</sup> Myriam Jouhar, « Documentation vaudoise », BCUL, consulté le 8 juillet 2022, <https://www.bcu-lausanne.ch/collections-patrimoine/documentation-vaudoise/>.

<sup>107</sup> Voir : [https://db-prod-bcul.unil.ch/dbbcu/docvd/lec\\_rub.php?Type=1](https://db-prod-bcul.unil.ch/dbbcu/docvd/lec_rub.php?Type=1); la plupart de ces sites sont archivés de manière pérenne par la Bibliothèque nationale suisse via le projet e-helvetica.

<sup>108</sup> Voir : <https://patrinum.ch/collection/Municipality>

<sup>109</sup> Voir : <https://patrinum.ch/search?cc=PersonnalitesVD&ln=fr&c=PersonnalitesVD>

<sup>110</sup> Jouhar, « Documentation vaudoise ».

Valdensia<sup>111</sup> pour créer automatiquement quelques 800 articles Wikipédia sur des personnalités vaudoises<sup>112</sup>. A noter que ce projet dérogeait au principe de Wikipédia selon lequel les articles doivent contenir un texte original et non être copiés. Ce versement a été unique et les notices ont donc évolué de manière distincte entre la base de la BCUL et Wikipédia. A la suite de ce projet, d'autres articles Wikipédia en lien avec des personnalités vaudoises ont été créés, cette fois manuellement et avec des textes originaux. Ces projets en lien avec Wikipédia ont permis de largement diffuser quelques données de la base, mais dans son ensemble, cette dernière reste peu connue des usagers. Il serait notamment intéressant que les données de cette base soient moissonnées par d'autres catalogues, comme le Dictionnaire Historique de la Suisse (DHS)<sup>113</sup>. Un autre projet intéressant serait de l'ouvrir au crowdsourcing via une plateforme tierce.

## 2.2.6 La Réserve précieuse

La Réserve précieuse regroupe plusieurs types de documents, dont des imprimés des 15<sup>e</sup> et 16<sup>e</sup> siècles et des incunables (souvent enrichis d'annotations manuscrites), des publications consacrées au canton de Vaud, publiées ou rédigées par des Vaudois avant le 20<sup>e</sup> siècle, des livres rares (notamment sélectionnés pour leurs illustrations ou leur reliure), et des livres modernes à tirage limité ou unique (souvent enrichis de gravures).<sup>114</sup> Ces documents sont catalogués sur Renouvaud et disponibles en consultation surveillée.

Depuis 2012, une sélection de plus de 100'000 documents antérieurs à 1871 et libres de droit sont accessibles au format numérique sur Google Books<sup>115, 116</sup>. Par la suite, la mise à disposition des documents de la Réserve précieuse au format numérique a continué avec le projet Scriptorium (2.1.1). Depuis 2021, la BCUL participe également à e-rara<sup>117</sup>, la plateforme suisse pour les imprimés du 15<sup>e</sup> au 20<sup>e</sup> siècle (9.1.3), en y mettant à disposition quelques 120 imprimés du 16<sup>e</sup> siècle.

La Réserve précieuse est également à l'origine de la création de 4 bases de données internes, qui sont toujours accessibles au public et utilisées dans le cadre de la recherche sur le patrimoine imprimé vaudois. Il s'agit de Biblos 18<sup>118</sup>, une base bibliographique des ouvrages publiés ou imprimés à Lausanne au 18<sup>e</sup> siècle ; de Fleuron<sup>119</sup>, en lien direct avec la base précédente ; de R.I.E.C.H<sup>120</sup>, le répertoire des éditeurs et imprimeurs suisses actifs du 16<sup>e</sup> au 18<sup>e</sup> siècle ; et enfin de Passe-Partout<sup>121</sup>. Également en lien avec Fleuron, cette dernière base

---

<sup>111</sup> Voir : <https://fr.wikipedia.org/wiki/Projet:Valdensia>

<sup>112</sup> Jeannette Frey, « Le projet Valdensia de la BCU Lausanne: une première expérience GLAM avec Wikimedia », arbidio, mars 2015, <https://www.arbido.ch/fr/edition-article/2015-1/glam-und-wikimedia/le-projet-valdensia-de-la-bcu-lausanne-une-premiere-experience-glam-avec-wikimedia>.

<sup>113</sup> Voir : <https://hls-dhs-dss.ch/>

<sup>114</sup> Marc Tiefenauer, « Réserve précieuse », BCUL, consulté le 8 juillet 2022, <https://www.bcu-lausanne.ch/collections-patrimoine/reserve-precieuse/>.

<sup>115</sup> Jeannette Frey, « Googlos – un partenariat public-privé pour la mise en ligne du patrimoine imprimé », arbidio, février 2016, <https://www.arbido.ch/fr/edition-article/2016/delocalise-achete-externalise/googlos-un-partenariat-public-privé-pour-la-mise-en-ligne-du-patrimoine-imprime>.

<sup>116</sup> Des liens vers ces documents numériques sont intégrés au catalogue Renouvaud.

<sup>117</sup> Page de présentation de la BCUL sur e-rara : [https://www.e-rara.ch/lac1\\_g](https://www.e-rara.ch/lac1_g)

<sup>118</sup> Voir : <https://db-prod-bcul.unil.ch/biblos/intro.php>

<sup>119</sup> Voir : <https://db-prod-bcul.unil.ch/ornements/scripts/index.html> ; Fleuron répertorie les ornements typographiques présents dans les imprimés du 16<sup>e</sup> au 18<sup>e</sup> siècle à Lausanne.

<sup>120</sup> Voir : <https://db-prod-bcul.unil.ch/rieich/intro.php>

<sup>121</sup> Voir : <https://bcutodai.unil.ch/BCUTodai/app/Todai.do;jsessionid=74ED5BE8D44BDF65E937F42653364DC3>

est un outil de recherche identifiant des ornements typographiques (16<sup>e</sup> au 18<sup>e</sup> siècle) par comparaison automatisée de leur forme ou à l'aide de mots-clés descriptifs. Elle permet de retrouver l'atelier d'imprimerie en lien avec un ornement donné.<sup>122</sup>

## 2.2.7 La Phonothèque

Un fonds d'environ 2600 documents sonores (disques compacts et vinyles) en relation avec le canton de Vaud (musiciens, groupes ou labels) sont disponibles sur le catalogue Renouvaud.<sup>123</sup> Cette collection inclut la musique classique, le jazz, la musique pop, le rock et la chanson. Elle est complétée par des disques pour enfants, ainsi que de la musique du monde et de la musique de films.<sup>124</sup> De même que pour la documentation vaudoise, tous les documents sont prêtables.

Cette collection est complétée par 3 bases de données importées sur Patrinum. Il s'agit de la base des labels vaudois (117 notices), de la base des musiciens vaudois (577 notices), ainsi que de la base des ensembles musicaux vaudois (139 notices).

## 2.2.8 Synthèse relative à l'état des fonds et plateformes externes

Comme il l'a été démontré, la BCUL fournit un effort important pour numériser et diffuser ses collections patrimoniales numériques sur le web. En 2011, elle a utilisé la base des autorités vaudoises pour créer 800 articles Wikipédia. En 2012, près de 100'000 imprimés antérieurs à 1871 et livres de droit ont été mis à disposition sur Google Books. Cette même année a aussi vu le lancement de Scriptorium, qui rend disponible la presse vaudoise numérisée. Puis, en 2018, c'est le lancement de Patrinum qui héberge toutes les collections numériques restantes. Enfin, on peut noter la participation de la BCUL aux plateformes e-codices, dès 2020, et à e-rara, dès 2021.

Parmi ces nombreux projets, le cas de la presse vaudoise, intégralement disponible sur Scriptorium en incluant une reconnaissance OCR de tous les textes, est exceptionnel autant au niveau du nombre de documents traités (plus de 560'000 fascicules), que des conventions, qui permettent toutes la publication des données et le téléchargement pour un usage privé. Le projet actuel, qui vise à uniformiser les conventions pour permettre l'usage des données dans le cadre du data mining, s'inscrit parfaitement dans l'idée de rendre les données réutilisables pour les humanités numériques.

Concernant les fonds du Dépôt légal numérique et de la Documentation vaudoise, les documents numériques sont presque exclusivement sous droit d'auteur, bien que des réflexions soient en cours pour permettre la publication de certaines collections. Aucune réutilisation des données n'est donc possible pour ces trois collections, à l'exception des métadonnées et des notices d'autorités. Il en va de même du côté de la Phonothèque, où seules des notices d'autorités sont disponibles sur Patrinum. Toutefois, il serait intéressant de rendre disponibles les métadonnées et notices d'autorités, pour chacun de ces fonds, dans le cadre d'une publication OAI-PMH documentée.

---

<sup>122</sup> Tiefenauer, « Réserve précieuse ».

<sup>123</sup> Suzanne Kaufmann, « Patrimoine sonore », BCUL, consulté le 20 juillet 2022, <https://www.bcu-lausanne.ch/collections-patrimoine/patrimoine-sonore/>.

<sup>124</sup> Kaufmann, part. Musiciens vaudois.

Pour la Réserve précieuse, les documents numérisés, à l'exception de ceux mis à disposition sur Scriptorium, sont uniquement disponibles sur Google Books ou e-rara (120 documents). Google Books propose une interface attractive pour les usagers et la reconnaissance OCR appliquée permet la recherche plein texte, même dans les anciens imprimés. Par contre, la plateforme est peu compatible avec les principes FAIR : pas d'identifiant pérenne, des métadonnées restreintes (sujettes à erreur car souvent créées à partir d'une reconnaissance automatique des contenus) et n'incluant pas la licence relative à chaque document, pas d'usage de standards ou de vocabulaire contrôlé. Toutefois, les (méta)données sont accessibles via des APIs<sup>125</sup> gratuites, bien documentées, et très réutilisées ;<sup>126</sup> le cadre de cette réutilisation n'est pas contre pas celui des humanités numériques. Afin de faire profiter la communauté scientifique de ces documents, il serait donc intéressant de les publier également sur une plateforme respectant les principes FAIR. Du côté d'e-rara, c'est une plateforme parfaitement FAIR, dont les données sont partagées via IIF et OAI-PMH.

Les collections des Archives musicales, de l'Iconopôle et du service des Manuscrits et archives ont récemment été migrées sur Patrinum. Les fonds continuent à être numérisés en fonctions des projets de recherche et des opportunités. Peu de fonds sont sous licence libre, et une partie importante des notices sont cachées pour les utilisateurs, car de qualité insatisfaisante suite à leur migration. En effet, les bases FileMaker ne possédant pas d'arborescence, plusieurs notices pouvaient exister pour un même document. Ce fait a complexifié la migration et, en attendant d'être dédoublonnées, les notices restent masquées dans Patrinum. Le service des Manuscrits et archives a choisi de laisser les anciennes bases à disposition du public dans l'intervalle.

Du côté des plateformes externes, le service des Manuscrits et archives participe à e-codices. De plus, des réflexions sont en cours pour que les documents bénéficient d'une reconnaissance textuelle. L'Iconopôle participe déjà au catalogue CCSA pour les affiches et participera probablement à Memobase, qui est une plateforme largement optimisée pour la réutilisation des données, ainsi qu'à notreHistoire.ch<sup>127</sup>, qui repose sur le principe du crowdsourcing. Enfin, les Archives musicales participent à RISM Suisse, une plateforme également très adaptée à la réutilisation des données.

Les plateformes e-codices et Memobase respectent les principes FAIR. Concernant RISM et CCSA, seule l'indication de la licence dans les métadonnées est manquante. Les données de RISM sont accessibles en Linked Open Data, tandis que Memobase utilise les technologies du linked data, mais sans que les données soient systématiquement ouvertes.

## 2.3 État des plateformes relatives à la BGE

La Bibliothèque de Genève (BGE) est composée de 4 sites, à savoir : Bastions, Centre d'iconographie, Musée Voltaire et La Musicale. Comme la BCUL, la BGE est en charge du dépôt légal de son canton et possède de nombreux fonds patrimoniaux, tels que des imprimés

---

<sup>125</sup> Voir : <https://developers.google.com/books/docs/v1/using>

<sup>126</sup> Voir la liste des utilisateurs : <https://developers.google.com/books/casestudies>

<sup>127</sup> Voir : <https://notrehistoire.ch>

anciens, des manuscrits, des fonds d'archives, des documents iconographiques, des affiches, des cartes, des plans, des documents audio-visuels ou encore des partitions.<sup>128</sup>

La BGE partage ces documents en les mettant physiquement à disposition ainsi que numériquement pour une partie d'entre eux. Cette mise à disposition numérique s'effectue via diverses plateformes internes et externes.

Au niveau de la mise à disposition des ressources numériques, la BGE suit une politique hybride. D'une part, elle développe sa présence numérique au niveau régional. Elle a, à ce titre, cofondé une plateforme de recherche multibases pour la mise à disposition des manuscrits et archives qui inclut : les Archives de la Ville de Genève (AVG), l'Institut et Musée Voltaire (IMV), le Muséum Genève (MHNG) et le Conservatoire (CJBG).<sup>129</sup> Les ressources iconographiques sont cataloguées sur un système commun avec le Musée d'art et d'histoire de Genève<sup>130</sup>, mais elles sont mises à disposition via un frontend spécifique à la BGE<sup>131</sup>.

D'autre part, l'institution cultive sa présence sur les plateformes nationales lorsqu'elles ont suffisamment d'ampleur. On peut citer e-rara, e-codices, ou encore CCSA.

### 2.3.1 Plateforme des manuscrits et archives

Utilisée pour la mise à disposition des fonds Voltaire<sup>132</sup> et des fonds d'archives privés<sup>133</sup> dont de nombreux papyrus, la plateforme abrite en tout près de 20'000 documents numérisés publiquement accessibles et téléchargeables. Elle repose sur le logiciel Ligeo archives<sup>134</sup> de la société Empreinte Digitale. Lancée en 2021, la plateforme a pour objectif principal de rendre accessible les notices et documents numériques aux usagers. Les professionnels ne cataloguent pas directement sur la plateforme, mais ont conservé leur ancien processus de travail, ce qui génère un fichier au format XML/EAD<sup>135</sup> (soit qui respecte la norme ISAD(G)), automatiquement importé sur Ligeo archives. Les fichiers numérisés sont quant à eux importés manuellement ; les formats PDF et JPEG sont acceptés et certains PDF bénéficient d'une reconnaissance OCR. Chaque document bénéficie d'un identifiant pérenne ARK (Archival Resource Key)<sup>136</sup>.

L'ensemble du catalogue des manuscrits et archives est uniquement consultable sur cette plateforme, qui permet également de stocker des notices non visibles pour les usagers. Elle est donc également utile aux professionnels en tant que système de gestion. Par contre, elle ne permet pas de pérenniser les documents numériques sur le long terme.

La plateforme continue à être activement développée et va notamment être étendue à d'autres institutions. En outre, il est prévu que des documents bénéficiant d'une reconnaissance HTR soient prochainement inclus. La plateforme dispose d'une visionneuse

---

<sup>128</sup> « Connaître la Bibliothèque - Présentation », Bibliothèque de Genève, consulté le 14 juillet 2022, <http://institutions.ville-geneve.ch/fr/bge/connaître-la-bibliotheque/la-bibliotheque/presentation/>.

<sup>129</sup> Voir : <https://archives.bge-geneve.ch/n/recherche-multi-institutions/n:98>

<sup>130</sup> Voir : <https://collections.geneve.ch/mah>

<sup>131</sup> Voir : [bge-geneve.ch/iconographie](https://bge-geneve.ch/iconographie)

<sup>132</sup> Voir : <https://voltaire.bge-geneve.ch>

<sup>133</sup> Voir : <https://archives.bge-geneve.ch>

<sup>134</sup> Voir : <https://www.ligeo-archives.com>

<sup>135</sup> Voir : <https://www.loc.gov/ead>

<sup>136</sup> Voir : <https://arks.org>

avancée qui permet notamment de régler l'éclairage et le contraste. Il est de plus aisé de signaler une erreur en lien avec une page ou un document donné.

Les documents numériques peuvent être librement téléchargés en haute résolution et sont réutilisables sous la licence CC BY-NC<sup>137</sup>.<sup>138</sup> Concernant les possibilités d'export par lot, les documents seront très prochainement disponibles via IIF. Quant aux métadonnées, elles pourraient être accessibles via OAI-PMH, mais cette fonction n'est actuellement pas activée et elles ne sont pas non plus téléchargeables manuellement depuis l'interface utilisateur. En outre, il n'y a pas d'APIs qui soient mises à disposition.

### 2.3.2 Plateforme iconographique

Près de 43'000 documents iconographiques numériques sont accessibles depuis cette plateforme. En backend, c'est le système MuseumPlus<sup>139</sup> de la société Zetcom qui est utilisé. Le frontend actuel date de 2020 et est un développement spécifique à la BGE effectué sur Drupal. La plateforme sert à la fois de système de gestion en interne et de catalogue public pour la mise à disposition des notices et des documents numériques.

Lorsque des documents numérisés sont encore sous droit d'auteur, ce qui est souvent le cas, une vignette est tout de même mise à disposition sur la plateforme. Les documents numérisés sont accessibles au format JPEG. Toutefois, la plateforme pourrait également gérer des documents multimédias si cela était nécessaire.

La plateforme est constamment enrichie de nouvelles ressources, dont près de 300 documents ajoutés en juin 2022. La BGE commence notamment à recevoir des lots importants de fonds nés numériques, ce qui crée de nouvelles problématiques, telles que le besoin de trouver des solutions pour automatiser le traitement et l'import des données, ou celui de remédier au déséquilibre créé entre le nombre de documents récents et le nombre de documents anciens, qui sont numérisés petit à petit.

La réutilisation des données n'est autorisée que dans le cadre privé<sup>140</sup>. Les métadonnées peuvent être téléchargées au format PDF, tandis que les images sont librement téléchargeables en résolution moyenne. L'onglet *Ma sélection* permet de télécharger une liste d'œuvres au format PDF ou de commander les images en haute résolution.

### 2.3.3 Plateformes externes

La BGE participe à de nombreuses plateformes (voir 9.1.3 pour la liste complète), dont :

- E-codices : 122 manuscrits ont été versés par la BGE<sup>141</sup>, qui participe à cette plateforme depuis plusieurs années. Les numérisations sont effectuées en externe (institution agréée) et le catalogage par l'équipe de e-codices.

---

<sup>137</sup> Voir : <https://creativecommons.org/licenses/by-nc/2.0>

<sup>138</sup> « Conditions d'utilisation », Bibliothèque de Genève - Manuscrits et archives privées, consulté le 17 juillet 2022, <https://archives.bge-geneve.ch/n/conditions-d-utilisation/n:121>.

<sup>139</sup> Voir : <https://www.zetcom.com>

<sup>140</sup> « Conditions générales d'utilisation », Bibliothèque de Genève Iconographie, consulté le 19 mai 2022, <https://bge-geneve.ch/iconographie/conditions-generales-dutilisation>.

<sup>141</sup> Ressources de la BGE sur e-codices : [http://e-codices.ch/en/search/?iCurrentPage=1&sQueryString=&sSearchField=fullText&sSortField=score&aSelectedFacets=%5bcollection\\_facet%5d%5b%5d=Gen%C3%A8ve%2C+Biblioth%C3%A8que+de+Gen%C3%A8ve](http://e-codices.ch/en/search/?iCurrentPage=1&sQueryString=&sSearchField=fullText&sSortField=score&aSelectedFacets=%5bcollection_facet%5d%5b%5d=Gen%C3%A8ve%2C+Biblioth%C3%A8que+de+Gen%C3%A8ve)

- E-rara : près de 6'300 documents ont été déposés par la BGE<sup>142</sup> via des versements réguliers. Un processus automatique a été mis en place pour la récupération automatique des métadonnées issues de SLSP. Un travail de structuration effectué manuellement est par contre toujours requis. La numérisation des images est effectuée en interne et le transfert est automatisé.
- Catalogue Collectif Suisse des Affiches (CCSA) : près de 26'000 affiches ont été versées<sup>143</sup> de manière régulière. Le processus de versement est géré par la BN. La plupart des affiches étant sous droit, la BGE rend accessible publiquement les vignettes, tandis que les images en haute résolution apparaissent comme protégées.
- notreHistoire.ch<sup>144</sup> : environ 16'000 documents iconographiques libres de droit tels que des photographies et des illustrations ont été mis à disposition des utilisateurs.
- Onstage<sup>145</sup> : environ 2'300 ont été déposés en 2020 par La Musicale. La migration a été gérée par le RISM. Aucun autre versement n'est planifié pour le moment.
- Wikipédia : l'institution contribue à la création d'articles. À cet effet, elle organise de manière régulière des Wikiathons. De nombreux articles ont notamment été créés pour les personnalités genevoises.

Concernant la presse, les deux titres suivants ont été numérisés grâce à la BGE : le *Journal de Genève*, disponible sur la plateforme [letempsarchives.ch](http://letempsarchives.ch)<sup>146</sup>, et la *Tribune de Genève*, disponible sur [e-newspaperarchives](http://e-newspaperarchives)<sup>147</sup>. Si des magazines ou revues devaient être numérisés, ils seraient probablement mis à disposition sur [e-periodica](http://e-periodica)<sup>148</sup>.

La BGE n'envisage pour le moment pas de participer à des catalogues comme Europeana, Memobase ou Swisscollections<sup>149</sup> pour des raisons de ressources humaines et de coût financier.

## 2.4 Plateformes en liens avec les réseaux suisses de bibliothèques

### 2.4.1 Swisscollections

Swisscollections se présente comme une plateforme complémentaire à Swisscovery<sup>150</sup> (plateforme nationale des bibliothèques gérée par la société SLSP) en agréant des fonds spéciaux qui nécessitent des outils de recherche spécifiques.<sup>151</sup> On y trouve quelques

<sup>142</sup> Lien BGE sur E-rara : [https://www.e-rara.ch/nav/index?&facets=domainGroup%3D%22gep\\_g%22](https://www.e-rara.ch/nav/index?&facets=domainGroup%3D%22gep_g%22)

<sup>143</sup> Lien BGE sur CCSA :

[https://www.posters.nb.admin.ch/discovery/collectionDiscovery?vid=41SNL\\_53\\_INST:posters&collectionId=8133219970003978&lang=fr](https://www.posters.nb.admin.ch/discovery/collectionDiscovery?vid=41SNL_53_INST:posters&collectionId=8133219970003978&lang=fr)

<sup>144</sup> Voir : <https://notrehistoire.ch/@bge>

<sup>145</sup> Voir : <https://onstage.rism.digital/>

<sup>146</sup> Voir : <https://www.letempsarchives.ch/>

<sup>147</sup> Voir : <https://www.e-newspaperarchives.ch/?a=cl&cl=CL1&sp=TDG&l=fr>

<sup>148</sup> Voir : <https://www.e-periodica.ch>

<sup>149</sup> Voir : <https://swisscollections.ch>

<sup>150</sup> Voir : <https://swisscovery.slsp.ch>

<sup>151</sup> « À propos de swisscollections », swisscollections, consulté le 3 juin 2022, <https://swisscollections.ch/About>.

1'865'000 ressources telles que des fonds d'archives, des documents iconographiques, des livres rares ou anciens, des manuscrits, des cartes, des enregistrements musicaux, des films ou encore des partitions.<sup>152</sup> Lancée en décembre 2020, la plateforme est gérée par l'association Swisscollections, elle-même issue de l'association HAN créée en 2011. Elle regroupe actuellement les ressources de 14 bibliothèques et archives<sup>153</sup>.

Au niveau technique, la plateforme est basée sur les outils open source VuFind<sup>154</sup> pour l'interface utilisateur, Apache Solr<sup>155</sup> pour le moteur de recherche, Apache Kafka<sup>156</sup> pour la gestion du flux des données et Kubernetes<sup>157</sup> pour l'infrastructure.<sup>158</sup> Les notices des documents sont dans un format MARC enrichi d'informations complémentaires telles que la licence ou le descriptif du contenu et de la structure du document. La plateforme ne stocke pas de documents numériques mais pointe vers des plateformes externes comme e-codices. Les métadonnées sont accessibles via le protocole SRU. Le code source de la plateforme est accessible sur GitLab<sup>159</sup> et un wiki germanophone est disponible pour la documentation.<sup>160</sup>

## 2.4.2 Plateformes en lien avec RERO+

### 2.4.2.1 Sonar

Sonar<sup>161</sup> (Swiss Open Access Repository) est un système de gestion pour les documents numériques proposé par la fondation RERO+.<sup>162</sup> Ce système, promu par Swissuniversities, a initialement été conçu comme un projet d'agrégation et de préservation à long terme pour les publications scientifiques suisses en open access. Aujourd'hui, il s'est transformé en Institutional Repository as a Service (IRaaS), dont l'usage est plus large.<sup>163</sup>

Le portail partagé héberge plus de 316'000 documents dont environ 160'000 en open access. Ces documents sont très majoritairement des articles scientifiques et des travaux d'étudiants. Le système n'est actuellement pas utilisé pour des documents comme des fonds d'archives, des manuscrits ou des ressources iconographiques.

Le système permet de gérer des comptes usagers avec différents rôles. En outre, on peut noter les caractéristiques suivantes : les documents disposent d'un identifiant ARK ; il est possible de limiter l'accès aux documents en fonction d'une plage d'adresses IP ; les données sont accessibles et modifiables via des APIs RESTful complètes.<sup>164</sup> Les métadonnées peuvent également être disponibles via le protocole OAI-PMH.

---

<sup>152</sup> « À propos de swisscollections ».

<sup>153</sup> Voir la liste : <https://swisscollections.ch/Partners>

<sup>154</sup> Voir : <https://vufind.org/vufind>

<sup>155</sup> Voir : <https://solr.apache.org>

<sup>156</sup> Voir : <https://kafka.apache.org>

<sup>157</sup> Voir : <https://kubernetes.io>

<sup>158</sup> « À propos de swisscollections ».

<sup>159</sup> Interface utilisateurs : <https://gitlab.com/swissbib/classic/vufind> ; traitement des données : <https://gitlab.com/swissbib/swisscollections>

<sup>160</sup> Voir le wiki : <https://ub-basel.atlassian.net/wiki/spaces/swisscollections/overview>

<sup>161</sup> Voir : <https://sonar.rero.ch>

<sup>162</sup> Voir : <https://www.rero.ch>

<sup>163</sup> « Sonar IRaaS », RERO+, consulté le 18 juillet 2022, <https://www.rero.ch/produits/sonar#demo>.

<sup>164</sup> « Sonar », swissuniversities, consulté le 18 juillet 2022, <https://www.swissuniversities.ch/fr/p-5-services/sonar>.

En l'état actuel, des adaptations, telles que l'ajout de certains champs supplémentaires, seraient nécessaires pour le catalogage des fonds d'archives ou les manuscrits (et il serait évidemment utile de répercuter ces modifications dans les filtres de recherche). Par contre, le système possède déjà un module de préservation à long terme pour les documents numériques, ce qui le positionne favorablement pour héberger des documents relatifs au dépôt légal numérique par exemple. Concernant les types de documents, Sonar peut héberger tout type de fichier mais est actuellement optimisé pour le texte, soit des fichiers au format PDF, XML ou TXT.

#### 2.4.2.2 RERO doc et RERO ILS

RERO doc<sup>165</sup> était la bibliothèque numérique du réseau RERO (REseau ROmand) qui est devenu la Fondation RERO+. Encore active, la plateforme RERO doc abrite quelques 81'500 documents numériques, dont plus de 57'000 articles et une minorité de documents patrimoniaux tels que des cartes géographiques, des partitions, des enregistrements sonores et des documents iconographiques. Il est prévu de migrer ces documents sur le nouveau catalogue collectif,<sup>166</sup> lancé en juillet 2021, qui utilise le système de gestion RERO ILS.<sup>167</sup>

Cette décision a été prise afin de réunir sur un même catalogue tous les documents des bibliothèques affiliées, ainsi qu'afin d'éviter la création de notices distinctes pour les documents numériques et les documents physiques. En outre, le système RERO ILS présente l'avantage, par rapport à Sonar, de suivre le standard RDA qui évolue dans le sens du modèle IFLA LRM<sup>168</sup> (pour la structuration des données par entités).

## 2.5 Mise en perspective entre la BCUL et la BGE

Il est intéressant de constater que la BCUL comme la BGE ont pris le parti de développer leurs propres plateformes et non de participer à des plateformes collectives plus importantes. Dans le cas de la BGE, cela s'explique notamment par le fait qu'il s'agit d'une bibliothèque qui était dans le réseau RERO et est maintenant affiliée à SLSP. Elle ne faisait donc pas partie de l'association HAN. En outre, la conception d'une plateforme multibases pour les institutions de la ville de Genève s'inscrit dans son objectif de proposer un catalogue au niveau régional.

Du côté de la BCUL, Patrinum a émergé d'un triple besoin très spécifique : conservation à long terme, système de gestion pour les fonds spéciaux et mise à disposition du public. Or aucune autre plateforme ne répond à ces trois critères.

Le fait de chercher à répondre à ces trois critères est en soit un parti pris différent de celui de la BGE, qui s'explique par la mise en place d'un dépôt légal numérique à la BCUL, alors que le dépôt légal de la BGE concerne exclusivement le patrimoine imprimé.

La BGE a donc choisi l'option d'utiliser un système tiers pour la conservation à long terme des fichiers et de créer une plateforme avec pour objectif principal la mise à disposition du public. Un effort spécifique a ainsi été fourni sur cet aspect. En outre, la BGE a choisi de toujours utiliser des plateformes parfaitement adaptées au type de documents en termes de norme

---

<sup>165</sup> Voir : <https://doc.rero.ch>

<sup>166</sup> Voir : <https://bib.rero.ch>

<sup>167</sup> Voir : <https://www.rero.ch/produits/ils>

<sup>168</sup> Voir : <https://www.transition-bibliographique.fr/enjeux/definition-ifla-lrm>

pour les métadonnées, d'où le fait que les Manuscrits et archives ne sont pas sur la même plateforme que les documents iconographiques. Outre le respect des normes, cette solution permet de bénéficier d'outils de recherche et de navigation sur le catalogue public qui sont directement adaptés aux métadonnées (affichage de l'arborescence et filtres de recherche spécifiques), alors que cela fait l'objet de développements a posteriori pour Patrinum.

La solution Patrinum, de son côté, évite de doubler les plateformes pour la préservation à long terme et la mise à disposition du public. De plus, elle englobe tous les types de documents, étant également adaptée aux ressources multimédia. Elle est ainsi en adéquation avec les besoins de la BCUL qui possède beaucoup de collections avec des types de documents variés dont la numérisation progresse en parallèle, selon les projets de recherche. A l'inverse, la BGE a fourni un gros fort pour la numérisation des fonds de Manuscrits et archives (environ 20'000 documents accessibles en ligne) ainsi que des fonds iconographiques (près de 43'000 documents accessibles en ligne). La BCUL, de son côté, propose moins d'une centaine de documents numérisés sur Patrinum pour les Manuscrits et archives, et environ mille documents pour l'Iconopôle.

Par contre, du côté des imprimés anciens et de la presse numérisée, la BCUL a fourni très tôt un effort important qui s'est concrétisé par la mise à disposition de 100'000 ouvrages sur Google Books et le lancement de Scriptorium en 2012. Quant à la BGE, si elle propose plus de 6'000 imprimés sur e-rara, seuls deux quotidiens sont accessibles en ligne en ce qui concerne la presse.

Du côté des plateformes externes, on voit que les deux institutions participent souvent aux mêmes catalogues, dont e-codices, e-rara, CCSA ou encore à egodocuments et Lumières.Lausanne. La BCUL a plus tendance à effectuer des versements uniques, ce qui suit la logique de la priorisation des numérisations par projet, alors que la BGE a pu mettre en place des versements réguliers, entre autres, sur e-codices, e-rara et CCSA. Le nombre de documents que la BGE met à disposition sur ces trois plateformes est également bien plus importants, ce qui s'explique notamment par l'arrivée récente de la BCUL sur ces plateformes.

En outre, on peut noter la participation de la BGE à notreHistoire.ch ainsi qu'à Onstage, tandis que la BCUL prévoit de participer à Memobase ainsi qu'à notreHistoire.ch dans un futur proche.

Pour la BCUL, la plus importante participation à une plateforme externe est de très loin Google Books. Dans une moindre mesure, on peut mentionner les 800 articles de Wikipédia issus des notices d'autorités vaudoises ainsi que le millier de documents versés au CCSA.

De manière générale, il serait intéressant pour la BCUL de renforcer sa présence sur les catalogues externes déjà utilisés en effectuant des versements réguliers.

## 3 Le domaine de la musique écrite

### 3.1 Domaine choisi

Pour les raisons mentionnées en introduction (1.4), cette section se concentre sur les projets en lien avec la musicologie et plus particulièrement nécessitant des corpus numériques de partitions (2.2.1).

### 3.2 L'IA dans le domaine musical

Dès les années 1970, un lien fort se crée entre technologies et musique qui annonce l'ouverture des humanités numériques au domaine de la musique.<sup>169</sup> C'est par exemple à cette époque que la station de travail musicale 4X est inventée.<sup>170</sup> Capable d'analyser et de synthétiser de nouvelles sonorités en temps réel, elle sera utilisée dès 1981 par Pierre Boulez.<sup>171</sup> Comme le montre cette invention, on commence dès cette époque à développer et programmer des machines dont l'objectif est de reproduire les capacités cognitives du musicien.<sup>172</sup> Ces recherches se subdivisent en deux axes, l'un étant le développement d'outils qui assistent ou imitent le musicien dans le cadre de sa performance sur scène (temps réel), tandis que l'autre s'attache à faciliter le travail du compositeur (temps différé), par exemple en fournissant des outils logiciels permettant d'automatiser certaines tâches relatives à la composition (naissance de la Composition Assistée par Ordinateur, CAO).<sup>173</sup>

Développé durant cette période par l'ircam (Institut de Recherche et Coordination Acoustique/Musique), le logiciel Max<sup>174</sup> (ou Pure data<sup>175</sup> pour sa variante open source) permet de capter les signaux, produits par des instruments musicaux spécialement conçus, et de déclencher des processus sonores définis par rapport à la partition.<sup>176</sup> L'avènement de ce type d'outil va nécessiter de résoudre la problématique du suivi de partitions, la difficulté pour la machine étant de comprendre quand un processus doit être déclenché alors que les musiciens jouent avec des variations, par exemple au niveau du rythme.<sup>177</sup>

En 2007, naît une nouvelle génération de systèmes de suivi de partition en temps réel.<sup>178</sup> Antescofo<sup>179</sup>, un module de Max et Pure data, fonctionne à l'aide d'une machine d'écoute capable d'analyser le signal acoustique produit par des instruments standards.<sup>180</sup> Cette nouvelle génération d'outils est rendue possible par plusieurs facteurs dont : l'avènement du

---

<sup>169</sup> Gérard Assayag, *Évolution musicale : entre art et technologies (Recherche musicale & intelligence digitale)* | partie 3, entretien réalisé par Dominique Moulon, MOOC Digital Paris, décembre 2017, <https://moocdigital.paris/cours/recherche-musicale-intelligence-digitale/evolution-musicale-entre-art-technologies>.

<sup>170</sup> « Sons Dessus Dessous #3 - La station d'informatique musicale 4X », 3 décembre 2017, <https://www.ircam.fr/article/detail/sons-dessus-dessous-3-la-station-dinformatique-musicale-4x/>.

<sup>171</sup> « Sons Dessus Dessous #3 - La station d'informatique musicale 4X ».

<sup>172</sup> Assayag, *Évolution musicale*.

<sup>173</sup> Ibid.

<sup>174</sup> Voir : <https://cycling74.com/products/max>

<sup>175</sup> Voir : <http://puredata.info>

<sup>176</sup> Assayag, *Évolution musicale*.

<sup>177</sup> Ibid.

<sup>178</sup> Ibid.

<sup>179</sup> Voir : <https://forum.ircam.fr/projects/detail/antESCOFO>

<sup>180</sup> Assayag, *Évolution musicale*.

format MP3 (dès les années 1990), qui facilite l'échange de fichiers audio par internet et donc rend possible la création de grandes bases de données musicales en ligne ; et l'application des algorithmes de machine learning à ces grandes bases afin d'en extraire des informations structurelles de haut niveau ; c'est le début du MIR (Music Information Retrieval).<sup>181</sup>

Aujourd'hui, une troisième génération d'outils voit le jour. A ce titre, On peut mentionner SoMax<sup>182</sup>, capables d'improviser en temps réel avec les musiciens de manière non supervisée.<sup>183</sup> Mais les sujets de recherche actuels sont en réalité d'une grande variété, parmi lesquels on peut évoquer les thématiques suivantes :

- La modélisation du geste corporel dans sa relation au son. On peut mentionner la captation du geste du musicien et son association au son produit. Dans le cadre de spectacles (concert, danse, etc.), c'est la captation du geste de l'artiste qui peut déclencher des effets visuels ou sonores.<sup>184</sup> On peut également mentionner l'usage du geste comme vecteur de l'écriture permettant de composer en temps réel des structures sonores.<sup>185</sup> L'expérimentation de Google Semi-Conductor<sup>186</sup> reprend cette thématique en permettant de diriger un orchestre (tempo, volume et instruments) en bougeant les bras devant une webcam.
- La spatialisation du son : il s'agit notamment de donner au son une acoustique virtuelle en fonction de descripteurs objectifs ou subjectif (son brillant, mat, etc.).<sup>187</sup>
- La conception de nouveaux types d'instruments électroniques : il s'agit d'équiper des instruments acoustiques standards de capteurs et actionneurs mécaniques capable de transformer le comportement mécanique et la vibration de l'instrument.<sup>188</sup> La nouveauté est ici de transformer le son non pas par l'intermédiaire de haut-parleurs, mais directement sur l'instrument afin de conserver la diffusion sonore d'une performance acoustique.<sup>189</sup>
- La synthèse de sons, notamment celle de la voix<sup>190</sup>. Ces recherches permettent par exemple de transformer en temps réel la nature de la voix en conservant une sonorité naturelle et non synthétique.<sup>191</sup> Dans le champ de la synthèse sonore, on peut

---

<sup>181</sup> Gérard Assayag, Du signal temps réel à l'intelligence digitale (Recherche musicale & intelligence digitale) | partie 4, entretien réalisé par Dominique Moulon, MOOC Digital Paris, décembre 2017, <https://moocdigital.paris/cours/recherche-musicale-intelligence-digitale/signal-temps-reel-lintelligence-digitale>.

<sup>182</sup> Voir : <https://www.ircam.fr/projects/pages/omax>

<sup>183</sup> Gérard Assayag, Entre espaces acoustiques et cognitifs (Recherche musicale & intelligence digitale) | Partie 5, entretien réalisé par Dominique Moulon, MOOC Digital Paris, décembre 2017, <https://moocdigital.paris/cours/recherche-musicale-intelligence-digitale/entre-espaces-acoustiques-cognitifs>.

<sup>184</sup> « Analyse et reconnaissance du geste », ircam, consulté le 28 juillet 2022, <https://www.ircam.fr/projects/pages/analyse-et-reconnaissance-du-geste>.

<sup>185</sup> Gérard Assayag, Geste, voix et corporéité (Recherche musicale & intelligence digitale) | partie 6, entretien réalisé par Dominique Moulon, MOOC Digital Paris, consulté le 26 juillet 2022, <https://moocdigital.paris/cours/recherche-musicale-intelligence-digitale/geste-voix-corporeite>.

<sup>186</sup> Google Creative Lab, « Semi-Conductor », Experiments with Google, décembre 2018, <https://experiments.withgoogle.com/semi-conductor>.

<sup>187</sup> Assayag, Entre espaces acoustiques et cognitifs (Recherche musicale & intelligence digitale) | Partie 5.

<sup>188</sup> Assayag, Geste, voix et corporéité (Recherche musicale & intelligence digitale) | partie 6.

<sup>189</sup> Ibid.

<sup>190</sup> « Analyse et synthèse des sons », ircam, consulté le 28 juillet 2022, <https://www.ircam.fr/recherche/equipes-recherche/anasy>.

<sup>191</sup> Assayag, Geste, voix et corporéité (Recherche musicale & intelligence digitale) | partie 6.

également mentionner Blob opera<sup>192</sup>, une autre expérimentation de Google permettant de faire chanter de manière interactive des voix d'opéra qui s'harmonisent entre elles automatiquement.

- La classification de la musique pour améliorer les recommandations faites aux usagers des grandes bibliothèques musicales.<sup>193</sup>
- La transcription automatique des enregistrements audios en partitions. On peut mentionner les auteurs Shibata et Nakamura, qui proposent une nouvelle méthode basée sur l'utilisation de réseaux de neurones profonds et la quantification du rythme basée sur un modèle statistique.<sup>194</sup> Appliqués à des musiques interprétées au piano, les résultats de cette méthode sont très satisfaisants<sup>195</sup> et ouvrent la voie à des applications permettant une transcription semi-automatisée (corrigée par l'humain lorsque le taux de correspondance est moyen). Les trois bases de données utilisées dans cette étude pour l'entraînement de l'IA sont : MIDI Aligned Piano Sounds database (MAPS), MuseScore et print-gakufu.<sup>196</sup>
- Le suivi de partition qui consiste à aligner un enregistrement audio avec une partition en mettant en évidence les notes jouées (3.3.7).

Comme on le constate, les axes de recherche actuels s'articulent principalement autour de la création et de la modification des sons, ainsi que de l'aide à la composition. Toutefois, la musique écrite continue à servir de référentiel lorsqu'il s'agit d'accompagner la performance d'un musicien, il y a donc tout un pan de recherche en lien avec l'association du son et de la notation musicale (3.3.7). Un autre domaine de recherche qui peut être exploré à partir d'enregistrements sonores ou de la musique écrite, est l'analyse de la structure des morceaux, notamment au niveau de l'harmonie, du rythme, etc.. Cela peut par exemple permettre d'identifier des schémas structurels spécifiques à un style de musique ou à un compositeur (3.4.3). On peut imaginer différentes applications concrètes en lien avec ce type d'analyses, comme la possibilité de transformer une mélodie pour correspondre à tel style musical ou tel compositeur. Ce type d'application existe d'ailleurs depuis longtemps dans le cadre des arts visuels (par exemple pour appliquer le style d'une peinture de Van Gogh à une photographie contemporaine). Une autre expérimentation de Google, Assisted Melody,<sup>197</sup> s'inscrit dans ce cadre en permettant d'enrichir une mélodie simple, créée par l'utilisateur, avec des harmonies dans le style de Bach, Beethoven ou Mozart.

Cet exemple montre bien que l'application de l'intelligence artificielle dans le domaine de la musique se développe, mais reste moins avancée que dans celui de l'image. Selon Gérard Assayag, qui dirige l'équipe Représentations musicales à l'ircam, la notion de temporalité est probablement l'une des causes majeures à la difficulté de ce développement.<sup>198</sup> Toutefois,

---

<sup>192</sup> « Blob Opera - On Tour! by David Li in collaboration with Google Arts & Culture », Experiments with Google, mai 2021, <https://experiments.withgoogle.com/blob-opera-on-tour>.

<sup>193</sup> Gail Bishop, « Music Datasets for Machine Learning », *Medium* (blog), 12 mars 2021, part. Classification, <https://gail-bishop.medium.com/music-datasets-for-machine-learning-a6cd8d707340>.

<sup>194</sup> Kentaro Shibata, Eita Nakamura, et Kazuyoshi Yoshii, « Non-Local Musical Statistics as Guides for Audio-to-Score Piano Transcription », *Information Sciences* 566 (août 2021): 1, <https://doi.org/10.1016/j.ins.2021.03.014>.

<sup>195</sup> Shibata, Nakamura, et Yoshii, 14.

<sup>196</sup> Voir : MAPS, <https://adasp.telecom-paris.fr/resources/2010-07-08-maps-database>; MuseScore, <https://musescore.org>; print-gakufu, <https://www.print-gakufu.com>.

<sup>197</sup> « Assisted Melody by Simon Doury, Artist in Residence at Google Arts & Culture Lab with Google Magenta », Experiments with Google, juin 2021, <https://experiments.withgoogle.com/assisted-melody>.

<sup>198</sup> Assayag, Du signal temps réel à l'intelligence digitale (Recherche musicale & intelligence digitale) | partie 4.

l'avènement du deep learning et l'augmentation rapide du nombre de bases musicales disponibles en ligne offrent de nouvelles possibilités.

Aujourd'hui, les derniers systèmes d'IA dans les arts visuels permettent de créer des images selon différents médiums (peinture, 3D, photographie, etc.) à partir de termes en langage naturel.<sup>199</sup> Si l'équivalent était possible dans la sphère musicale, on pourrait composer une œuvre à partir de mots ; mais on constate immédiatement que la tâche est plus complexe. En effet, d'une part, il est nécessaire de maîtriser cette notion de temporalité. D'autre part, il est nécessaire de trouver des mots pour décrire une musique, ce qui est un exercice peu habituel. Toutefois, on peut imaginer qu'il sera possible dans un horizon pas si lointain d'aller jusqu'à la conception de systèmes IA capables de composer des pièces que l'on puisse véritablement confondre avec des compositions humaines à partir de consignes simples, telles que : le style musical, éventuellement des compositeurs ou interprètes, des instruments, des descripteurs d'ambiance (épique, dansant, ...), etc.. A ce titre, on peut d'ailleurs mentionner le projet Flow Machines<sup>200</sup> de Sony, qui, grâce à un réseau de neurones entraînés sur 13'000 chansons pop, a notamment créé *Daddy's Car*, une chanson dans le style des Beatles entièrement conçue par IA.<sup>201</sup>

### 3.3 Normes et technologiques

Plusieurs standards et formats spécifiques ont été développés pour la musique écrite. On peut mentionner le standard MEI, utilisable autant pour les métadonnées que pour l'encodage des fichiers eux-mêmes, et le format MusicXML, utilisable pour les fichiers uniquement. La reconnaissance OCR est remplacée par la reconnaissance OMR (Optical Music Recognition), qui ouvre des perspectives intéressantes pour l'analyse des documents numérisés. Il faut également mentionner les APIs IIF a/v, qui sont disponibles depuis 2020 et ont permis d'étendre les fonctionnalités IIF à différents usages pour les ressources audio-visuelles. Enfin, le suivi de partition est un domaine de recherche actuel qui relie plusieurs des éléments précédemment évoqués.

#### 3.3.1 MEI

Très similaires à TEI (Text Encoding Initiative), MEI (Music Encoding Initiative)<sup>202</sup> est un système d'encodage à destination des partitions musicales afin de rendre leur structure lisible par la machine.<sup>203</sup> Il s'agit d'un format open-source reposant sur le langage XML, créé dans le but de devenir un standard unicode international pour les partitions.<sup>204</sup> MEI est structuré par module dont chacun est dédié à un usage applicatif spécifique.<sup>205</sup> Les deux modules les plus importants sont le module partagé et le module d'en-tête, qui définit et groupe toutes les métadonnées.<sup>206</sup> En outre, ce module permet également de mettre en œuvre le modèle FRBR

---

<sup>199</sup> Voir : <https://openai.com/dall-e-2>

<sup>200</sup> Voir : <http://www.flow-machines.com>

<sup>201</sup> Blanca Sanjuanbenito, « Artificial Intelligence Brings The Beatles and Bach Back to Life », *BBVA* (blog), 01 2020, <https://www.bbva.com/en/artificial-intelligence-resuscitates-beatles-bach/>.

<sup>202</sup> Voir : site officiel, <https://music-encoding.org> ; code source, <https://github.com/music-encoding>.

<sup>203</sup> Akademie der Wissenschaften und der Literatur Mainz, « What is MEI? », MEI, 2022, part. An introduction to MEI, <https://music-encoding.org/about/>.

<sup>204</sup> *De-Re-M.E.I. : A Look at the Music Encoding Initiative Schema*, 2022, <https://youtu.be/GBcMEjogCxl>.

<sup>205</sup> Keil et Pugin, « Das Internationale Quellenlexikon der Musik, RISM », part. 6.

<sup>206</sup> *Ibid.*, part. 6.

ou d'encapsuler des incipits musicaux dans différents formats.<sup>207</sup> MEI permet également de décrire les métadonnées des documents dans un format structuré ainsi que de décrire leurs relations avec le document numérique ou des ressources liées, comme un enregistrement sonore.<sup>208</sup> De plus, il permet aussi de décrire les partitions anciennes, comme les partitions médiévales ou de la Renaissance.<sup>209</sup> Verena Monier, responsable des Archives musicales, précise au sujet de MEI que le standard n'est malheureusement pas encore suffisamment abouti pour qu'une description formelle soit utilisée au niveau international, d'où le choix d'utiliser plutôt le format MusicXML à la BCUL pour faciliter la réutilisation des fichiers stockés sur Patrinum.

### 3.3.2 MIDI

Musical Instrument Digital Interface (MIDI) est un protocole de communication et un format de fichier introduit dans les années 1980.<sup>210</sup> Le protocole MIDI décrit les interfaces et les connecteurs qui relient les instruments électroniques aux ordinateurs et autres appareils audio.<sup>211</sup> Un fichier MIDI repose sur le Standard MIDI File (SMF) qui, proche d'une partition numérique, formalise les informations telles que la hauteur des notes, leur début, leur fin, leur durée, et leur volume. Par contre, les informations, telles que le phrasé des notes, le doigté, le coup d'archet, etc., sont absentes, ce qui ne permet pas au SMF de remplacer le format MEI ou MusicXML.<sup>212</sup> Le MIDI présente par contre l'avantage de pouvoir facilement être converti en audio ou en partition à l'aide de nombreux logiciels de musique.<sup>213</sup> Cette technologie permet donc depuis longtemps de suivre une partition en générant de l'audio, mais il s'agit de sons de synthèses exécutés de manière automatique, soit loin d'égaliser la musicalité ou l'expressivité d'une interprétation humaine.<sup>214</sup> Enfin, on peut noter que le MIDI est un format de fichiers très léger et facile à traiter qui est donc bien adapté pour la recherche MIR.<sup>215</sup>

### 3.3.3 MusicXML

MEI, tout comme MusicXML, est un format ouvert permettant d'encoder les notations musicales (notes, portées, clefs, etc.) en se basant sur le langage XML. Ces deux formats sont suffisamment développés pour intégrer toutes les informations présentes dans une partition imprimée, contrairement au MIDI. Par contre, MusicXML et MEI poursuivent des objectifs différents : MusicXML est avant tout un format d'échange de fichiers,<sup>216</sup> alors que MEI devrait devenir un standard international. En outre, MEI possède la capacité de décrire les métadonnées en plus des fichiers, ce qui n'est pas le cas de MusicXML. Toutefois, le format le plus pérenne pour le stockage des partitions numériques est MusicXML, car MEI peine à

---

<sup>207</sup> Ibid.

<sup>208</sup> Akademie der Wissenschaften und der Literatur Mainz, « What is MEI? », part. How is MEI different from MusicXML?

<sup>209</sup> Ibid.

<sup>210</sup> « Musical Instrument Digital Interface », in *Wikipédia*, 31 mars 2022,

[https://fr.wikipedia.org/w/index.php?title=Musical\\_Instrument\\_Digital\\_Interface&oldid=192442209](https://fr.wikipedia.org/w/index.php?title=Musical_Instrument_Digital_Interface&oldid=192442209).

<sup>211</sup> Lin Hao Lee, « Musical Score Following and Audio Alignment » (arXiv, 6 mai 2022), 27, <http://arxiv.org/abs/2205.03247>.

<sup>212</sup> Lee, 27.

<sup>213</sup> Ibid.

<sup>214</sup> Ibid., 28.

<sup>215</sup> Ibid., 27.

<sup>216</sup> « Introduction à MusicXML – La partition numérique », part. MusicXML, qu'est-ce que c'est ?, consulté le 19 juin 2022, <https://www.partitionnumerique.com/technos/musicxml/>.

s'imposer au niveau international.<sup>217</sup> MusicXML a l'avantage de garantir la compatibilité du fichier avec la plupart des logiciels d'édition et de lecture de partition.<sup>218</sup> En outre, MusicXML est un sur-ensemble de MIDI dans le sens où les fichiers peuvent aussi être interprétés dans des logiciels de musique.<sup>219</sup> MusicXML peut donc être facilement synthétisé en audio, extrait en MIDI et converti en partitions, comme le montre la Figure 1.<sup>220</sup>

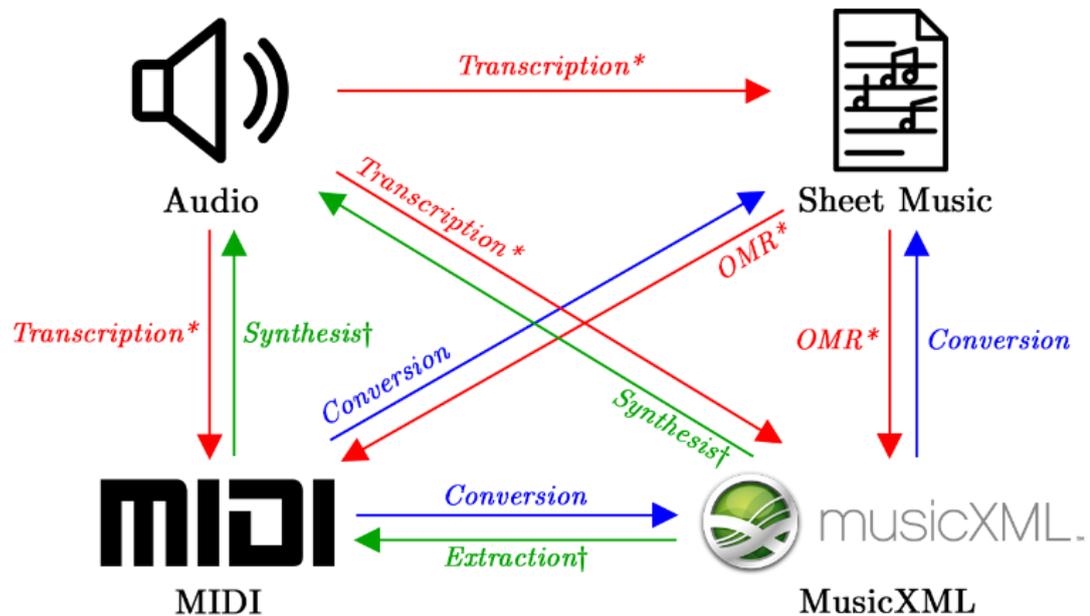


Figure 1: Illustration des méthodes de conversion courantes entre audio, partitions, MIDI et MusicXML. Les liens rouges (marqués d'un \*) indiquent des conversions avec pertes et difficiles à effectuer ; les liens verts (marqués d'un †) indiquent des conversions sans perte et également faciles ; les liens bleus (non marqués) indiquent des conversions avec perte et faciles. Source : Lin Hao Lee, « Musical Score Following and Audio Alignment » (arXiv, 6 mai 2022), 27.

### 3.3.4 Incipits musicaux

L'idée utiliser les incipits musicaux à des fins de recherche, par exemple pour faciliter l'identification de documents dont les sources sont anonymes ou incertaines, remonte aux années 1980.<sup>221</sup> L'intégration de la recherche par incipit dans RISM est donc un élément qui a été mis en place dès les débuts du catalogue (un champ MARC est d'ailleurs spécifiquement dédié à ce type d'informations) et qui a été grandement amélioré depuis la mise en place du plugin open source Verovio<sup>222,223</sup> Klaus Keil et Laurent Pugin mentionnent plusieurs algorithmes de recherche dédiés aux incipits, mais ils précisent que la pertinence des résultats dépend des besoins des utilisateurs et qu'il serait utile d'offrir des interfaces de recherche flexibles (recherche par exemple de similitudes rythmiques, d'intervalle, etc.).<sup>224</sup>

<sup>217</sup> Lee, « Musical Score Following and Audio Alignment », 29.

<sup>218</sup> « MusicXML for Exchanging Digital Sheet Music », makemusic, consulté le 2 août 2022, <https://www.musicxml.com/>.

<sup>219</sup> Lee, « Musical Score Following and Audio Alignment », 28.

<sup>220</sup> Ibid.

<sup>221</sup> Keil et Pugin, « Das Internationale Quellenlexikon der Musik, RISM », part. 7.

<sup>222</sup> Voir : <https://www.verovio.org>

<sup>223</sup> Keil et Pugin, « Das Internationale Quellenlexikon der Musik, RISM », part. 7.

<sup>224</sup> Ibid.

Du côté de la BCUL, les incipits ne sont pas renseignés dans les notices et Patrinum n'offre pas de possibilité de recherche pour ce type d'information. L'intérêt d'ajouter ou non de telles fonctionnalités reste à déterminer, mais on peut toutefois noter qu'il existe une librairie Python incipit<sup>225</sup> permettant d'extraire automatiquement ces éléments à partir de documents numérisés sous format PDF ou image (une post-correction manuelle est probablement nécessaire).

### 3.3.5 APIs IIF audio et vidéo

IIF est un ensemble de normes ouvertes qui permet de mettre à disposition des images de manières interopérable sur le web ; c'est aussi la communauté internationale qui conçoit ces normes en développant les APIs IIF.<sup>226</sup> Les spécifications IIF s'alignent sur les normes générales du Web générales.<sup>227</sup> Les documents sont mis à disposition sur les serveurs, accompagnés d'un manifeste qui répertorient les informations telles que : comment afficher l'objet lui-même et quelles informations l'accompagnent (dont les métadonnées ou les informations sur la structure, telles que l'ordre des pages pour un livre).<sup>228</sup> La manière d'afficher l'image (zoom, rotation, niveaux de gris ou couleur) est renseignée dans les paramètres de l'URL désignant l'image,<sup>229</sup> ce qui permet à n'importe quelle visionneuse d'image compatible (Mirador, Universal Viewer, Diva.js) de l'afficher sans qu'il soit nécessaire de la copier.<sup>230</sup>

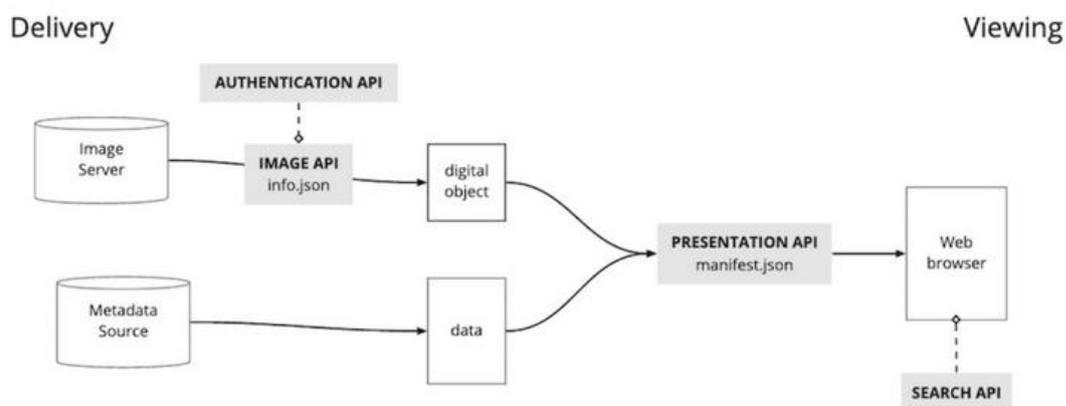


Figure 2: Diagramme du fonctionnement des APIs principales de IIF, Source : « How It Works - Introduction », IIF, <https://iif.io/get-started/how-iif-works/>.

En 2016 un groupe de travail dédié aux spécifications techniques en lien avec les ressources audios/vidéos (a/v) a été créé et un ensemble de fonctionnalités a été intégré dans la version 3.0<sup>231</sup> des API IIF.<sup>232</sup> Il s'agit notamment de permettre d'annoter et afficher des annotations

<sup>225</sup> Voir : <https://libraries.io/pypi/incipit>

<sup>226</sup> « Home », IIF, consulté le 5 août 2022, <https://iif.io/>.

<sup>227</sup> « How It Works », IIF, consulté le 5 août 2022, <https://iif.io/get-started/how-iif-works/>.

<sup>228</sup> Ibid.

<sup>229</sup> Ibid.

<sup>230</sup> IIF-C. Staff, « How to Use IIF Resources and Image Viewers », IIF Resource Guides, 7 avril 2020, [https://iif.io/guides/using\\_iif\\_resources/](https://iif.io/guides/using_iif_resources/).

<sup>231</sup> Voir : <https://iif.io/api/presentation/3.0>

<sup>232</sup> « IIF A/V Technical Specification Group », IIF, consulté le 1 août 2022, <https://iif.io/community/groups/av/av-tsg/>.

sur un fichier a/v, de lire plusieurs ressources a/v synchronisées en même temps, de mettre en surbrillance ou de spécifier une région ou un point d'intérêt dans un fichier a/v, et de jouer un fichier a/v en mettant en surbrillance une région dans une image à un temps donné ; ceci afin de pouvoir lire un morceau de musique tout en utilisant la surbrillance pour suivre les images de la partition musicale correspondante.<sup>233</sup>

### 3.3.6 OMR

#### 3.3.6.1 Fonctionnement et théorie

L'Optical Music Recognition (OMR) est définissable comme le domaine de recherche qui étudie comment la notation musicale peut être reconnue et interprétée par la machine.<sup>234</sup> Proche de l'OCR, dont elle est inspirée, l'OMR implique non seulement la reconnaissance des symboles représentés sur la partition musicale, mais également une interprétation sémantique de ces symboles. Les informations à identifier en lien avec les notes sont principalement la hauteur et la durée, mais de nombreuses informations complémentaires peuvent également être présentes, telles que le volume, le mode d'exécution (par exemple legato) ou d'autres annotations comme les marques d'archet, les doigtés, etc.. À cela s'ajoute encore les informations contextuelles, comme la clef ou la cadence.<sup>235</sup>

Différentes conventions d'écritures musicales existent, dont la plus courante est la Common Western Music Notation (CWMN). Il existe toutefois également différents systèmes historiques, comme la notation mensurale (utilisée en Europe entre le 13<sup>e</sup> et le 16<sup>e</sup> siècle) ou d'autres systèmes de notation actuels, spécifiques à un type d'instrument, par exemple les tablatures pour les instruments à cordes. En outre, il existe différentes taxonomies à l'intérieur d'un même système. Ainsi, Calvo-Zaragoza, Hajič et Pacha proposent de classer les partitions utilisant la CWMN comme suit : monophoniques (une mélodie simple sur une portée), homophonique (une voix principale mais composée de plusieurs notes simultanées sur une portée), polyphonique (plusieurs voix sur une seule portée), et pianoform (plusieurs voix et plusieurs portées).<sup>236</sup>

Comme on le voit, les notations qui peuvent être soumises aux systèmes OMR varient beaucoup et un système donné ne peut évidemment pas identifier tous les types de notation. A l'heure actuelle, la plupart des applications OMR sont conçues pour identifier la notation CWMN, tandis que quelques travaux portent sur les notations historiques comme la mensurale<sup>237</sup> ou les Neumes<sup>238</sup>.

Les systèmes OMR divisent généralement le traitement en quatre étapes : le prétraitement (amélioration du contraste, suppression du bruit, etc.), la détection des symboles (recherche

---

<sup>233</sup> Sheila Rabun, « IIF Update: Interoperability for Audio & Video Resources », CLIR, 17 août 2017, <https://rethinking.clir.org/blog/iif-update-interoperability-for-audio-video-resources/>.

<sup>234</sup> Jorge Calvo-Zaragoza, Jan Hajič Jr., et Alexander Pacha, « Understanding Optical Music Recognition », *ACM Computing Surveys* 53, n° 4 (31 juillet 2021): 3, <https://doi.org/10.1145/3397499>.

<sup>235</sup> Calvo-Zaragoza, Hajič Jr., et Pacha, 4.

<sup>236</sup> Ibid., 13.

<sup>237</sup> Jorge Calvo-Zaragoza, Alejandro H. Toselli, et Enrique Vidal, « Handwritten Music Recognition for Mensural Notation with Convolutional Recurrent Neural Networks », *Pattern Recognition Letters* 128 (1 décembre 2019): 115-21, <https://doi.org/10.1016/j.patrec.2019.08.021>.

<sup>238</sup> L.W.G. Barton, « The NEUMES Project: digital transcription of medieval chant manuscripts », in *Second International Conference on Web Delivering of Music, 2002. WEDELMUSIC 2002. Proceedings.*, 2002, 211-18, <https://doi.org/10.1109/WDM.2002.1176213>.

et classification), l'assemblage des symboles et l'ajout d'une couche de compréhension sémantique, l'encodage dans un format de sortie (par exemple MIDI pour la lecture ou MusicXML/MEI pour l'édition ultérieure).<sup>239</sup>

L'arrivée des systèmes basés sur le deep learning a toutefois permis de supprimer certaines étapes comme le pré-traitement, ou d'en simplifier d'autres (classification associée à la segmentation).<sup>240</sup> Tous les problèmes techniques ne sont néanmoins pas résolus.<sup>241</sup>

- La mauvaise détection des symboles : une solution consiste à travailler avec des images de meilleure qualité ou à faire intervenir une vérification humaine lorsque le système a un taux de confiance moyen dans la reconnaissance du symbole.
- La reconstruction sémantique : cette étape est traitée avec des méthodes qui diffèrent selon les recherches et peu généralisables, car elles ne couvrent souvent pas tous les aspects de la notation.

On peut également mentionner quelques difficultés qui ne sont pas directement liées aux systèmes OMR, à savoir<sup>242</sup> :

- La petite taille de la communauté qui étudie ces questions et le manque de coordination entre les travaux de recherche.
- L'absence de définition d'un format de sortie, ce qui rend difficile de comparer les méthodes entre elles et donc d'évaluer leur efficacité.

Un dernier point à soulever est que le nombre de partitions numérisées reste faible en comparaison des textes numérisés, ce qui ne permet pas d'entraîner les systèmes OMR (notamment les réseaux de neurones) avec autant d'efficacité que l'ont été les systèmes l'OCR. Toutefois, on peut mentionner le catalogue MuseScore (3.5), qui, avec ses quelques 1'500'000 partitions lisibles par la machine,<sup>243</sup> améliore grandement la situation.

### 3.3.6.2 Logiciels et mise en pratique

Malgré ces difficultés, on trouve aujourd'hui plusieurs systèmes OMR autant propriétaires qu'open source. Parmi ces derniers, on peut citer OpenOMR<sup>244</sup>, Aruspix<sup>245</sup> (conçu pour les imprimés anciens) ou Audiveris<sup>246</sup> ; ce dernier a l'avantage d'être conçu pour permettre à l'utilisateur de corriger directement les erreurs de détections OMR. OpenOMR et Audiveris sont applicables aux partitions CWMN uniquement, mais on trouve d'autres systèmes très spécifiques : par exemple destinés aux tablatures pour luths, à la notation byzantine, ou aux neumes.<sup>247</sup>

---

<sup>239</sup> Calvo-Zaragoza, Hajič Jr., et Pacha, « Understanding Optical Music Recognition », 22-23.

<sup>240</sup> Ibid., 22-23.

<sup>241</sup> Ibid., 26-27.

<sup>242</sup> Ibid., 26.

<sup>243</sup> Seulement 340'000 documents sont dans le domaine public, le reste est disponible sur abonnement payant, l'autorisation d'exploitation pour des systèmes IA reste donc à vérifier.

<sup>244</sup> Voir : <https://sourceforge.net/projects/openomr>

<sup>245</sup> Voir : <http://www.aruspix.net>

<sup>246</sup> Voir : <https://github.com/Audiveris/audiveris>

<sup>247</sup> Jan Hajič Jr., « Optical Music Recognition Using Deep Learning » (Prague: Charles University, 2020), 21, [https://ufal.mff.cuni.cz/~hajicj/2020/docs/GAUK1444217\\_techreport.pdf](https://ufal.mff.cuni.cz/~hajicj/2020/docs/GAUK1444217_techreport.pdf).

L'étude récente « OMR-assisted transcription: a case study with early prints » compare le temps passé pour la transcription de partitions en notation mensurale en utilisant, d'un côté, le temps passé à effectuer des transcriptions manuelles en employant différents logiciels de transcription, et, de l'autre côté, le temps passé à produire des transcriptions en utilisant des logiciels OMR incluant une post-correction manuelle. Les logiciels utilisés pour la correction OMR assistée sont Aruspix<sup>248</sup> (créé spécifiquement pour les partitions des 16<sup>e</sup> et 17<sup>e</sup> siècles) et MurRET<sup>249</sup> (qui nécessite de créer et entraîner son propre modèle). Le résultat montre que l'usage d'Aruspix est clairement la solution la plus efficace, tandis que l'usage de MuRET donne un résultat mitigé.<sup>250</sup>

De manière générale, cette étude fait bien ressortir la plus-value que les systèmes OMR peuvent déjà apporter en complémentarité avec une correction manuelle, y compris pour la reconnaissance de documents utilisant une notation non standard. En outre, comme un utilisateur doit de toute manière intervenir pour la correction finale, il serait utile de lui permettre de corriger les erreurs dès le moment de la détection des caractères, ce qui pourrait améliorer directement le système en lui faisant apprendre de ses erreurs.<sup>251</sup> A ce titre, on peut mentionner le système OMR Rodan,<sup>252</sup> qui met à disposition des utilisateurs un outil de classification interactif permettant de passer entre un mode automatique et un mode manuel pour la classification des caractères.<sup>253</sup>

Une autre possibilité, proposée par Hankinson, est de s'accommoder d'une reconnaissance OMR imparfaite mais suffisante pour faire des recherches web sur le contenu des partitions ainsi que pour mener des analyses de corpus. Selon lui, il est en effet suffisant pour les humains d'utiliser la partition scannée.<sup>254</sup> Il s'agit d'une solution intermédiaire qui est intéressante dans un premier temps, mais pose néanmoins des limitations par rapport à différents usages finaux. Une partition dans un format exploitable, mais avec des erreurs, ne peut pas être transposée automatiquement dans une autre tonalité, être éditée en braille, ou être affichée sur une tablette dans un format responsive. On peut d'ailleurs mentionner l'existence de nouvelles applications à destination des orchestres et des chœurs qui permettent au chef de diriger en donnant certaines indications directement via la partition, et, aux musiciens de voir uniquement les portées relatives à leur instrument ainsi que de ne plus avoir à tourner les pages.<sup>255</sup> Enfin, une reconnaissance OMR imparfaite pose également problème pour la plupart des solutions de suivi de partition (3.3.7).

### 3.3.7 Le suivi de partition

Le suivi de partition (score following) consiste à permettre l'écoute d'une pièce musicale tout en suivant la partition relative à cette pièce (mise en surbrillance des notes jouées). Comme mentionné précédemment, dès les années 1980 il a été possible de suivre une partition issue

---

<sup>248</sup> Voir : <https://www.aruspix.net/about>

<sup>249</sup> Voir : <https://github.com/HISPAMUS/muret>

<sup>250</sup> María Alfaro-Contreras et al., « OMR-Assisted Transcription: A Case Study with Early Prints », 2021, 5.

<sup>251</sup> Hajič Jr., « Optical Music Recognition Using Deep Learning », 19.

<sup>252</sup> Voir : <https://github.com/ddmal/rodan> ; <http://ddmal.music.mcgill.ca/e2e-omr-documentation>

<sup>253</sup> Minh Anh Nguyen, « Updates to RODAN Gamera Interactive Classifier » (Montreal, Canada, 28 mai 2018), 9, <https://simssa.ca/blog/Workshop-on-SIMSSA-XIV/>.

<sup>254</sup> Andrew Hankinson, « Optical Music Recognition » (Bodleian Libraries, University of Oxford, 2017), 42, [https://tm.web.ox.ac.uk/sites/default/files/tm/documents/media/hankinson\\_omr.pdf](https://tm.web.ox.ac.uk/sites/default/files/tm/documents/media/hankinson_omr.pdf).

<sup>255</sup> Bettina Bhend, « Mit Deep Learning zum digitalen Notenständler », Naratek, 11 décembre 2019, part. Einleitung, <https://www.naratek.com/de/2019/mit-deep-learning-zum-digitalen-notenstaender>.

d'un fichier MIDI exécutée à l'aide de sons de synthèse (3.3.2). Aujourd'hui, il est toutefois possible d'aller beaucoup plus loin grâce au développement de l'IA. Il est ainsi possible de lier une performance musicale humaine (enregistrement audio, vidéo ou performance en temps réel) avec une partition lisible par la machine (par exemple au format MEI ou MusicXML) (3.2).

Ce changement de paradigme a été rendu possible par l'évolution importante des outils et technologies appliqués dans ce domaine. La recherche pour l'alignement audio – partition a commencé dès les années 1980. A cette époque, des algorithmes qui fonctionnaient principalement sur la détection des hauteurs de notes et utilisaient des méthodes de correspondance basées sur les chaînes de caractères ont été développés.<sup>256</sup> Ces méthodes ont été améliorées vers la fin des années 1990 avec des recherches de correspondances complémentaires basées sur le rythme et le tempo.<sup>257</sup> A partir des années 2000 les chercheurs s'inspirent du domaine du Speech Recognition (ou reconnaissance de la parole).<sup>258</sup> Les deux solutions qui émergent alors sont basées sur le Modèle de Markov Caché (MMC) puis la Dynamic Time Warping (DTW ou déformation temporelle dynamique).<sup>259</sup> La plupart des logiciels développés jusqu'à cette époque, à l'instar d'Antescofo qui est également basé sur le MMC, sont des outils d'accompagnement pour le musicien.<sup>260</sup>

Depuis les années 2010, on peut mentionner le site MIREX (Music Information Retrieval Evaluation eXchange)<sup>261</sup> sur lequel quelques solutions d'alignement ont été soumises ; toutefois, elles n'ont pas encore abouties à des avancées majeures.<sup>262</sup> On peut également mentionner le travail de Henkel, Kelz et Widmer<sup>263</sup> qui a la particularité de se baser sur des partitions sous forme d'images et d'y appliquer un algorithme de Reinforcement Learning, alors que jusqu'ici toutes les méthodes décrites nécessitaient des partitions lisibles par la machine (MIDI, MusicXML, etc.).<sup>264</sup> Or, comme vu précédemment (3.3.6), l'OMR est encore un domaine en développement et la limitation de partitions en open access dans un format lisible par la machine selon les périodes constitue une limitation. Malheureusement, ce système a pour l'instant des problèmes de performance lorsqu'il est appliqué à d'autres sets de partitions et nécessiterait donc des tests complémentaires.<sup>265</sup> Enfin, on constate une tendance de la recherche actuelle à améliorer la piste audio via un prétraitement (suppression des bruitages, de la réverbération, etc.), afin d'améliorer le son avant d'effectuer l'alignement.<sup>266</sup>

Pour la mise en application des technologies de suivi de partition, plusieurs solutions d'alignement sont open source, à l'instar de celles exploitées par Lin Hao Lee dans son rapport « Musical Score Following and Audio Alignment ». Les solutions étudiées dans ce cadre, dont

---

<sup>256</sup> Lee, « Musical Score Following and Audio Alignment », 37.

<sup>257</sup> Ibid., 37.

<sup>258</sup> Ibid., 38.

<sup>259</sup> Ibid.

<sup>260</sup> Ibid., 39.

<sup>261</sup> Voir : [https://www.music-ir.org/mirex/wiki/MIREX\\_HOME](https://www.music-ir.org/mirex/wiki/MIREX_HOME)

<sup>262</sup> Lee, « Musical Score Following and Audio Alignment », 42.

<sup>263</sup> Florian Henkel, Rainer Kelz, et Gerhard Widmer, « Learning to Read and Follow Music in Complete Score Sheet Images » (arXiv, 21 juillet 2020), <http://arxiv.org/abs/2007.10736>.

<sup>264</sup> Lee, « Musical Score Following and Audio Alignment », 43-44.

<sup>265</sup> Ibid., 44.

<sup>266</sup> Ibid.

la plus performante, basée sur les technologies Constant Q-Transform (CQT) et DTW, sont accessibles sur GitHub<sup>267</sup> sous la licence GPLv3.<sup>268</sup>

Outre l'usage du suivi de partition pour l'accompagnement et la co-improvisation, d'autres types d'applications pourraient utiliser ces technologies, telles que des applications pour tourner automatiquement les pages des partitions numériques<sup>269</sup> ou pour effectuer la synchronisation automatique des sous-titres avec le jeu des acteurs dans le cadre de représentations d'opéra.<sup>270</sup> Pour les bibliothèques et archives, un usage intéressant serait la mise en relation d'enregistrements audios et vidéos avec des partitions numériques. En effet, si des grandes bases de partitions, comme RISM, et de grandes bases de fichiers audiovisuels, comme Memobase, permettent d'accéder aux documents numériques de manière FAIR dans des formats appropriés (par exemple MusicXML pour les partitions), on peut imaginer la création de nouveaux méta-catalogues réunissant diverses partitions et diverses interprétations des mêmes œuvres et offrant la possibilité de suivre une partition choisie avec une interprétation musicale choisie.

## 3.4 Projets innovateurs

### 3.4.1 PerformScore par le MACLab

Le Music and Audio Computing Lab (MACLab)<sup>271</sup> de la Graduate School of Culture Technology de Kaist (Corée du Sud) a réalisé en 2017 un système web de démonstration<sup>272</sup> permettant d'aligner des enregistrements audios d'une même partition exécutés par différents pianistes. Cette interface permet ainsi de comparer le jeu des musiciens en permutant entre plusieurs pistes audios.<sup>273</sup>

L'approche AMT (Automatic Music Transcription) est utilisée pour permettre un alignement automatisé entre la partition et les enregistrements.<sup>274</sup> Cette approche permet de transformer les enregistrements audios en fichiers MIDI qui sont ensuite alignés avec les partitions, également en MIDI, grâce à la méthode DTW.<sup>275</sup>

Les fichiers utilisés sont issus du méta-catalogue IMSLP (3.5) et du corpus de piano vienna-4x22,<sup>276</sup> tandis que les fichiers des partitions MIDI sont issus de la base de données de pianos MAPS et de vienna-4x22. Les fichiers utilisés sont disponibles sur le site du MACLab.<sup>277</sup>

---

<sup>267</sup> Voir : <https://github.com/flippy-fyp/flippy>

<sup>268</sup> Lee, « Musical Score Following and Audio Alignment », chap. 11.

<sup>269</sup> Ibid., 19.

<sup>270</sup> Andreas Arzt et al., « Artificial Intelligence in the Concertgebouw » (Autriche, 2015), 6, <https://www.ijcai.org/Proceedings/15/Papers/343.pdf>.

<sup>271</sup> Voir : <https://mac.kaist.ac.kr/>

<sup>272</sup> Voir : <https://jdasam.github.io/PerformScore>

<sup>273</sup> « Score and Performance Alignment », Music and Audio Computing Lab, consulté le 2 juillet 2022, [https://mac.kaist.ac.kr/score\\_performance\\_alignment.html](https://mac.kaist.ac.kr/score_performance_alignment.html).

<sup>274</sup> L'approche AMT est appliquée grâce à deux réseaux de neurones récurrents.

<sup>275</sup> Taegyun Kwon, Dasaem Jeong, et Juhan Nam, « Audio-to-Score Alignment Of Piano Music Using RNN-Based Automatic Music Transcription » (Finlande, 2017), 1, <https://mac.kaist.ac.kr/pubs/KwonJeongNam-smc2017.pdf>.

<sup>276</sup> Voir : [https://repo.mdw.ac.at/projects/IWK/the\\_vienna\\_4x22\\_piano\\_corpus](https://repo.mdw.ac.at/projects/IWK/the_vienna_4x22_piano_corpus)

<sup>277</sup> Voir : <https://mac.kaist.ac.kr/~ilcobo2/alignWithAMT/#ref2>

Figure 3: MACLab de Kaist, démonstration web du projet "Score and Performace Alignment". Partition : Chopin, Etude op. 10 5 ; interprètes : Cortot, Horowitz, Paderewki et Vasary. La transparence des notes correspond à leur volume, il est possible de passer dynamiquement d'un interprète à l'autre.

### 3.4.2 Prototype IIF-AV-player par le DDMAL

Le Distributed Digital Music Archives & Libraries Lab (DDMAL)<sup>278</sup>, situé à la Schulich School of Music (McGill University in Montreal, Quebec, Canada), a produit en 2018 un prototype d'alignement audio/vidéo – partition<sup>279</sup> (Figure 4 : Etude en F mineur, Op. 10 n° 9, Chopin Frédéric. Démonstration des APIs IIF pour la mise en relation de la partition et d'une représentation musicale par Valentina Lisitsa. Source : <https://ddmal.music.mcgill.ca/IIF-AV-player>. Figure 4). Dans cet exemple, les APIs IIF sont utilisées pour mettre à disposition un enregistrement audio et un enregistrement vidéo où chaque mesure a été annotée manuellement et peut donc être synchronisée avec la partition (format MEI, affichée en SVG grâce à l'outil Verovio).<sup>280</sup> Ainsi, pour chaque mesure annotée du fichier audio ou vidéo, les notes de la partition correspondante sont mises en surbrillance.<sup>281</sup>

<sup>278</sup> Voir : <https://ddmal.music.mcgill.ca>

<sup>279</sup> Voir démonstration : <https://ddmal.music.mcgill.ca/IIF-AV-player> ; code source, <https://github.com/DDMAL/IIF-AV-player>

<sup>280</sup> Andy Irving, IIF AV, 3 juillet 2020, 17:23, <https://youtu.be/1noRWCvyUrg?t=987>.

<sup>281</sup> Ibid.

Par rapport à l'exemple précédent du MACLab de Kaist, ce prototype est innovant car il utilise les APIs IIIF. En revanche, la segmentation des fichiers est effectuée manuellement, ce qui implique donc de combiner cette solution à des systèmes de suivis automatiques pour un usage à grande échelle.

Figure 4 : Etude en F mineur, Op. 10 n° 9, Chopin Frédéric. Démonstration des APIs IIIF pour la mise en relation de la partition et d'une représentation musicale par Valentina Lisitsa. Source : <https://ddmal.music.mcgill.ca/IIIF-AV-player>.

### 3.4.3 Musiclibs

Musiclibs<sup>282</sup> est un projet de méta-catalogue de partitions lancé par SIMSSA (Single Interface for Music Score Searching and Analysis) en 2017.

SIMSSA est un projet international lancé en 2011 auquel participe d'ailleurs le DDMAL.<sup>283</sup> Ce projet a été fondé avec deux objectifs de taille, à savoir : « apprendre aux ordinateurs à extraire les symboles musicaux de partitions musicales ayant été numérisées » et « réunir en un seul lieu virtuel les collections de bibliothèques et de musées du monde entier ». <sup>284</sup> SIMSSA travaille activement au développement d'une interface web permettant à la fois d'appliquer

<sup>282</sup> Voir : <https://www.musiclibs.net>

<sup>283</sup> « Presentations », SIMSSA, consulté le 6 août 2022, <https://simssa.ca/activities/presentations/>.

<sup>284</sup> « Accueil », SIMSSA, consulté le 3 août 2022, <https://simssa.ca/>.

une reconnaissance OMR aux partitions chargées, mais aussi d'effectuer des analyses et des recherches sur ces dernières.<sup>285</sup>

Musiciens s'inscrit bien dans ces objectifs, mais semble malheureusement avoir été laissé de côté, l'interface en ligne n'étant pas fonctionnelle. Toutefois, les solutions techniques sur lesquelles le projet repose sont très intéressantes. Il s'agit de récupérer les manifestes des partitions via IIF, d'appliquer une reconnaissance OMR grâce à l'outil web Rodan développé par le DDMAL (3.3.6.2), de créer un fichier MEI permettant l'analyse et la recherche dans le contenu de la partition, et enfin de mettre à disposition dans l'interface Musiciens à la fois le document scanné et le document MEI (Figure 5).<sup>286</sup> À terme, l'objectif était de faire de Musiciens une interface de recherche *full-music*,<sup>287</sup> permettant aux utilisateurs de rechercher les métadonnées fournies par les institutions externes, et le contenu de la notation musicale obtenue par OMR.<sup>288</sup>

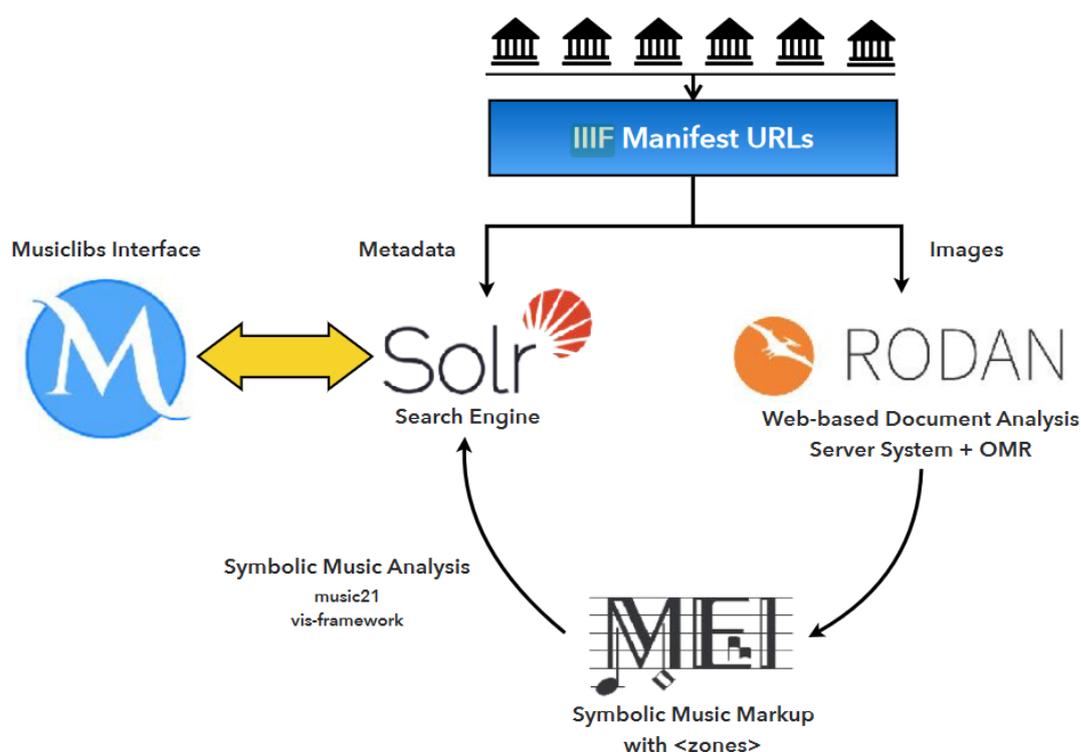


Figure 5: Musiclibris Recognition & Indexing Process, Sources : Andrew Hankinson, « Optical Music Recognition » (Bodleian Libraries, University of Oxford, 2017), [https://tm.web.ox.ac.uk/sites/default/files/tm/documents/media/hankinson\\_omr.pdf](https://tm.web.ox.ac.uk/sites/default/files/tm/documents/media/hankinson_omr.pdf).

<sup>285</sup> « About the SIMSSA project », SIMSSA, consulté le 3 août 2022, <https://simssa.ca/about/>.

<sup>286</sup> Hankinson, « Optical Music Recognition », 52.

<sup>287</sup> 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.

<sup>288</sup> « About », Musiclibris, part. The Future, consulté le 6 août 2022, <https://www.musiclibris.net/about/>.

### 3.4.4 Distant listening

Distant listening est un projet du Digital and Cognitive Musicology Lab (DCML) de l'EPFL financé par le Fonds National Suisse (FNS), qui a pour objectif l'étude du développement de l'harmonie entre 1700 et 2000.<sup>289</sup>

Actuellement en cours (2019-2023), ce projet a pour but la modélisation des structures harmoniques de la tonalité courante, de la tonalité étendue (19<sup>e</sup> siècle) et des nouvelles formes harmoniques utilisées dans les musiques pop, rock et jazz.<sup>290</sup> Cette recherche requiert dans un premier temps la création de corpus musicaux annotés ainsi que le développement d'un modèle d'analyse harmonique.<sup>291</sup> Pour atteindre son objectif elle cherche à combiner 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).<sup>292</sup>

Plusieurs billets de blogs et articles scientifiques ont déjà été publiés concernant les résultats d'analyse.<sup>293</sup> On peut mentionner "Tracing historical changes in the exploration of tonal space", qui via un corpus de 2000 pièces (75 compositeurs) entre 1350 et 1940 et l'application de méthodes statistiques et d'apprentissage automatique (ici la réduction de dimension), 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.<sup>294</sup>

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.<sup>295</sup> Appliquées à des corpus importants, comme c'est le cas ici, cela permet d'étudier à grande échelle les changements stylistiques dans l'évolution de la musique au cours de l'histoire.<sup>296</sup> Cette technologie permettrait de comprendre les spécificités stylistiques propres à des corpus ou à des compositeurs spécifiques.

Une phase importante de ce projet est la préparation du corpus. Ce qui a nécessité de résoudre différentes problématiques telles que la limitation du nombre d'œuvre disponibles, ou l'absence de normes communément utilisées dans la notation de l'harmonie ; ce qui peut être attribué à des différences conceptuelles dans l'interprétation des caractéristiques harmoniques, ainsi qu'à des problèmes de communication entre chercheurs.<sup>297</sup> Pour parer à ce problème spécifique, le DCML a développé une norme pour la notation de l'harmonie<sup>298</sup> ainsi qu'un processus de travail utilisant GitHub.<sup>299</sup> Ce processus a été mis en place afin de

---

<sup>289</sup> « Distant Listening », EPFL, consulté le 3 août 2022, <https://www.epfl.ch/labs/dcml/projects/distant-listening/>.

<sup>290</sup> Ibid.

<sup>291</sup> Ibid.

<sup>292</sup> Ibid.

<sup>293</sup> Ibid., part. Research output.

<sup>294</sup> Fabian C. Moss, « Tracing Historical Changes in the Exploration of Tonal Space », EPFL, 28 septembre 2020, <https://www.epfl.ch/labs/dcml/tracing-historical-changes/>.

<sup>295</sup> Ibid.

<sup>296</sup> Ibid.

<sup>297</sup> « Representing Harmony: Goals and Challenges », EPFL, consulté le 3 août 2022, <https://www.epfl.ch/labs/dcml/workshops/representing-harmony/>.

<sup>298</sup> Voir : <https://github.com/DCMLab/standards>

<sup>299</sup> Voir : <https://github.com/DCMLab/corelli>

permettre aux professionnels de co-annoter les corpus et de trouver un consensus en cas d'avis divergents.<sup>300</sup>

Concernant la problématique du nombre d'œuvres, un projet spécifique a été lancé pour la constitution d'un corpus incluant le développement d'outils spécifiques.

#### 3.4.4.1 Corpus Project

Afin de parer à la rareté de grands corpus incluant l'annotations des harmonies tonales et des contrepoints, ce projet a été lancé afin de réunir des partitions et de les annoter en utilisant des normes nouvellement développées par le DCML.<sup>301</sup>

Concrètement, il s'est agi : de réunir des partitions issues du web dans différents formats (MusicXML, Humdrum, etc.), de les convertir si nécessaire au format Muscore3 (mscz) en utilisant une librairie Python développée à cet effet (la librairie ms3)<sup>302</sup>, d'annoter manuellement les harmonies dans les partitions, et d'annoter les cadences dans des tableaux TSV.<sup>303</sup> La librairie ms3 permet également d'exporter les éléments des partitions dans des tableaux TSV (notes, mesures, harmonies) ou d'ajouter des éléments annotés dans des fichiers TSV (comme les cadences) à des partitions.<sup>304</sup>

Tous les documents annotés n'ont pas encore été mises à disposition du public, mais il est prévu que ce soit le cas.<sup>305</sup> Le but de ce projet est de permettre la réutilisation de ces ressources pour d'autres projets d'humanités numériques.<sup>306</sup> Actuellement, seulement une partie des partitions est disponible sur GitHub.<sup>307</sup>

#### 3.4.5 Le projet Neuma

Neuma<sup>308</sup> est un méta-catalogue en devenir. Il agrège déjà plusieurs sources, dont RISM, et il propose le téléchargement des partitions au format PDF, MEI et MusicXML. Le lecteur Verovio<sup>309</sup>, intégré au catalogue, permet de lire les partitions.

Neuma est issu d'un projet de recherche (2009-2011) financé par l'Agence Nationale Française de la Recherche. La plateforme est actuellement maintenue par l'institut IReMus et le Musicology lab.<sup>310</sup> Elle poursuit le triple objectif de : donner accès à des partitions rares dans une édition soignée et homogène ; fournir des outils pour la recherche et l'analyse ; et enfin

---

<sup>300</sup> Johannes Hentschel et al., « A semi-automated workflow paradigm for the distributed creation and curation of expert annotations », in *Proceedings of the 22nd International Society for Music Information Retrieval Conference*, 2021, 262, <https://doi.org/10.5281/zenodo.5624417>.

<sup>301</sup> « Corpus Project », EPFL, consulté le 4 juin 2022, <https://www.epfl.ch/labs/dcml/projects/corpus-project/>.

<sup>302</sup> Voir : <https://github.com/johentsch/ms3>

<sup>303</sup> *Creating and Evaluating an Annotated Corpus Using the Library ms3 (Johannes Hentschel)*, 2020, <https://www.youtube.com/watch?v=UBY3wuIS4wc>.

<sup>304</sup> Ibid.

<sup>305</sup> Ibid.

<sup>306</sup> « Corpus Project ».

<sup>307</sup> Voir : [https://github.com/DCMLab/dcml\\_corpora](https://github.com/DCMLab/dcml_corpora)

<sup>308</sup> Voir : <http://neuma.huma-num.fr>

<sup>309</sup> Voir : <https://www.verovio.org>

<sup>310</sup> « Contact », Neuma, consulté le 2 août 2022, <http://neuma.huma-num.fr/home/contact>.

de 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.<sup>311</sup>

C'est également une plateforme qui vise à permettre aux usagers de se créer un compte, de créer leurs propres collections, d'annoter les partitions existantes et d'ajouter de nouvelles partitions. C'est donc une plateforme qui vise à réunir une communauté d'utilisateurs actifs qui participeront à l'élargissement des collections et des connaissances qui y sont associées.<sup>312</sup>

Enfin, la plateforme propose un service d'APIs REST<sup>313</sup> qui permet l'échange de documents JSON selon le schéma OpenAPI.<sup>314</sup>

Les éléments novateurs de cette plateforme sont : la possibilité de lire les partitions directement dans le catalogue, la mise à disposition des partitions dans des formats exploitables par la machine, la volonté d'encourager le crowdsourcing sur une plateforme de ce type et la mise à disposition d'APIs bien documentées permettant d'extraire des sets de partitions exploitables à des fins d'analyse. Bien que le projet initial soit ancien (2009-2011), de nombreux éléments ont été ajoutés à la plateforme entre février et juillet 2022. On peut donc espérer que cette dernière va continuer à se développer et proposer toutes les fonctionnalités précédemment évoquées dans un délai court.

### 3.4.6 ELVIS project

ELVIS (Electronic Locator of Vertical Interval Successions) est un projet qui a été initialement financé par la subvention Digging into Data Challenge. Aujourd'hui, il est activement maintenu et développé par la branche de recherche et d'analyse du projet SIMSSA. ELVIS, comme SIMSSA, est basé à la Schulich School of Music de la McGill University (Montreal, Canada).<sup>315</sup>

Le projet ELIS englobe plusieurs solutions distinctes dont la base de données ELVIS et le framework VIS

#### 3.4.6.1 ELVIS database

La base de données ELVIS<sup>316</sup> a été lancée en janvier 2012.<sup>317</sup> Aujourd'hui, elle réunit près de 3000 pièces dans un format lisible par la machine.<sup>318</sup> Les fichiers sont en général disponibles au format MusicXML et PDF et sont parfois également proposés dans d'autres formats, comme MEI ou MIDI.

Les utilisateurs doivent se créer un compte pour accéder aux fichiers. Ils peuvent également uploader leurs propres fichiers de musique. Probablement pour cette raison, les métadonnées ne reposent pas sur un standard (titre, auteur, dates de composition, nombre de voix, genres, instrumentation, langue du texte, lieu d'origine, sources, tags, vocal/instrumental/mixte, sacré/laïque/non catégorisé, commentaire) ; en outre, une arborescence peut être insérée (mouvement). Les utilisateurs peuvent créer et modifier des collections qui sont visibles par

---

<sup>311</sup> « Accueil », Neuma, consulté le 2 août 2022, <http://neuma.huma-num.fr/>.

<sup>312</sup> « Présentation », Neuma, consulté le 28 juillet 2022, <http://neuma.huma-num.fr/home/presentation>.

<sup>313</sup> Voir : <http://neuma.huma-num.fr/rest/swagger>

<sup>314</sup> « Services », Neuma, consulté le 28 juillet 2022, <http://neuma.huma-num.fr/home/services>.

<sup>315</sup> « About the Elvis Database », ELVIS, consulté le 3 août 2022, <https://database.elvisproject.ca/about/>.

<sup>316</sup> Voir : <https://database.elvisproject.ca>

<sup>317</sup> « About the Elvis Database ».

<sup>318</sup> Ibid.

tous. Ils peuvent également créer des lots de documents et les télécharger (avec arborescence ou non) ; seul bémol, l'option pour créer un lot directement à partir des résultats d'une recherche n'est pas disponible pour les comptes standards. Les lots téléchargés sont accompagnés d'un fichier de métadonnées. Enfin, une application web<sup>319</sup> disponible sur GitHub permet d'uploader, rechercher et télécharger les ressources d'ELVIS database.

En résumé, ce catalogue est collaboratif et permet aux utilisateurs d'accéder à des données choisies dans des formats permettant l'application de méthodes d'analyse. Il est par exemple possible d'effectuer une recherche sur toutes les pièces sacrées composées en Italie entre 1500 et 1700, d'obtenir les documents au format XML et de les analyser à l'aide d'un logiciel open source tel que music21<sup>320</sup> (un outil spécifiquement développé pour effectuer des analyses de corpus musicaux).<sup>321</sup>

Les seuls éléments que l'on peut véritablement déplorer par rapport aux principes FAIR sont l'absence de la licence dans les métadonnées des documents, et, de manière plus générale, le fait de ne pas s'être basé sur une norme existante pour les métadonnées.

On peut noter qu'il est annoncé qu'ELVIS database évolue en SIMSSA database<sup>322</sup>. Cette dernière devrait être améliorée au niveau du modèle de données, en se focalisant notamment sur l'accessibilité et l'intégrité des données.<sup>323</sup> SIMSSA database, devrait, comme Musiclibs, permettre la recherche sur le contenu des partitions en plus des métadonnées ainsi qu'offrir des fonctionnalités d'extraction pour les outils MIR. De plus, les données devraient être mises à disposition au format LOD.<sup>324</sup> SIMSSA database promet donc une très bonne compatibilité avec les principes FAIR et la réutilisation des données en général.

### 3.4.6.2 Le Framework VIS

VIS<sup>325</sup> est un framework gratuit sous licence AGPLv3+ développé en Python et utilisant music21 et pandas<sup>326</sup> (un outil pour l'analyse de données). L'objectif est de permettre la création de systèmes flexibles et simples à développer destinés à l'analyse de la musique écrite dans des formats exploitables.<sup>327</sup>

## 3.5 Méta-catalogues de partitions

Les méta-catalogues RISM (2.2.1.1), ELVIS (3.4.6) et Neuma (3.4.5) ont été vus jusqu'ici. Ils ont en commun un caractère institutionnel. 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

---

<sup>319</sup> Voir : <https://github.com/ELVIS-Project/elvis-database>

<sup>320</sup> Voir : <http://web.mit.edu/music21>

<sup>321</sup> « About the Elvis Database ».

<sup>322</sup> Voir : <https://github.com/ELVIS-Project/simssadb>

<sup>323</sup> Gustavo Polins Pedro et Andrew Kam, « SIMSSA Database », 4, [https://simssa.ca/assets/files/SIMSSADB\\_May2018.pdf](https://simssa.ca/assets/files/SIMSSADB_May2018.pdf).

<sup>324</sup> Yaolong Ju et al., « SIMSSA DB », 2019, 4, <https://www.semanticscholar.org/paper/SIMSSA-DB-Ju-McKay/b1cc3cadabe60c622bf3afe008ccfd505c74ce4d>.

<sup>325</sup> Voir : <https://vis-framework.readthedocs.io/en/v3.0.5>

<sup>326</sup> Voir : <https://pandas.pydata.org>

<sup>327</sup> « About This Documentation — VIS Framework Documentation (Version 3.0.5) », VIS API, consulté le 4 août 2022, <https://vis-framework.readthedocs.io/en/v3.0.5/>.

au niveau mondial. C'est avant tout un catalogue de métadonnées (1'601'000 notices), dont environ 6% sont rattachées à un document numérique (103'000 documents numériques).

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 (PDF généré à partir des autres fichiers), à l'inverse de RISM. Ces deux catalogues ont comme objectif de permettre la réutilisation de leurs documents numériques à des fins de recherche et d'analyses de corpus musicaux, mais leurs métadonnées sont moins détaillées (surtout celles de Neuma) et ne respectent pas de standard. RISM ne propose pas de comptes lecteurs, alors qu'au contraire Neuma et ELVIS vont jusqu'à permettre aux utilisateurs de participer au catalogue. Neuma a comme objectif supplémentaire de rendre accessibles les partitions rares et est le seul catalogue à intégrer un lecteur de partition. ELVIS met à disposition près de 3000 partitions numériques complètes alors que Neuma n'en propose encore que très peu (souvent uniquement les incipits sont disponibles). Actuellement, on peut voir que Neuma a du potentiel, mais c'est un projet en cours qui requiert encore beaucoup d'améliorations et de développements au niveau de l'interface, ainsi que l'ajout d'une grande quantité de données. À l'inverse, RISM et ELVIS sont des projets aboutis et stables. Enfin, ELVIS servira de socle au grand projet à venir SIMSSA database, qui devrait être le premier système à mettre des données à disposition dans un format lisible par la machine, sous la forme de LOD et de manière pleinement compatible avec les principes FAIR.

Outre ces trois catalogues, il convient encore de mentionner IMSLP<sup>328</sup> (International Music Score Library Project) et MuseScore<sup>329</sup> 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 RISM, ELVIS et Neuma.

IMSLP, également connu sous le nom de Bibliothèque musicale Petrucci, est un projet lancé en 2006 et géré par la société à but non lucratif Project Petrucci LLC.<sup>330</sup> Ce catalogue met à disposition près de 668'000 et 75'000 enregistrements audios qui couvrent toutes les périodes jusqu'à nos jours, dont majoritairement le 5<sup>e</sup> au 18<sup>e</sup> siècle. IMSLP accueille des partitions dans tous les styles de musique, y compris le jazz ou le rock, toutefois il est nécessaire que les partitions soient sous licence libre ou dans le domaine public. On peut encore mentionner qu'IMSLP a fusionné avec Werner Icking Music Archive (WIMA) en 2012.<sup>331</sup> De plus, des projets sont en cours pour intégrer de nouvelles collections issues des institutions suivantes : Sibley Music Library (University of Rochester), Gaylord Music Library (Washington University), Royal Danish Library (Det Kongelige Bibliotek).<sup>332</sup>

MuseScore est avant tout un logiciel libre pour l'édition de partitions. Le catalogue de MuseScore était donc initialement limité aux utilisateurs du logiciel afin de leur permettre de partager leurs travaux.<sup>333</sup> Aujourd'hui, le catalogue en ligne compte près de 1'500'000 partitions qui peuvent toutes être consultées en ligne et dont près de 340'000 sont dans le

---

<sup>328</sup> Voir : <https://imslp.org>

<sup>329</sup> Voir : <https://musescore.com/sheetmusic>

<sup>330</sup> « About », IMSLP, consulté le 5 août 2022, <https://imslp.org/wiki/IMSLP:About>.

<sup>331</sup> « WIMA », IMSLP, consulté le 5 août 2022, <https://imslp.org/wiki/IMSLP:WIMA>.

<sup>332</sup> « Contributor Portal », IMSLP, consulté le 5 août 2022, [https://imslp.org/wiki/IMSLP:Contributor\\_Portal](https://imslp.org/wiki/IMSLP:Contributor_Portal).

<sup>333</sup> « MuseScore », in *Wikipedia*, 25 juillet 2022,

<https://en.wikipedia.org/w/index.php?title=MuseScore&oldid=1100370114>.

domaine public. Ces dernières peuvent être téléchargées librement dans les formats Muscore, MusicXML, MIDI et PDF, tandis que les partitions non libres sont téléchargeables sur abonnement. Le logiciel MuseScore inclut une fonctionnalité web pour la reconnaissance OMR basée sur Audiveris,<sup>334</sup> ce qui permet aux utilisateurs d'importer des partitions scannées. Tous les types de musiques sont présents avec une majorité pour le classique, les musiques de films, le pop, le rock et le folk.

En 2017, MuseScore et IMSLP ont lancé l'initiative OpenScore<sup>335</sup>. Cette dernière visait à créer des fichiers Muscore et MusicXML sous licence CCO<sup>336</sup> à partir des partitions IMSLP dans le domaine public ; ceci en s'appuyant sur la communauté pour améliorer les transcriptions automatiques.<sup>337</sup> Il s'agit d'un projet toujours en cours, les partitions étant disponibles sur le site de MuseScore. En plus du projet principal, il y a également deux groupes de transcriptions annexes qui sont : OpenScore Lieder Corpus<sup>338</sup> (chansons françaises, allemandes et anglaises) et OpenScore Braille<sup>339</sup> (exportations de partitions MusicXML pour la conversion en braille).<sup>340</sup>

## 3.6 Mise en perspective avec les Archives musicales de la BCUL

### 3.6.1 Normes et formats

Les partitions numérisées mises à disposition sur Patrinum par les Archives musicales sont maintenant toujours au format TIFF pour la sauvegarde à long terme. Des fichiers JPEG sont générés automatiquement pour les lecteurs (une fonctionnalité permet de créer un PDF avec l'ensemble des fichiers JPEG liés à une partition). Par contre, il n'y a aucun projet en cours ou à venir pour effectuer une reconnaissance OMR de ces documents afin de les proposer également au format MEI ou MusicXML.

De leur côté, les partitions nées-numériques disposent d'une copie au format MusicXML également chargée sur Patrinum. Toutefois, les utilisateurs n'ont actuellement accès qu'au document PDF.

La solution choisie pour conserver les partitions nées numériques est donc tout à fait adéquate et a même permis de proposer très rapidement un set de partitions pour compléter le corpus utilisé dans le cadre du projet Distant listening (3.4.4.1). Il serait par contre intéressant d'également rendre accessibles les partitions au format MusicXML sur Patrinum.

De manière générale, on constate que les nouveaux projets d'envergure comme SIMSSA ont une volonté forte de mettre à disposition des partitions dans des formats exploitables. A ce titre, on peut aussi mentionner la BnF (Bibliothèque Nationale de France), qui a lancé une étude en 2019 pour analyser la possibilité d'effectuer une reconnaissance OMR massive des partitions de Gallica<sup>341</sup>, et qui devrait se concrétiser par la mise en place d'un éditeur en ligne

---

<sup>334</sup> Ibid.

<sup>335</sup> Voir : <https://musescore.com/openscore>

<sup>336</sup> Voir : <https://creativecommons.org/share-your-work/public-domain/cc0>

<sup>337</sup> « Liberating Sheet Music », OpenScore, consulté le 5 août 2022, <https://openscore.cc>.

<sup>338</sup> Voir : <https://musescore.com/OpenScore-Lieder-Corpus>

<sup>339</sup> Voir : <https://musescore.com/openscore-braille>

<sup>340</sup> « MuseScore ».

<sup>341</sup> Voir : <https://gallica.bnf.fr>

pour la correction collaborative.<sup>342</sup> Toutefois, aucune solution fonctionnelle ne semble être actuellement en place du côté institutionnel. À l'inverse, IMSLP et MuseScore ont réussi à avancer de manière très concrète sur le projet OpenScore en mettant en place une solution de crowdsourcing qui fonctionne et qui a permis de mettre en ligne plus de 1000 partitions transcrites.<sup>343</sup> On peut néanmoins remarquer, que même dans ce cadre, l'avancée du projet n'est pas simple et que des équipes ont dû être recrutées en plus des bénévoles pour la transcription, comme en témoigne le commentaire suivant (réponse à un usager qui se demande pourquoi le projet principal avance moins vite que le projet Lieder corpus) :

“The Lieder Corpus has received funding from two universities (Cambridge and Cornell), which allowed us to bring a team of reviewers on board to help with that project. The scores on the main project are longer and more complicated, so transcribers and reviewers need extra training before they make a meaningful contribution. However, we are starting to move some of the more experienced people across from the Lieder Corpus to help accelerate the main project.”<sup>344</sup>

À l'heure actuelle, il n'existe donc pas de solution simple pour transcrire les partitions scannées sans engendrer un travail manuel très important. On peut néanmoins remarquer que la participation à IMSLP permettrait d'une part de diffuser largement les partitions numérisées de la BCUL, et d'autre part, de participer à OpenScore, qui reste le projet de transcription le plus avancé actuellement. Une autre solution serait d'appliquer une reconnaissance OMR non vérifiée mise à disposition sur Patrinum en parallèle avec les documents JPEG, mais cette option n'offrirait pas vraiment de plus-value directe, à moins d'adapter l'interface de Patrinum à la recherche *full music*. Enfin, il est aussi simplement possible d'attendre que SIMSSA database se développe ; dans ce cas, une mise à disposition des documents sur Patrinum via IIF serait suffisante, la reconnaissance OMR étant effectuée du côté de SIMSSA.

Du côté des métadonnées, le format MARC de Patrinum correspond au format du méta-catalogue RISM, avec la nuance que Patrinum sort un peu du standard avec l'adaptation de MARC à la norme ISAD(G). Klaus Keil et Laurent Pugin mentionnent à ce sujet un équilibre à trouver entre les formats spécifiques à la musique et ceux utilisés pour les ressources textuelles en général, car les premiers permettent de tenir compte de toutes les spécificités liées au type de document, tandis que les seconds sont plus adaptés pour le partage des données avec des institutions non spécialisées.<sup>345</sup> À ce titre, il existe un outil disponible sur GitHub<sup>346</sup> pour la conversion MARC – MEI développé dans le cadre de RISM.<sup>347</sup> Dans la mesure où Patrinum permet l'export en MARC/XML, le format des métadonnées actuel est donc tout à fait adéquat au type de document.

---

<sup>342</sup> Philippe Rigaux, « Projet GioQoso - Gestion de la qualité des partitions musicales ouvertes », 2019, 16, [https://www.cnrs.fr/mi/IMG/pdf/philippe\\_rigaux\\_-\\_gioqoso\\_-\\_mastodons2019.pdf](https://www.cnrs.fr/mi/IMG/pdf/philippe_rigaux_-_gioqoso_-_mastodons2019.pdf).

<sup>343</sup> « 1000 Scores in the Lieder Corpus! », OpenScore : Kickstarter, 28 mars 2021, <https://www.kickstarter.com/projects/openscore/openscore-join-the-sheet-music-revolution>.

<sup>344</sup> Ibid.

<sup>345</sup> Keil et Pugin, « Das Internationale Quellenlexikon der Musik, RISM », part. 6.

<sup>346</sup> Voir : <https://github.com/music-encoding/encoding-tools>

<sup>347</sup> Keil et Pugin, « Das Internationale Quellenlexikon der Musik, RISM », part. 6.

### 3.6.2 Interopérabilité et méta-catalogues

Il serait souhaitable d'atteindre un millier de partitions numérisées accessibles en open access sur Patrinum pour que des projets puissent être développés uniquement à partir des ressources des Archives musicales. Il est donc particulièrement important de mettre à disposition les données via des méta-catalogues, afin qu'elles soient incluses dans les corpus inter-institutionnels utilisés pour la recherche. Malheureusement il n'existe pas, à l'heure actuelle, de méta-catalogue national suisse pour les partitions. Au niveau international, le projet RISM est très intéressant avec sa mise à disposition des métadonnées en LOD et son adéquation presque parfaite avec les principes FAIR. Lorsqu'il sera possible d'y ajouter les partitions du 20<sup>e</sup> siècle, ce sera donc un catalogue incontournable pour ce domaine. D'ici là, il serait utile que Patrinum publie les manifestes IIF des partitions.

Concernant les autres méta-catalogues évoqués jusqu'ici, une publication dans Neuma ne fait pas vraiment sens, sauf pour d'éventuelles partitions rares. Quant à ELVIS, seuls les documents nés numériques disponibles au format MusicXML pourraient y être publiés. En fonction du nombre de documents disponibles en open access dans ce format, une collaboration pourrait donc être intéressante. Toutefois, si SIMSSA database avance rapidement, une publication des fichiers MusicXML via IIF (en utilisant la propriété seeAlso)<sup>348</sup> sera probablement suffisante.

IMSLP est un catalogue d'ampleur internationale incontournable pour les musiciens. C'est aussi le seul catalogue à mettre à disposition des enregistrements audios en plus des partitions, d'où l'usage de ses ressources dans certains travaux de recherche. Pour diffuser au mieux les collections aux utilisateurs finaux, il serait donc utile d'y participer, bien qu'IMSLP ne respecte pas les principes FAIR.

### 3.6.3 Projet de suivi de partition ?

Les Archives musicales ont un partenariat avec la Phonothèque nationale suisse<sup>349</sup> concernant la conservation des documents sonores. En effet, seuls les documents vaudois en lien avec la musique écrite sont conservés à la BCUL (partitions, fonds d'archives, etc.), tandis que la Phonothèque inclut dans son périmètre les documents sonores vaudois. Un projet intéressant serait donc d'utiliser les technologies de suivi de partition pour permettre l'écoute des documents audios en parallèle avec la partition, de manière similaire au projet PerformScore (3.4.1).

Toutefois, un tel projet requerrait non seulement un travail important pour la préparation des données (format MusicXML ou MEI requis pour les partitions, nécessité de disposer des fichiers audios au format numérique), mais aussi pour la mise en œuvre technique (méthode d'alignement à choisir, set d'entraînement à trouver ou créer, entraînement d'un réseau de neurones, etc.), ainsi que pour la conception d'une plateforme web adaptée. Le fait qu'il n'ait pas été possible de trouver des exemples de projets autres que des prototypes montre d'ailleurs bien la complexité d'un tel travail.

Il s'agit donc d'un projet qui n'est dans tous les cas pas réalisable par une institution seule, mais qui pourrait probablement le devenir dans le cadre d'un partenariat, par exemple avec

---

<sup>348</sup> Voir : <https://iif.io/api/presentation/2.1/#seealso>

<sup>349</sup> Voir : <https://www.fonoteca.ch>

la BN et le DCML de l'EPFL. On peut également noter que l'amélioration constante des solutions techniques pour le suivi de partition, ainsi que l'augmentation du nombre de ressources audio-visuelles en ligne (on peut penser à Memobase), devraient faciliter la mise en œuvre de tels projets.

## 4 Comment améliorer la mise à disposition ?

L'état des lieux a permis de mettre en évidence plusieurs points d'amélioration sur différents niveaux. Pour améliorer la mise à disposition au niveau régional, un travail de documentation pourrait être fait directement sur les plateformes internes de la BCUL, à savoir Patrinum et Scriptorium. Au niveau national, la BCUL participe déjà aux grandes plateformes suisses, mais les données qui y sont publiées sont quantitativement faibles. Enfin, au niveau international, la BCUL participe principalement à Google Books, il serait donc intéressant d'envisager d'autres méta-catalogues.

### 4.1 Réutilisation depuis les plateformes internes

Sans qu'il soit question de développements supplémentaires, les deux plateformes internes de la BCUL offrent déjà plusieurs possibilités pour la mise à disposition des données qui ne sont actuellement pas exploitées. Il s'agit de : la publication des métadonnées via le protocole OAI-PMH ; de la publication des manifestes IIF ; et de la mise à disposition de la documentation concernant les APIs.

Du côté d'OAI-PMH, il serait possible d'ajouter une simple mention du les sites Patrinum de Scriptorium, afin de préciser que les données sont disponibles sur demande via ce protocole (la BN procède de cette manière).<sup>350</sup> Pour aller plus loin, il serait également possible de créer une page de documentation détaillée, comme par exemple celle de Memobase.<sup>351</sup>

Concernant IIF, la création de manifestes est déjà prévue lorsque les APIs seront fonctionnelles pour Patrinum. Concernant Scriptorium, il serait intéressant d'également utiliser cette option. À noter que cela nécessiterait par contre d'ajouter la licence aux métadonnées.

Enfin, à propos des APIs, celles nativement proposées par TIND (Patrinum) permettent facilement d'effectuer des recherches pointues dans les données. En effet, les critères d'une recherche effectuée depuis l'interface publique se retrouvent dans l'URL de la page des résultats. La pratique recommandée est donc simplement de copier la portion d'URL concernée et de l'intégrer au point de terminaison de l'API accompagnée des paramètres choisis (format de fichier à obtenir, champs MARC à inclure, etc.). En outre, l'API permet également d'accéder au fichier et aux métadonnées des fichiers. L'usage des APIs nécessite d'avoir un compte utilisateur avec le rôle approprié et il est notamment possible de limiter l'accès à la recherche ou à l'export des documents et métadonnées. Enfin, les droits de

---

<sup>350</sup> Voir l'exemple de la BN : <https://www.nb.admin.ch/snl/fr/home/informations-professionnels/mettre-a-disposition/opendata>

<sup>351</sup> Voir l'exemple de Memobase : <https://ub-basel.atlassian.net/wiki/spaces/MD/pages/694124708>

l'utilisateur concernant les documents accessibles sont répercutés dans l'API. Si une collection est cachée pour un usager, ce dernier n'y aura donc pas non plus accès par API.

Du côté de Scriptorium, MedialINFO propose des API RESTful qui permettent d'accéder aux métadonnées et au texte OCR. Des recherches précises, combinant une série de critères, peuvent également être effectuées. Contrairement à Patrinum, il n'est par contre pas possible de limiter l'accès à une collection donnée en fonction des droits du compte utilisateur.

Concernant la mise à disposition de la documentation API, cette dernière a actuellement un statut confidentiel. Il n'est donc pas possible de la rendre publique comme le fait par exemple la Biodiversity Heritage Library<sup>352</sup>, qui propose même un formulaire uniquement dédié à la demande d'une clef API<sup>353</sup>. Toutefois, on peut imaginer ajouter une mention sur les sites de Patrinum et de Scriptorium pour informer les utilisateurs que des APIs sont disponibles sur demande. Pour aller plus loin, il serait intéressant de mettre à disposition une documentation pour les lecteurs disposant de comptes avec des droits spécifiques.

## 4.2 Les méta-catalogues Suisses

La BCUL participe déjà aux grandes plateformes suisses telles que e-codices, e-rara, RISM suisse, CCSA et projette de participer à Memobase. Or, tous ces méta-catalogues respectent les principes FAIR où en sont proches. Les Manuscrits et archives, les Archives musicales, l'Iconopôle et la Réserve précieuse participent donc chacun à au moins un méta-catalogue suisse adapté à la réutilisation des données. On constate par contre que le nombre de documents mis à disposition sur ces méta-catalogues ne représente qu'une petite partie de ces collections. Il serait donc intéressant, dans la mesure des ressources disponibles, de régulariser les versements sur ces plateformes.

Concernant les collections qui ne sont pas adaptées aux catalogues actuellement utilisés par la BCUL, on pourrait imaginer sélectionner de nouveaux méta-catalogues, comme Swisscollections pour les fonds d'archives, les partitions ou encore les cartes géographiques. Du côté de la presse, les plateformes E-newspaperarchives ou E-periodica, qui proposent chacune le partage via IIF, seraient également intéressantes. Toutefois, la plateforme Connectome est actuellement le projet incontournable pour la mise à disposition des données au niveau suisse.

### 4.2.1 Le projet Connectome

Le projet Connectome a été lancé en 2020 et a pour objectif de faciliter et d'accélérer la réutilisation des données dans les domaines des sciences sociales et humaines (SSH).

#### 4.2.1.1 Prémices

Dans un premier temps, le SWITCH Innovation Lab "Repositories & Data Quality" a été mandaté afin de faire un état des pratiques de recherche dans le domaine des sciences humaines et sociales. Dans le cadre d'une collaboration avec ZHAW Zurich University of Applied Sciences, le SWITCH Innovation Lab a produit un rapport concernant la réutilisation

---

<sup>352</sup> Voir l'exemple de la BHL : <https://about.biodiversitylibrary.org/tools-and-services/developer-and-data-tools/#APIs>

<sup>353</sup> Voir le formulaire pour la clef API : <https://www.biodiversitylibrary.org/getapikey.aspx>

des données par les chercheurs et sur les critères de leurs choix concernant les sources.<sup>354</sup> Or, du côté patrimonial, l'enquête a montré que les données issues des institutions GLAM ne sont pas toujours disponibles pour les chercheurs.<sup>355</sup> Enfin, la variété des bases de données sources mentionnées comme pertinentes pour les sciences humaines fait ressortir l'importance d'interconnecter les bases, comme l'ont exprimé à plusieurs reprises les répondants à l'enquête.<sup>356</sup>

#### 4.2.1.2 Description et technologies

Le projet pilote du Connectome a été lancé avec les sept partenaires : DaSCH, FORS, EPFL Blue Brain, eXascale Infolab, SATW,SAGW et SWITCH.<sup>357</sup> Intimement liés avec les principes de l'open science et des données FAIR, les objectifs du projet en 2022 sont notamment :<sup>358</sup>

- Fournir des solutions complètes pour exploiter l'ensemble des données ouvertes au niveau suisse.
- Faire progresser les données ouvertes en récoltant, reliant et harmonisant les métadonnées de plusieurs fournisseurs de données.
- Construire un service de découverte et permettre aux services tiers de tirer parti du graphe de connaissances.

Concernant son fonctionnement, le Connectome récolte les métadonnées issues de réservoirs externes et les stocke en interne. Par contre, les documents eux-mêmes ne sont pas récupérés. Les premiers réservoirs récoltés étaient : SNSF P3 (Swiss National Science Foundation)<sup>359</sup>, Innosuisse Aramis<sup>360</sup>, opendata.swiss<sup>361</sup> et Unpaywall<sup>362</sup> ainsi qu'OpenAIRE<sup>363</sup> et Europeana (ces deux derniers à des fins de test). La liste des fournisseurs s'allonge depuis 2021 avec, entre autres, SWISSUbase<sup>364</sup> et DaSCH<sup>365</sup>.

Au niveau de l'infrastructure technique, les données sont collectées via différents moyens tels que le protocole OAI ou l'usage d'APIs et de dumps<sup>366</sup>. Les données sont ensuite nettoyées et harmonisées sur la base de l'ontologie RESCS<sup>367</sup>, puis enrichies avec Wikidata<sup>368</sup> et finalement

---

<sup>354</sup> Nicolai Hauf et Martin Jaekel, Data reuse in Social Science and Humanities, entretien réalisé par SWITCH, 30 avril 2021, <https://www.switch.ch/stories/Data-reuse-in-Social-Science-and-Humanities/>.

<sup>355</sup> Ibid., 1.

<sup>356</sup> Hauf et al., « Data Reuse in the Social Sciences and Humanities », 1.

<sup>357</sup> Cornelia Puhze, « Le connectome passe en phase pilote », SWITCH, 3 mars 2020, <https://www.switch.ch/fr/stories/Connectome-pilot-phase/>.

<sup>358</sup> « The Connectome Project », SWITCH, consulté le 2 juin 2022, <https://www.switch.ch/connectome/>.

<sup>359</sup> Voir : <https://www.snf.ch/en/Gs0FKuNCqLzQQsjs/news/new-online-database-simplifies-access-to-80-000-projects>

<sup>360</sup> Voir : <https://www.aramis.admin.ch>

<sup>361</sup> Voir : <https://opendata.swiss>

<sup>362</sup> Voir : <https://unpaywall.org>

<sup>363</sup> Voir : <https://www.openaire.eu>

<sup>364</sup> Voir : <https://www.swissubase.ch>

<sup>365</sup> Voir : <https://www.dasch.swiss>

<sup>366</sup> Un dump est un fichier contenant la structure et le contenu de la base de données.

<sup>367</sup> Voir : <https://rescs.org>

<sup>368</sup> Voir : <https://www.wikidata.org>

intégrées au graphe de connaissance du Connectome (usage du framework Blue Brain Nexus<sup>369</sup>).<sup>370</sup> Ce graphe est structuré en LOD.

Le Connectome est un projet encore en développement dont les étapes en cours sont la désambiguïsation des noms d'auteurs et l'ajout de la provenance aux métadonnées. Il est également prévu que l'ontologie RESCS soit alignée sur d'autres ontologies spécifiques aux domaines métier.<sup>371</sup> L'usage de l'intelligence artificielle fait partie intégrante du projet, et elle est notamment utilisée pour analyser, enrichir et organiser les données du graphe.<sup>372</sup>

Enfin, le Connectome offre des APIs<sup>373</sup> très poussées qui permettent d'interroger les données avec beaucoup de précision, ce qui facilite largement la réutilisation des données pour tout type de projet<sup>374</sup>.

#### 4.2.1.3 La plateforme de découverte AskReco

Le Connectome a également pour objectif de mettre à disposition du public, et en particulier des chercheurs, une plateforme qui permettra d'effectuer des recherches dans le graphe de données. AskReco est le prototype de cette plateforme. Il est prévu que les usagers puissent y faire des recherches sur des ensembles de données afin de découvrir des connexions utiles pour de nouvelles recherches.<sup>375</sup>

Plusieurs chercheurs et testeurs pilotes ont aidé à identifier les besoins et les processus de travail en lien avec les données de la recherche.<sup>376</sup> Ainsi, l'interface web permettra prochainement d'identifier des publications avec les ensembles de données qui y sont liées, puis, de retrouver des publications en lien avec ces ensembles.<sup>377</sup> En outre, il sera possible de personnaliser l'algorithme de recherche (notamment concernant les scores de pertinence ainsi que les choix des champs à prendre en considération).<sup>378</sup>

Une plateforme opérationnelle devrait prochainement être accessible, tandis que l'objectif actuel est de continuer à identifier de nouvelles fonctionnalités pour les utilisateurs<sup>379</sup>.

#### 4.2.1.4 Le Connectome et la BCUL

Dans un premier temps, l'objectif de Connectome était de donner accès aux données de la recherche. Aujourd'hui, ce projet s'est étendu aux données issues des institutions GLAM comme les bibliothèques et les archives. Dans ce cadre, la BCUL participera au Connectome en tant que fournisseur de données. Ces données sont pour le moment limitées à un corpus

---

<sup>369</sup> Voir : <https://bluebrainnexus.io>

<sup>370</sup> « An ecosystem of open linked research data creates new opportunities », SWITCH, 2021, <https://www.switch.ch/connectome/serviceproviders/#collapse-2221ffe7-01b3-11ec-a1a3-5254009dc73c-2>.

<sup>371</sup> « An ecosystem of open linked research data creates new opportunities », part. Current Key features for using the Knowledge Graph.

<sup>372</sup> Ibid., part. Connectome AI to enrich value extraction.

<sup>373</sup> Connectome API : <https://connectome-test01-api.switch.ch/v1/swagger>

<sup>374</sup> « An ecosystem of open linked research data creates new opportunities », part. Current Key features for using the Knowledge Graph.

<sup>375</sup> « New powerful ways to search for, discover and organise linked open data », SWITCH, consulté le 12 juillet 2022, <https://www.switch.ch/connectome/researchers/>.

<sup>376</sup> Sebastian Sigloch, « Discover and Use Linked Open Data », SWITCH, 23 novembre 2020, <https://www.switch.ch/stories/Discover-linked-open-data/>.

<sup>377</sup> Ibid.

<sup>378</sup> Ibid.

<sup>379</sup> Ibid.

restreint de documents issus du service des Manuscrits et archives. Des métadonnées au format MARC/ISAD(G) de Patrinum ont été transformées au format RIC-O<sup>380</sup> (Records In Contexts Ontology) afin d'être versées au Connectome (qui dispose déjà d'un mapping entre RIC-O et RESCS). Il est aujourd'hui question que la BCUL fournisse également d'autres données et qu'elle participe aussi en tant qu'utilisatrice des données du Connectome.

Plusieurs idées ont émergé à ce sujet, telles qu'enrichir Patrinum avec certaines données archivistiques du Connectome, ou agréger les données issues des dépôts légaux cantonaux et de la Confédération sur une seule plateforme.<sup>381</sup>

De manière générale il serait nécessaire que d'autres institutions similaires à la BCUL versent leurs données au Connectome pour que des projets de ce type puissent voir le jour. En outre, il serait très utile que le Connectome donne également accès aux images et aux documents multimédias via les APIs IIIF et non uniquement aux métadonnées. Plus particulièrement pour la recherche, l'accès aux textes issus des documents numérisés (reconnaissance OCR, HTR ou OMR pour les partitions) ouvrirait énormément de possibilités. Par exemple, si les textes de Scriptorium étaient disponibles sur le Connectome, il serait possible d'enrichir la base des personnalités vaudoises avec toutes les mentions de l'une ou l'autre de ces personnes dans la presse.<sup>382</sup> De plus, si le Connectome intégrait aussi les textes issus des autres plateformes relatives à la presse, comme E-newspaperarchives, cela donnerait encore plus de pertinences aux données.<sup>383</sup> Ainsi, on pourrait imaginer créer des notices d'autorités directement à partir du texte intégral des articles et retracer le parcours d'un auteur en croisant ses publications et les articles de presse qui le mentionnent.<sup>384</sup>

Comme on le constate, le Connectome ouvrira énormément de possibilités pour la réutilisation des données dès le moment où suffisamment d'institutions y auront publié leurs ressources. Le défi actuel du projet est donc de réussir à intégrer ces données, sachant que cela demande un travail important de transformation, les institutions utilisant souvent des normes différentes, et parfois personnalisées, à l'instar de Patrinum.

### 4.3 Les méta-catalogues internationaux

La BCUL participe aux méta-catalogues internationaux Google books, Wikipédia et Biodiversity Heritage Library (BHL).<sup>385</sup> Il ne s'agit néanmoins pas des seuls catalogues utilisés à l'échelle mondiale où la BCUL est présente. En effet, e-codices publie ses données sur la plateforme Europeana,<sup>386</sup> de même que Memobase.<sup>387</sup> Les méta-catalogues e-rara et e-codices publient leurs données sur Gallica,<sup>388</sup> et BHL publie ses données sur Internet Archive.<sup>389</sup> En outre, Gallica

---

<sup>380</sup> Voir : [https://www.ica.org/standards/RiC/RiC-O\\_v0-2](https://www.ica.org/standards/RiC/RiC-O_v0-2)

<sup>381</sup> Ces idées ont vu le jour lors de la séance « Switch Research Data Connectome – applications à la BCUL » à laquelle ont participé les responsables des collections, IDHEAP, 5 avril 2022.

<sup>382</sup> Ibid.

<sup>383</sup> Ibid.

<sup>384</sup> Ibid.

<sup>385</sup> Voir : <https://www.biodiversitylibrary.org>, description en annexe 9.1.3.

<sup>386</sup> Voir : <https://www.europeana.eu> ; « Portails de manuscrits », e-codices, consulté le 22 juin 2022, [https://www.e-codices.unifr.ch/fr/about/mss\\_portals](https://www.e-codices.unifr.ch/fr/about/mss_portals).

<sup>387</sup> « Participer à Memobase », part. Pourquoi participer à Memobase?

<sup>388</sup> Voir : <https://gallica.bnf.fr> ; « e-codices », Gallica, consulté le 13 juillet 2022, <https://gallica.bnf.fr/edit/und/e-codices> ; « e-rara.ch », Gallica, consulté le 13 juillet 2022, <https://gallica.bnf.fr/edit/und/e-rarach>.

<sup>389</sup> Voir : <https://archive.org> ; « Biodiversity Heritage Library », Internet Archive, 22 juillet 2022, <https://archive.org/details/biodiversity>.

participe de manière conséquente à plusieurs méta-catalogues d'envergure tels que : Europeana (environ 1.3 millions de documents),<sup>390</sup> Internet Archive (environ 4000 documents),<sup>391</sup> Wikisource (près de 8 millions de documents)<sup>392</sup> et Wikimedia Commons (170'000 documents).<sup>393</sup> Comme on le constate, les données publiées par la BCUL sur e-codices et e-rara sont extrêmement diffusées, non seulement de manière directe par ces deux catalogues, mais également indirectement, grâce à Gallica.

De façon générale, être présent sur un méta-catalogue international amène des avantages conséquents. D'une part, cela a un impact important sur la consultation des ressources par le grand public, comme en témoigne les quelques 25 millions d'accès annuels aux documents disponibles sur Google Books.<sup>394</sup> D'autre part, les méta-catalogues sont généralement orientés vers la réutilisation des données et proposent donc des outils accessibles et bien documentés dans ce sens (APIs, IIF, SPARQL, OAI-PMH). Ces éléments font de ces plateformes des ressources largement utilisées dans le cadre des projets en humanités numériques, comme vont le montrer les sous-chapitres suivants.

Afin d'identifier les catalogues internationaux qui pourraient être pertinents concernant les collections patrimoniales de la BCUL, la suite de ce chapitre propose une description sommaire des plateformes incontournables dans ce domaine. Les éléments analysés sont notamment le public cible, l'adéquation par rapport aux principes FAIR et l'existence de projet en humanités numériques basés sur ces sources. Les bibliothèques numériques présentées sont Gallica, Google Arts et Culture, Europeana, Hathi Trust, Internet Archives, Wikimedia Commons et Wikisource (un tableau récapitulatif se trouve en annexe 9.1.4).

### 4.3.1 Gallica

Gallica est le catalogue numérique de la BnF et de ces partenaires (élargi à une dizaine d'institutions externes comme la Bibliothèque et Archives Canada ou la Library of Congress).<sup>395</sup> Gallica compte aujourd'hui plus de 9 millions de ressources de tout type (images, cartes, enregistrements audios, vidéos, partitions, objets, manuscrits, presse, et évidemment livres) sous licence CC BY-NC. La plateforme permet de présenter ses collections au grand public à travers son blog,<sup>396</sup> qui propose notamment des expositions virtuelles (mention « À la loupe »). Elle s'adresse toutefois également aux chercheurs en mettant à disposition une recherche avancée et des filtres pour limiter les résultats. Les métadonnées sont décrites selon les standards Dublin Core et MPEG-21 DIDL qui restent relativement détaillés.<sup>397</sup> Enfin, Gallica se distingue surtout par les nombreux outils proposés pour la réutilisation des données (APIs, SPARQL,<sup>398</sup> IIF et OAI-PMH), par les nombreux documents bénéficiant d'une reconnaissance

---

<sup>390</sup> Voir : recherche National Library of France sur Europeana, <https://www.europeana.eu/fr/search?page=1&qf=PROVIDER%3A%22National%20Library%20of%20France%22&query=&view=grid>

<sup>391</sup> « Gallica », Internet Archive, 11 août 2022, <https://archive.org/details/bnfgallica>.

<sup>392</sup> Voir : <https://fr.wikisource.org> ; « Wikisource:Gallica », Wikisource, 11 juillet 2022, <https://fr.wikisource.org/wiki/Wikisource:Gallica>.

<sup>393</sup> Voir : <https://commons.wikimedia.org> ; « Category:Files from Gallica », Wikimedia Commons, 17 août 2022, [https://commons.wikimedia.org/wiki/Category:Files\\_from\\_Gallica](https://commons.wikimedia.org/wiki/Category:Files_from_Gallica).

<sup>394</sup> Frey, « Googlos – un partenariat public-privé pour la mise en ligne du patrimoine imprimé », part. Bilan.

<sup>395</sup> « Bibliothèque et Archives Canada », Gallica, 11 août 2022, <https://gallica.bnf.fr/edit/und/bibliotheque-et-archives-canada-0> ; « Library of Congress », Gallica, 11 août 2022, <https://gallica.bnf.fr/edit/und/library-congress>.

<sup>396</sup> Voir : <https://gallica.bnf.fr/blog/recherche?query=2860&mode=desktop>

<sup>397</sup> Descriptif du MPEG-21 DIDL : <https://www.bnf.fr/fr/mpeg-21-didl-digital-item-declaration>.

<sup>398</sup> Les données sont proposées au format RDF depuis : <https://data.bnf.fr>.

OCR, et sa collaboration avec tous les autres méta-catalogues choisis pour cette analyse, à l'exception de Google Arts and Culture. La plateforme répond aux principes FAIR à l'exception de la licence, qui n'est pas renseignée de manière explicite dans les métadonnées.

### 4.3.2 Google Arts and Culture

Google Arts and Culture<sup>399</sup> est un projet dont l'objectif est de créer une sorte de musée virtuel en ligne donnant accès à la culture au niveau mondial. Les premières institutions à y participer étaient donc des musées. Néanmoins un certain nombre de bibliothèques possédant des fonds iconographiques l'on rejoint, à l'instar de la British Library.<sup>400</sup> Cette plateforme propose des outils très variés pour aider l'utilisateur à découvrir les collections, tels que des jeux, des outils de zoom immersifs dans une œuvre (nommés « Art up close »)<sup>401</sup> ou encore des visites virtuelles de nombreux bâtiments. Il n'y a par contre pas d'option de recherche avancée disponible, ni de solution pour filtrer les résultats. De plus, les métadonnées sont peu détaillées. La plateforme n'est donc vraiment pas orientée pour la recherche.

Des APIs<sup>402</sup> permettent d'accéder aux données et les applications qu'elles permettent de développer peuvent être mises à disposition sur la plateforme Experiments with Google.<sup>403</sup> Les documents présents sur la plateforme sont sous licence libre ou affranchis,<sup>404</sup> mais aucune licence claire ne précise le cadre de leur réutilisation.

### 4.3.3 Europeana

La plateforme Europeana vise à rendre disponible le patrimoine culturel européen. Comme Google Arts and Culture, elle regroupe avant tout des ressources iconographiques, soit plus de 30 millions d'images. Les textes sont la 2<sup>ème</sup> ressource la plus importante, avec environ 22 millions de documents, puis viennent les enregistrements audios, les vidéos et finalement les objets 3D.<sup>405</sup> Les documents disponibles sont sous licence CC0 ou dans le domaine public et la licence est explicitement renseignée dans les métadonnées.

Comme Gallica, Europeana propose des expositions virtuelles et des articles de blog.<sup>406</sup> Le méta-catalogue n'est pas directement ouvert au crowdsourcing mais des plateformes annexes comme Crowdheritage<sup>407</sup> offrent la possibilité de lancer des campagnes de participation citoyenne sur une thématique précise. L'interface ne dispose pas d'une option de recherche avancée, néanmoins de nombreux filtres permettent de raffiner les résultats. En outre, si le format des métadonnées est spécifique à Europeana, il reste relativement détaillé et très bien

---

<sup>399</sup> Pour Google Arts and Culture, voir : <https://artsandculture.google.com>.

<sup>400</sup> Voir : <https://artsandculture.google.com/partner/the-british-library>.

<sup>401</sup> Voir un exemple : <https://artsandculture.google.com/story/mQRcpFw21RhJw>

<sup>402</sup> Voir : <https://github.com/googleartsandculture>

<sup>403</sup> Voir : <https://experiments.withgoogle.com/collection/arts-culture>

<sup>404</sup> « Frequently asked questions - Cultural Institute Platform Help », Cultural Institute Platform Help, 8 août 2022, <https://support.google.com/culturalinstitute/partners/answer/6002688?hl=en#zippy=%2Chow-about-copyrighted-content>.

<sup>405</sup> « About Us », Europeana, consulté le 14 juin 2022, <https://www.europeana.eu/en/about-us>.

<sup>406</sup> Voir : <https://www.europeana.eu/fr/stories>

<sup>407</sup> Voir : <https://crowdheritage.eu/en>

documenté.<sup>408</sup> De manière comparable à Gallica, les données sont accessibles via plusieurs canaux (APIs, IIF et SPARQL)<sup>409</sup> et leur utilisation est visible à travers de nombreux projets.<sup>410</sup>

Europeana est pleinement compatible avec les principes FAIR.<sup>411</sup> De plus, des mesures ont récemment été définies afin d'améliorer la qualité des métadonnées.<sup>412</sup>

### 4.3.4 Hathi Trust

Hathi Trust<sup>413</sup> est un méta-catalogue qui regroupe quelques 17,6 millions de ressources textuelles dont environ 7 millions sont dans le domaine public.<sup>414</sup> Le nombre d'ouvrages francophones est par contre limité à 646'000 documents. La plateforme a comme objectif de préserver à long terme ces documents ainsi que de les rendre accessibles pour la recherche scientifique.<sup>415</sup> Les ressources agrégées sont issues de Google, Internet Archive, Microsoft ainsi que des institutions membres, ces dernières étant principalement basées aux USA.<sup>416</sup>

Ce méta-catalogue s'adresse essentiellement aux chercheurs des institutions affiliées. Les documents proposés bénéficient d'une reconnaissance OCR et des outils de recherche avancés sont disponibles. Les métadonnées sont au format MARC.

Concernant l'accessibilité des ressources, des restrictions importantes sont malheureusement présentes concernant le téléchargement des documents (il faut procéder page par page) pour le public externe. De plus, certains documents peuvent ne pas du tout être accessibles pour ces usagers. La licence est renseignée dans les métadonnées de chaque document et une page explicative précise le cadre de la réutilisation.<sup>417</sup>

Les données sont disponibles via APIs et OAI-PMH.<sup>418</sup> De plus, des outils spécifiquement développés par le Hathi Trust Research Center (HTRC) permettent de sélectionner des corpus de données et offrent des outils d'analyse et de visualisation.<sup>419</sup> La plupart de ces outils sont accessibles aux chercheurs non affiliés et peuvent être utilisés librement à des fins de recherche.<sup>420</sup>

Hathi Trust est assez proche des principes FAIR, mais n'utilise pas d'identifiants pérennes.

---

<sup>408</sup> Le format Europeana Data Model (EDM) est utilisé. Voir : <https://pro.europeana.eu/page/edm-documentation>.

<sup>409</sup> Voir : <https://pro.europeana.eu/page/apis>

<sup>410</sup> Voir : <https://pro.europeana.eu/page/generic-services-projects>

<sup>411</sup> Antoine Isaac et Nuno Freire, « Europeana and the FAIR Principles for Research Data », Europeana Pro, 1 février 2022, <https://pro.europeana.eu/post/europeana-and-the-fair-principles-for-research-data>.

<sup>412</sup> Ibid., part. What will Europeana do in the near future?

<sup>413</sup> Voir : <https://www.hathitrust.org>

<sup>414</sup> « Welcome to HathiTrust! », HathiTrust Digital Library, consulté le 6 juillet 2022, <https://www.hathitrust.org/about>.

<sup>415</sup> « Our Digital Library », HathiTrust Digital Library, consulté le 10 août 2022, [https://www.hathitrust.org/digital\\_library](https://www.hathitrust.org/digital_library).

<sup>416</sup> « Our Digital Library ».

<sup>417</sup> Voir : [https://www.hathitrust.org/access\\_use](https://www.hathitrust.org/access_use)

<sup>418</sup> Voir : APIs, <https://www.hathitrust.org/data>; OAI-PMH, <https://www.hathitrust.org/oai>

<sup>419</sup> « HTRC Access and Use », HathiTrust Digital Library, consulté le 9 juillet 2022, [https://www.hathitrust.org/htrc\\_access\\_use](https://www.hathitrust.org/htrc_access_use).

<sup>420</sup> Ibid.

### 4.3.5 Internet Archive

Internet Archive est une plateforme mondiale dont le but est de sauvegarder la connaissance et d’y donner un accès universel à tous les publics.<sup>421</sup> Elle compte plus de 726 milliards de pages internet, 35 millions de ressources textuelles, 14,5 millions d’enregistrements audios, 8,4 millions de fichiers vidéo et 4,4 millions d’images.<sup>422</sup>

La plateforme est dotée d’une recherche simple agrémentée de filtres. Toutefois, le nombre d’entrée pour chaque filtre rend ces derniers difficilement exploitables. Les métadonnées, de leur côté, sont dans un format spécifique et relativement peu détaillé.<sup>423</sup> En outre, la plateforme contient un nombre de ressources tellement important que ces dernières peuvent difficilement être retrouvées via l’interface web. Pour ces raisons, la plateforme en ligne semble difficilement utilisable pour les chercheurs.

Internet Archive est complètement ouvert au crowdsourcing en permettant notamment aux utilisateurs d’ajouter du contenu. Le projet Building Library Together a pour but d’axer encore plus le catalogue sur la communauté, en améliorant les outils pour le téléversement, la description et l’organisation des ressources.<sup>424</sup> Par ce projet, le méta-catalogue souhaite se positionner comme un Wikipédia dédié aux collections médiatiques.<sup>425</sup>

Internet Archive a lancé en 2006 le projet Open library,<sup>426</sup> qui met à disposition 2 millions de livres électroniques, y compris des livres récents auxquels le droit d’auteur s’applique. Les bibliothèques participantes ont sélectionné des documents numériques dans leurs collections et les ont mis à disposition sur la plateforme.<sup>427</sup> Les lecteurs peuvent se créer un compte et emprunter gratuitement ces livres<sup>428</sup> pour une consultation en ligne ou hors ligne.<sup>429</sup> Les ressources dans le domaine public peuvent être consultées et téléchargées sans compte lecteur.<sup>430</sup> En outre, il s’agit d’une bibliothèque en ligne participative où tout utilisateur peut ajouter un document.<sup>431</sup> Enfin, ce projet a également pour but de répertorier tous les livres publiés en mettant à disposition une page web pour chacun d’entre eux (plus de 20 millions de livres disposent actuellement d’une page).<sup>432</sup>

De manière générale, Internet Archive partage ses données via des APIs dédiées ainsi que IIIF.<sup>433</sup> De plus, la plupart des livres numérisés disposent d’une reconnaissance OCR.<sup>434</sup> Les données sont donc accessibles à la réutilisation et plusieurs projets ayant recours aux

---

<sup>421</sup> « About the Internet Archive », Internet Archive, 14 juin 2022, <https://archive.org/about/>.

<sup>422</sup> « Internet Archive: Digital Library of Free & Borrowable Books, Movies, Music & Wayback Machine », consulté le 24 février 2022, <https://archive.org/>.

<sup>423</sup> Pour la description des métadonnées, voir : <https://internetarchive.readthedocs.io/en/stable/metadata.html>.

<sup>424</sup> « Internet Archive: Projects », Internet Archive, part. Building Library Together, consulté le 14 juin 2022, <https://archive.org/projects/>.

<sup>425</sup> Ibid.

<sup>426</sup> Voir : <https://openlibrary.org>

<sup>427</sup> « Borrowing Books Through Open Library », Open Library, 1 avril 2022, part. How does borrowing a book work through Open Library?, <https://openlibrary.org/help/faq/borrow>.

<sup>428</sup> Les ouvrages ne peuvent être consultés que par une seule personne à la fois.

<sup>429</sup> « Borrowing Books Through Open Library », part. How does borrowing a book work through Open Library?

<sup>430</sup> Voir un exemple : [https://openlibrary.org/works/OL1100007W/Le\\_tour\\_du\\_monde\\_en\\_quatre-vingts\\_jours?edition=ia%3ALeTourDuMondeEnQuatreVingtsJours](https://openlibrary.org/works/OL1100007W/Le_tour_du_monde_en_quatre-vingts_jours?edition=ia%3ALeTourDuMondeEnQuatreVingtsJours).

<sup>431</sup> « Internet Archive: Projects », part. Open Library.

<sup>432</sup> Ibid.

<sup>433</sup> Voir : APIs, <https://blog.archive.org/developers/> ; IIIF, <https://iiif.archivelab.org/iiif/documentation>.

<sup>434</sup> Voir documentation : <https://archive.org/services/docs/api/ocr.html>.

humanités numériques les exploitent.<sup>435</sup> Du côté FAIR, la licence est spécifiée dans les métadonnées, de même que la source. En revanche, les documents ne disposent pas d'un identifiant pérenne.

### 4.3.6 Wikimedia Commons et Wikisource

Wikimedia Commons (ou simplement Commons) héberge le contenu multimédia de l'ensemble des plateformes de la Fondation Wikimedia.<sup>436</sup> Par extension, c'est devenu une médiathèque en ligne qui va au-delà des projets Wikimedia, car tout document sous licence libre et possédant un intérêt encyclopédique ou éducatif peut y être déposé.<sup>437</sup> Aujourd'hui, elle regroupe plus de 85 millions de fichiers<sup>438</sup> dont plus de 77 millions d'images.<sup>439</sup>

De son côté, Wikisource est une bibliothèque numérique multilingue contenant des textes de tout type (littérature, poésie, etc.) qui ont en principe fait l'objet d'une publication papier<sup>440</sup> et sont compatibles avec la licence CC BY-SA.<sup>441</sup> Les documents sont au format DjVu,<sup>442</sup> soit permettant la recherche plein texte ainsi que l'export aux formats EPUB, MOBI et PDF. La plateforme compte notamment quelques 347'000 textes francophones et 560'000 textes anglophones.<sup>443</sup>

Comme les autres projets Wikimedia, ces deux plateformes reposent sur le crowdsourcing. En particulier Wikisource a besoin de sa communauté pour corriger les textes OCR de ses ressources.

Des APIs permettent de réutiliser les données des deux plateformes. De plus, Commons assure la mise à disposition des documents via IIIF.<sup>444</sup> Toutefois, la vraie force de ces deux plateformes repose sur leurs liens avec Wikipédia et Wikidata<sup>445</sup> ainsi que DBpedia<sup>446</sup> (seul projet à être externe à la Fondation Wikimédia). En particulier, les plateformes Wikidata et DBpedia permettent d'accéder aux données de Commons et Wikisource au format RDF via un point de terminaison SPARQL.<sup>447</sup> Les données de Wikipédia et Wikidata sont largement réutilisées dans une variété d'applications permettant l'analyse et la visualisation de corpus de données.<sup>448</sup>

---

<sup>435</sup> Voir quelques projets récents : <http://blog.archive.org/2022/02/09/library-as-laboratory-a-new-series-exploring-the-computational-use-of-internet-archive-collections/>.

<sup>436</sup> « Wikimedia Commons », in *Wikipédia*, 23 juin 2022, [https://fr.wikipedia.org/w/index.php?title=Wikimedia\\_Commons&oldid=194771531](https://fr.wikipedia.org/w/index.php?title=Wikimedia_Commons&oldid=194771531).

<sup>437</sup> Ibid.

<sup>438</sup> Commons regroupe notamment des photographies, des partitions, des cartes géographiques, des vidéos ou encore des objets 3D.

<sup>439</sup> « Media Statistics », Wikimedia Commons, 13 août 2022, <https://commons.wikimedia.org/wiki/Special:MediaStatistics>.

<sup>440</sup> Des exceptions sont appliquées pour les documents issus de fonds d'archives notamment.

<sup>441</sup> « Wikisource », in *Wikipédia*, 27 juillet 2022, <https://fr.wikipedia.org/w/index.php?title=Wikisource&oldid=195637185>.

<sup>442</sup> « DjVu », in *Wikipédia*, 3 mai 2022, <https://fr.wikipedia.org/w/index.php?title=DjVu&oldid=193380438>.

<sup>443</sup> Voir : page d'accueil en français, <https://fr.wikisource.org/wiki/Wikisource> ; page d'accueil en anglais, [https://en.wikisource.org/wiki/Main\\_Page](https://en.wikisource.org/wiki/Main_Page).

<sup>444</sup> Voir : APIs mediawiki, [https://www.mediawiki.org/wiki/API:Main\\_page](https://www.mediawiki.org/wiki/API:Main_page) ; IIIF, <https://iiif.io/guides/guides/wikipedia.org>.

<sup>445</sup> Voir : <https://www.wikidata.org>

<sup>446</sup> Voir : <https://www.dbpedia.org>

<sup>447</sup> La différence principale entre les deux plateformes est que DBpedia extrait des données structurées de Wikipédia et les publie au format RDF, tandis que Wikidata est nativement en RDF et sert de source aux différentes applications de la Fondation Wikimedia dont Wikipédia, Commons et Wikisource.

<sup>448</sup> Voir : <https://seealso.org> ; [https://www.wikidata.org/wiki/Wikidata:Tools/Visualize\\_data](https://www.wikidata.org/wiki/Wikidata:Tools/Visualize_data)

## 4.4 Synthèse

Comme l'a montré ce chapitre, de nombreuses options sont disponibles pour améliorer la mise à disposition des ressources de la BCUL. Ces solutions agissent à des niveaux complémentaires, car elles améliorent respectivement l'accessibilité des données à l'échelle régionale, suisse et mondiale.

L'optimisation de Scriptorium et Patrinum permettrait par exemple à un chercheur de l'UNIL, souhaitant récupérer toutes les photographies de Lausanne entre 1840 et 1950 sous licence libre, d'accéder facilement à ces ressources sans avoir de demande à effectuer. De plus, cela améliorerait également sur le long terme la situation des professionnels de la BCUL, qui n'auraient alors plus besoin d'extraire eux-mêmes les données pour répondre à ce type de demande. Enfin, du moment que toutes les ressources (documents et métadonnées) ne sont pas mises à disposition sur d'autres catalogues, ce qui aurait probablement un coût trop élevé, optimiser les plateformes internes est le seul moyen pour permettre la réutilisation de l'ensemble des données produites par la BCUL.

On peut par contre mentionner que l'implémentation de toutes les solutions proposées au chapitre 4.1 n'est pas forcément utile. Ainsi, la mise à disposition des métadonnées via OAI-PMH, y compris celles qui ne disposent pas d'un document numérique, pourrait suffire à compléter l'implémentation des manifestes IIF (qui concernent uniquement les documents numériques et leurs métadonnées).

Concernant la publication des données sur les méta-catalogues, une présence plus importante de la BCUL apporterait plusieurs avantages :

1. Les collections seraient plus visibles et accessibles pour le grand public
2. Les documents de la BCUL contribueraient à créer des collections plus riches et donc porteuses de plus de sens pour la recherche.
3. Les utilisateurs pourraient se réapproprier et réutiliser les ressources de la BCUL dans le cadre des catalogues ouverts au crowdsourcing.
4. Les données de la BCUL pourraient être exploitées via des outils que l'institution ne peut pas mettre en place à elle seule.
  - a. Les données pourraient être intégrées dans un graphe LOD et mises en relation avec d'autres données au niveau national et international
  - b. Des outils de recherches avancés seraient disponibles pour l'extraction de corpus pointus de données, par exemple un point de terminaison SPARQL.
  - c. Les ressources de la BCUL seraient automatiquement incluses dans les projets d'humanités numériques pour lesquels elles ont de la pertinence, contrairement à la situation actuelle où les responsables de collections doivent dans certains cas activement proposer leurs ressources (par exemple l'EPFL n'a pas fait d'office appel à la BCUL dans le cadre du projet Distant listening alors que des collections adaptées étaient disponibles).

Aucuns des méta-catalogues décrits au chapitre 4.3 n'est capable de répondre de manière optimale à tous les points susmentionnés en même temps. Par contre, certains d'entre eux permettent des compromis intéressants. De plus, il est également envisageable de participer à plusieurs d'entre avec des ressources choisies.

La plateforme Google Arts and Culture serait en parfaite adéquation avec le point 1, mais ne répondrait pas du tout aux autres. Participer à ce projet serait cependant intéressant pour valoriser des ressources à la fois esthétiques et bien documentées via des outils comme « Art up close ».

Europeana et Gallica sont des plateformes qui ne répondent pas aussi bien que Google Arts and Culture au point 1 mais restent quand même adaptées au grand public et ont l'avantage de répondre également aux autres points (à l'exception du 3), car elles proposent toutes les deux de nombreuses modalités de partage dont IIIF et SPARQL. Néanmoins, comme démontré au chapitre 4.3, les données présentes sur e-codices et e-rara sont déjà publiées sur ces deux catalogues. Aussi, une solution pourrait être de publier davantage de données sur e-codices et e-rara et de conclure un accord avec Gallica pour y publier des ressources qui n'entrent pas dans le périmètre de ces deux catalogues.<sup>449</sup>

En outre, participer à Wikisource et Commons (par l'intermédiaire de Gallica ou non) permettrait également de mettre à disposition des données au format LOD sur des plateformes largement réutilisées telles que Wikidata et DBpedia. De plus, l'écosystème Wikimédia étant particulièrement adaptés au crowdsourcing, y participer permettrait de répondre pleinement au point 3.

Concernant la plateforme Internet Archive, elle présente des avantages similaires aux projets Wikimédia, mais sans que les données ne soient structurées en RDF. Le projet Open Library serait par contre intéressant pour mettre à disposition des ouvrages imprimés adaptés à la lecture de loisir, car l'interface est plus conviviale pour cet usage que celles de Google Books, Wikisource ou e-rara. Quant à Hathi Trust, les outils de recherche et d'analyse proposés par le Hathi Trust Research Center constituent la plus-value principale de cette plateforme. Néanmoins, cela ne représente pas un avantage significatif par rapport à Gallica, qui pourrait accueillir les mêmes types de documents, mais dans un environnement majoritairement francophone et en offrant une mise à disposition LOD.

Pour terminer, il est utile de souligner que les objectifs du projet Connectome sont intimement liés à ceux proposés dans ce travail. Ce projet est donc probablement celui qui pourra, à terme, le mieux répondre aux points 2 et 4. Pour rappel, le Connectome devrait permettre d'harmoniser et interconnecter dans un graphe LOD l'ensemble des données ouvertes au niveau suisse. C'est donc sur cette plateforme que les données de la BCUL devraient pouvoir être le mieux interconnectées avec les ressources des institutions GLAM sur l'ensemble du territoire national. Or, un projet de ce type peut offrir des possibilités d'analyses innombrables.<sup>450</sup> De plus, le public ciblé étant précisément les chercheurs, et l'interface AskReco étant spécifiquement conçue à leur intention, ce projet devrait démultiplier l'utilisation des données qui seront mises à disposition pour la recherche.

---

<sup>449</sup> Europeana demande aux institutions qui souhaitent participer de passer par des agrégateurs dont Gallica fait partie.

<sup>450</sup> Par exemple, le graphe du Connectome devrait permettre d'analyser et de visualiser avec facilité la production patrimoniale vaudoise par rapport au reste de la Suisse.

## 5 Conclusion

Ce travail a permis de mettre évidence les lacunes dans la mise à disposition des données par les institutions patrimoniales, dans le sens où ces données ne sont pas exploitables dans le cadre de projets de recherche en lien avec les humanités numériques. Ces lacunes posent donc problème pour la réutilisation des ressources patrimoniales dans la sphère de la recherche et handicapent le développement de la connaissance dans les sciences humaines (1.2).

L'état des collections patrimoniales numériques de la BCUL et des plateformes utilisées pour les diffuser a permis de faire ressortir la richesse et la diversité des fonds ainsi que la pluralité des plateformes utilisées. 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 (2.2.8). Ces problématiques constituent un premier blocage pour l'accessibilité des ressources patrimoniales.

Les différentes plateformes utilisées pour diffuser ces ressources ont été analysées en regard des principes FAIR, l'objectif de ces derniers étant précisément de donner un cadre à la publication des données dans le but de faciliter leur réutilisation par les machines (1.1.3). Cette analyse a montré que les méta-catalogues d'envergure nationale, tels que e-codices, e-rara, RISM Suisse ou CCSA, respectent, ou sont proches de respecter, ces principes (2.2.8). En revanche, les catalogues maintenus en interne par la BCUL, soit Scriptorium et Patrinum, nécessiteraient plusieurs adaptations pour pouvoir être labellisés FAIR (2.1.3.1). Or, s'il ne fait pas sens de revenir sur certains choix techniques qui contrarient ces principes (par exemple l'adaptation du format MARC à la norme ISAD(G) implémenté dans Patrinum), plusieurs possibilités d'amélioration ont été identifiées pour favoriser l'accessibilité et l'interopérabilité des données (4.1).

L'état des lieux, qui inclut une comparaison entre la BCUL et la BGE, a également permis de montrer que les deux institutions sont dans une situation très proche. Toutes deux maintiennent deux portails en interne pour les collections patrimoniales qui ne permettent actuellement pas la réutilisation des données. De plus, elles publient sur des catalogues externes qui sont relativement similaires (2.5). On peut toutefois noter que la BGE publie un plus grand nombre de documents sur ces plateformes, ce qui tend plus à compenser les manques des portails internes.

Le chapitre 3 de ce travail opère un focus sur le patrimoine relatif à la musique écrite, afin de fournir une analyse plus détaillée du potentiel innovant relatif à ce type de document (3.2). Ces collections sont particulières, car leur exploitation par la machine requiert un écosystème technologique qui diffère des ressources textuelles. Pour cette raison, une partie de ce chapitre est consacrée à décrire les différentes technologies nécessaires à l'exploitation de la musique écrite (3.3). Par la suite, différents projets innovants sont identifiés. On peut noter qu'il s'agit dans la plupart des cas de prototypes, d'outils en cours de développement ou des projets de recherche et non de produits finis pleinement exploitables (3.4). Cet état de fait tient aux difficultés techniques, notamment au niveau de la reconnaissance OMR (3.3.6), qui empêchent encore de pleinement exploiter le potentiel de ces ressources (3.6.1). Plusieurs solutions sont néanmoins évoquées afin d'améliorer l'accessibilité et de permettre la réutilisation des collections de la BCUL concernées (3.6.2).

Le dernier chapitre se rapporte à la réutilisation du patrimoine numérique de la BCUL au sens large et s'attache à proposer des améliorations englobant l'ensemble de ces collections. Les propositions consistent notamment à : rendre plus interopérables les plateformes internes (4.1), publier plus de données sur les méta-catalogues suisses (4.2), et finalement à participer à des méta-catalogues de plus grande envergure (4.3). Ces propositions sont complémentaires, car elles rendent accessibles les données à des échelles différentes. Toutefois, il est peu probable que la BCUL dispose des ressources pour toutes les implémenter, le chapitre 4.4 met donc en évidence différentes solutions en précisant à quels besoins chacune d'entre elles peut répondre.

Pour revenir aux objectifs mentionnés en introduction, ce travail a largement décrit par quels moyens techniques les données de la BCUL pouvaient être mises à disposition, afin d'améliorer leur interopérabilité ainsi que pour permettre leur réutilisation en se rapprochant des principes FAIR (4).

L'utilité de cette mise à disposition est démontrée à travers le prisme de la musique écrite. À cet effet, le chapitre 3 présente les dernières avancées de l'intelligence artificielle dans le domaine de la musique, ce qui atteste des applications concrètes qui émanent de l'avancée de la recherche. Puis, ce chapitre détaille les formats de données et les projets qui permettent actuellement le développement de la connaissance en lien avec la musique écrite. On constate ainsi que les chercheurs ont besoin d'accéder à des corpus de données dans des formats exploitables par la machine, par exemple MEI ou MusicXML dans le cas des partitions. Ces ressources sont nécessaires pour permettre l'analyse des données par des algorithmes toujours plus complexes, tels que des réseaux de neurones.

La mise à disposition de grands sets de données est fondamentale pour permettre aux réseaux de neurones d'obtenir des résultats satisfaisants, ce qui implique de participer à des méta-catalogues réunissant un grand nombre de documents. De plus, les laboratoires de recherche spécifiques à un domaine sont généralement peu nombreux au niveau national, il est donc nécessaire de s'adresser aux chercheurs à l'échelle mondiale pour que tous les spécialistes d'un domaine aient accès aux ressources. Le partage des données sur des méta-catalogues internationaux apporte donc un grand plus à la recherche scientifique.

## 5.1 Discussion

On constate aujourd'hui un certain décalage entre les objectifs couramment mis en avant en bibliothèques, telles que la qualité des métadonnées, et les besoins concrets des chercheurs en humanités numériques. En effet, la pratique en bibliothèque est souvent de préférer ne donner un accès qu'aux collections parfaitement cataloguées, y compris si cela limite le nombre de ressources disponibles. De plus, les documents publiés sur les méta-catalogues sont souvent peu nombreux, car la publication sur ces plateformes a un coût annuel, généralement calculé en fonction du nombre de documents hébergés.

Ces éléments favorisent un partitionnement des collections numériques, qui restent ainsi majoritairement hébergées sur des catalogues internes, qui ne disposent donc pas des dernières technologies. Ainsi, seules les œuvres ayant une haute valeur patrimoniale sont publiées sur des catalogues internationaux. Or, si cette logique fonctionne lorsque l'on s'adresse au grand public, elle n'est souvent plus adéquate pour faire progresser la recherche

actuelle. En effet, si les chercheurs qui travaillent sans avoir recours aux humanités numériques n'ont besoin que des métadonnées, car on s'attend à ce qu'ils se déplacent pour consulter la ressource physique, ce n'est plus le cas des chercheurs qui travaillent en analysant des corpus de données.

Plus spécifiquement, les chercheurs ayant recours aux technologies des humanités numériques ont besoin de grands corpus de données dans des formats exploitables. Le niveau de description des métadonnées ne revêt alors plus une utilité fondamentale, car ces dernières ont pour principale fonction de permettre la sélection d'un set. Un format de métadonnées interopérables avec d'autres catalogues devient ainsi plus utile qu'un format très détaillé spécifique à un type de document.

Or, l'usage d'outils reposant sur les humanités numériques est de plus en plus courant, car la quantité de données à analyser devient trop conséquente pour un travail manuel. Ainsi, les métadonnées servent exclusivement à extraire des corpus, tandis que le contenu du document lui-même qui devient essentiel à l'analyse, qui est produite, non plus par l'humain, mais par un algorithme.

Évidemment ces nouveaux usages sont sujets à débats, et posent notamment des problèmes éthiques, car les intelligences artificielles qui analysent les données sont biaisées en fonction des données qu'on leur fournit. Par exemple, si on demande à une IA de créer une image représentant un médecin, elle va très probablement représenter un homme, car les images de femme médecin sont moins nombreuses sur internet et risquent d'être sous représentées dans le set qui aura entraîné cette IA. Il s'agit néanmoins d'une problématique qui va au-delà du périmètre de ce travail.

Pour revenir aux bibliothèques faces aux humanités numériques, et à la problématique du coût induit par le versement des objets numériques, comme les images en hautes définitions, sur les méta-catalogues, une solution, qui tend à se généraliser, est l'usage des APIs IIIF. Cette technologie est en effet parfaitement adaptée pour dépasser cette difficulté, puisqu'elle permet de rendre disponibles les images (et les ressources qui y sont liées, dont les annotations et surtout les fichiers exploitables obtenus par OCR ou OMR) sur des méta-catalogues sans que ceux-ci aient à les héberger. Cette solution a par contre le défaut d'obliger les institutions à fournir un travail de leur côté, pour mettre à disposition leurs données sous ce format. IIIF n'est donc pas encore une technologie suffisamment présente dans les bibliothèques numériques institutionnelles, comme en témoignent les plateformes de la BCUL et de la BGE, qui sont en train de la mettre en place sur leurs plateformes respectives. De plus, il est également indispensable que des méta-catalogues agrègent des manifestes IIIF de chaque institution locale, pour une mise à disposition centralisée et représentative à destination des chercheurs.

Concernant le format des métadonnées, les technologies LOD apportent également une solution concrète, car elles peuvent être appliquées à des données sources obtenues dans des formats variés, ce qui permet donc de ne pas sacrifier la qualité des métadonnées à leur interopérabilité. De plus, les humanités numériques ayant tendance à promouvoir les recherches interdisciplinaires, le format RDF est donc particulièrement adapté pour connecter des données issues de différents domaines. Si le concept de linked data remonte aux années 2000, on constate aujourd'hui qu'il se répand. Les grands méta-catalogues analysés dans ce

travail démontrent d'ailleurs ce fait, car RISM, Memobase, Connectome, Europeana, Gallica et Wikidata l'utilisent.

Ainsi un projet comme SIMSSA database (3.4.6.1) est représentatif de ce travail, car il devrait reposer sur les technologies IIIF et LOD, ce qui répond parfaitement aux nouveaux besoins de la recherche en humanités numériques, tout en étant en adéquation avec les moyens techniques et financiers des bibliothèques.

## 6 Bibliographie

### 6.1 Contexte général

Béquet, Gaëlle. Trois bibliothèques européennes face à Google - Aux origines de la bibliothèque numérique (1990-2010). Publications de l'Écoles des Chartes, 2015.

Collectif. Connaître et valoriser la création littéraire numérique en bibliothèque. Presses de l'enssib, 2019.

Dalbin, Sylvie, Emmanuelle Bermès, Antoine Isaac, Romain Wenz, Yann Nicolas, Tayeb Merabti, Anila Angjeli, et al. « Approches documentaires : priorité aux contenus ». Documentaliste-Sciences de l'Information 48, no 4 (2011): 42-59. <https://doi.org/10.3917/docsi.484.0042>.

Dempsey, Lorcan. « Library Collections in the Life of the User: Two Directions ». LIBER Quarterly: The Journal of the Association of European Research Libraries 26, n° 4 (2017): 338-59. <https://doi.org/10.18352/lq.10170>.

Granjon, Fabien. « Présentation du dossier : Les sciences humaines et sociales au prisme du digital turn ». Variations. Revue internationale de théorie critique, n° 19 (7 avril 2016). <https://doi.org/10.4000/variations.726>.

Martin, Chloé. Le guide des bibliothèques numériques: le guide essentiel des s@voirs numérisés du monde. FYP editions, 2010.

Mion Mouton, Fanny. « Bibliothèque numériques et coopération : comparaisons internationales ». Mémoire, Université de Lyon, 2013. <https://www.enssib.fr/bibliotheque-numerique/documents/60377-bibliotheques-numeriques-et-cooperation-comparaisons-internationales.pdf>.

Mkadmi, Abderrazak, et Imad Saleh. Bibliothèque numérique et recherche d'informations. Hermès Science publications, 2008.

Navarrete, Trilce. « Benefits of collaborative digitization projects in Europe. A dutch perspective ». Les Cahiers du numérique 11, no 1 (2015): 41-56. <https://www.cairn.info/revue-les-cahiers-du-numerique-2015-1-page-41.htm>.

Papy, Fabrice, et Cyril Jakubowicz. Bibliothèque numérique et innovation: 2ème édition revue et augmentée. ISTE Group, 2021.

Pastore, Graziella. « Les coopérations entre chercheurs et bibliothécaires dans le cadre des projets de numérisation de corpus documentaires ». Mémoire de diplôme de conservateur de bibliothèques, Université de Lyon, 2018. <https://www.enssib.fr/bibliotheque-numerique/documents/68136-les-cooperations-entre-chercheurs-et-bibliothe-caires-dans-le-cadre-des-projets-de-numerisation-de-corpus-documentaires.pdf>.

## 6.2 Domaine musical

Alfaro-Contreras, María, David Rizo, Jose M Iñesta, et Jorge Calvo-Zaragoza. « OMR-Assisted Transcription: A Case Study with Early Prints », 2021, 7.

Anh Nguyen, Minh. « Updates to RODAN Gamera Interactive Classifier ». Montreal, Canada, 28 mai 2018. <https://simssa.ca/blog/Workshop-on-SIMSSA-XIV/>.

Assayag, Gérard. Du signal temps réel à l'intelligence digitale (Recherche musicale & intelligence digitale) | partie 4. Entretien réalisé par Dominique Moulon. MOOC Digital Paris, décembre 2017. <https://moocdigital.paris/cours/recherche-musicale-intelligence-digitale/signal-temps-reel-lintelligence-digitale>.

Assayag, Gérard. Entre espaces acoustiques et cognitifs (Recherche musicale & intelligence digitale) | Partie 5. Entretien réalisé par Dominique Moulon. MOOC Digital Paris, décembre 2017. <https://moocdigital.paris/cours/recherche-musicale-intelligence-digitale/entre-espaces-acoustiques-cognitifs>.

Assayag, Gérard. Évolution musicale : entre art et technologies (Recherche musicale & intelligence digitale) | partie 3. Entretien réalisé par Dominique Moulon. MOOC Digital Paris, décembre 2017. <https://moocdigital.paris/cours/recherche-musicale-intelligence-digitale/evolution-musicale-entre-art-technologies>.

Assayag, Gérard. Geste, voix et corporéité (Recherche musicale & intelligence digitale) | partie 6. Entretien réalisé par Dominique Moulon. MOOC Digital Paris. Consulté le 26 juillet 2022. <https://moocdigital.paris/cours/recherche-musicale-intelligence-digitale/geste-voix-corporeite>.

Bell, Eamonn, et Laurent Pugin. « Heuristic and Supervised Approaches to Handwritten Annotation Extraction for Musical Score Images ». *International Journal on Digital Libraries* 20, n° 1 (1 mars 2019): 49-59. <https://doi.org/10.1007/s00799-018-0249-7>.

- Bhend, Bettina. « Mit Deep Learning zum digitalen Notenständer ». Naratek, 11 décembre 2019. <https://www.naratek.com/de/2019/mit-deep-learning-zum-digitalen-notenstaender>.
- Bishop, Gail. « Music Datasets for Machine Learning ». *Medium* (blog), 12 mars 2021. <https://gail-bishop.medium.com/music-datasets-for-machine-learning-a6cd8d707340>.
- Calvo-Zaragoza, Jorge, Jan Hajič Jr., et Alexander Pacha. « Understanding Optical Music Recognition ». *ACM Computing Surveys* 53, n° 4 (31 juillet 2021): 1-35. <https://doi.org/10.1145/3397499>.
- Calvo-Zaragoza, Jorge, Alejandro H. Toselli, et Enrique Vidal. « Handwritten Music Recognition for Mensural Notation with Convolutional Recurrent Neural Networks ». *Pattern Recognition Letters* 128 (1 décembre 2019): 115-21. <https://doi.org/10.1016/j.patrec.2019.08.021>.
- De-Re-M.E.I. : A Look at the Music Encoding Initiative Schema*, 2022. <https://youtu.be/GBcMEjogCxl>.
- Dorfer, Matthias, Florian Henkel, et Gerhard Widmer. « Learning to Listen, Read, and Follow: Score Following as a Reinforcement Learning Game ». *arXiv e-prints*, 1 juillet 2018. <https://ui.adsabs.harvard.edu/abs/2018arXiv180706391D>.
- Fujinaga, Ichiro, Andrew Hankinson, et Laurent Pugin. « Automatic Score Extraction with Optical Music Recognition (OMR) ». In *Springer Handbook of Systematic Musicology*, édité par Rolf Bader, 299-311. Springer Handbooks. Berlin, Heidelberg: Springer, 2018. [https://doi.org/10.1007/978-3-662-55004-5\\_16](https://doi.org/10.1007/978-3-662-55004-5_16).
- Hajič Jr., Jan. « Optical Music Recognition Using Deep Learning ». Prague: Charles University, 2020. [https://ufal.mff.cuni.cz/~hajicj/2020/docs/GAUK1444217\\_techreport.pdf](https://ufal.mff.cuni.cz/~hajicj/2020/docs/GAUK1444217_techreport.pdf).
- Hankinson, Andrew. « Optical Music Recognition ». Bodleian Libraries, University of Oxford, 2017. [https://tm.web.ox.ac.uk/sites/default/files/tm/documents/media/hankinson\\_omr.pdf](https://tm.web.ox.ac.uk/sites/default/files/tm/documents/media/hankinson_omr.pdf).
- Henkel, Florian, Rainer Kelz, et Gerhard Widmer. « Learning to Read and Follow Music in Complete Score Sheet Images ». arXiv, 21 juillet 2020. <http://arxiv.org/abs/2007.10736>.
- Irving, Andy. IIF AV, 3 juillet 2020. <https://youtu.be/1noRWCvyUrg?t=987>.
- Ju, Yaolong, Cory McKay, Emily Hopkins, Julie Cumming, Gustavo Polins Pedro, et Ichiro Fujinaga. « SIMSSA DB », 2019. <https://www.semanticscholar.org/paper/SIMSSA-DB-Ju-McKay/b1cc3cadabe60c622bf3afe008ccfd505c74ce4d>.

- Keil, Klaus, et Laurent Pugin. « Das Internationale Quellenlexikon der Musik, RISM: Ein Gemeinschaftsprojekt zum Nutzen und als Aufgabe für Forschung und Bibliotheken ». *Bibliothek Forschung und Praxis* 42, n° 2 (26 juillet 2018): 309-18. <https://doi.org/10.1515/bfp-2018-0042>.
- Keil, Klaus, et Jennifer Ward. « Applications of RISM data in digital libraries and digital musicology ». *International Journal on Digital Libraries* 20 (1 mars 2019). <https://doi.org/10.1007/s00799-016-0205-3>.
- Kwon, Taegyun, Dasaem Jeong, et Juhan Nam. « Audio-to-Score Alignment Of Piano Music Using RNN-Based Automatic Music Transcription », 6. Finlande, 2017. <https://mac.kaist.ac.kr/pubs/KwonJeongNam-smc2017.pdf>.
- Languin, Laurence. « Panorama des ressources documentaires en ligne ». In *Musique en bibliothèque*, 323-42. Bibliothèques. Paris: Éditions du Cercle de la Librairie, 2012. <https://doi.org/10.3917/elec.alix.2012.02.0323>.
- Lee, Lin Hao. « Musical Score Following and Audio Alignment ». arXiv, 6 mai 2022. <http://arxiv.org/abs/2205.03247>.
- Nakamura, Tomohiko, Eita Nakamura, et Shigeki Sagayama. « Real-Time Audio-to-Score Alignment of Music Performances Containing Errors and Arbitrary Repeats and Skips ». *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24, n° 2 (février 2016): 329-39. <https://doi.org/10.1109/TASLP.2015.2507862>.
- Pugin, Laurent. « Interaction Perspectives for Music Notation Applications ». In *Proceedings of the 1st International Workshop on Semantic Applications for Audio and Music*, 54-58. Monterey CA USA: ACM, 2018. <https://doi.org/10.1145/3243907.3243911>.
- Pugin, Laurent, et Claudio Bacciagaluppi. « An Analysis of Musical Work Datasets and their Current Level of Linkage ». In *7th International Conference on Digital Libraries for Musicology*, 32-39. DLfM 2020. New York, NY, USA: Association for Computing Machinery, 2020. <https://doi.org/10.1145/3424911.3425518>.
- Rigaux, Philippe. « Projet GioQoso - Gestion de la qualité des partitions musicales ouvertes », 2019. [https://www.cnrs.fr/mi/IMG/pdf/philippe\\_rigaux\\_-\\_qioqoso\\_-\\_mastodons2019.pdf](https://www.cnrs.fr/mi/IMG/pdf/philippe_rigaux_-_qioqoso_-_mastodons2019.pdf).
- Robineau, Régis. « Actualités et perspectives de IIF ». 24 mars 2021. <https://projet.biblissima.fr/fr/actualites-perspectives-iif>.
- Tuggener, Lukas, Ismail Elezi, Jurgen Schmidhuber, et Thilo Stadelmann. « Deep Watershed Detector for Music Object Recognition ». arXiv, 26 mai 2018. <https://doi.org/10.48550/arXiv.1805.10548>.

Ward, Jennifer A. « The use and reuse of RISM data in libraries ». *Journal of Electronic Resources Librarianship* 28, n° 2 (2 avril 2016): 129-33.  
<https://doi.org/10.1080/1941126X.2016.1167551>.

Wu, Fu-Hai Frank, et Jyh-Shing Roger Jang. « An Architecture for Optical Music Recognition of Numbered Music Notation ». In *Proceedings of International Conference on Internet Multimedia Computing and Service*, 241-45. ICIMCS '14. New York, NY, USA: Association for Computing Machinery, 2014.  
<https://doi.org/10.1145/2632856.2632930>.

### 6.3 Intelligence Artificielle

Benhamou, Yaniv, et Sarah Kenderdine. « Enjeux juridiques et éthiques / AI & GLAM collections: legal & ethical challenges to sharing GLAMs collections ». Université Paris-Saclay et la Bibliothèque nationale de France, 10 décembre 2021. <https://www.bnf.fr/fr/mediatheque/enjeux-juridiques-et-ethiques-ai-glam-collections-legal-ethical-challenges-sharing>.

Engel, Laurence, Mike Keller, et Aslak Sira Myhre. « Discussion conclusive autour des enjeux de l'IA en bibliothèque - Conclusion panel: libraries facing the AI challenge ». Université Paris-Saclay et la Bibliothèque nationale de France, 10 décembre 2021. <https://www.bnf.fr/fr/mediatheque/discussion-conclusive-autour-des-enjeux-de-lia-en-bibliotheque-conclusion-panel>.

Franke, Fabian, Konstanze Söllner, et Ludger Syré. « Künstliche Intelligenz Und Bibliotheken Bericht Vom Österreichischen Bibliothekartag 10.–13.9.2019 in Graz ». *ABI-Technik* 40, n° 1 (2020): 79-92.  
<https://doi.org/10.1515/abitech-2020-1009>.

Hutchinson, Tim. « Natural Language Processing and Machine Learning as Practical Toolsets for Archival Processing ». *Records Management Journal* 30, n° 2 (2020): 155-74. <http://dx.doi.org/10.1108/RMJ-09-2019-0055>.

Jaillant, Lise. *Archives, Access and Artificial Intelligence*. Digital Humanities Research. Bielefeld: University Press, 2022.  
<https://doi.org/10.1515/9783839455845>.

Verwayen, Harry. « Could AI and Data Mining Technologies Overcome Issues in Cultural Heritage? » *Europeana Pro*, 25 septembre 2018.  
<https://pro.europeana.eu/post/could-ai-and-data-mining-technologies-overcome-copyright-issues-in-cultural-heritage>.

### 6.4 Humanités numériques

Bahde, Anne. « Conceptual Data Visualization in Archival Finding Aids: Preliminary User Responses ». *portal: Libraries and the Academy* 17, n° 3 (2017): 485-506. <https://doi.org/10.1353/pla.2017.0031>.

- Balkun, Mary McAleer, et Marta Mestrovic Deyrup. *Transformative Digital Humanities: Challenges and Opportunities*. Digital Research in the Arts and Humanities. London: Routledge Taylor & Francis Group, 2020.
- Berry, David M. *Understanding Digital Humanities*. Houndmills, Basingstoke, Hampshire ; Palgrave Macmillan, 2012.
- Cemens Manz, Marian, Julien A. Raemy, Béatrice Gauvain, et Vera Chiquet. « Let's talk about standards – A write-up of a discussion on metadata standardization in the Digital Humanities ». Blog/ DH-CH, 18 janvier 2022. <https://dh-ch.ch/blog/posts/let's-talk-about-standards-a-write-up-of-a-discussion-on-metadata-standardization-in-the-digital-humanities.html>.
- Cheng, Yu-Jung, et Shu-Lai Chou. « Using Digital Humanity Approaches to Visualize and Evaluate the Cultural Heritage Ontology ». *Electronic Library* 40, n° 1/2 (2022): 83-98. <https://doi.org/10.1108/EL-09-2021-0171>.
- Eve, Martin Paul. *The Digital Humanities and Literary Studies*. First edition. The Literary Agenda. Oxford: University Press, 2022.
- Giuliano, Frédéric. « Humanités numériques et archives : la longue émergence d'un nouveau paradigme ». *Documentation et bibliothèques* 65, n° 2 (2019): 37-46. <https://doi.org/10.7202/1063788ar>.
- Hassani, Hossein, Christina Beneki, Stephan Unger, Maedeh Taj Mazinani, et Mohammad Reza Yeganegi. « Text Mining in Big Data Analytics ». *Big Data and Cognitive Computing* 4, n° 1 (mars 2020): 1. <https://doi.org/10.3390/bdcc4010001>.
- Ma, Rongqian, et Kai Li. « Digital Humanities as a Cross-disciplinary Battleground: An Examination of Inscriptions in Journal Publications ». *Journal of the American Society for Information Science and Technology* 73, n° 2 (2022): 172-87. <https://doi.org/10.1002/asi.24534>.
- Mounier, Pierre. « Manifeste des Digital Humanities ». *Journal des anthropologues. Association française des anthropologues*, n° 122-123 (1 décembre 2010): 447-52. <https://doi.org/10.4000/jda.3652>.
- Poole, Alex H., et Deborah A. Garwood. « “Natural allies”: Librarians, archivists, and big data in international digital humanities project work ». *Journal of Documentation* 74, n° 4 (1 janvier 2018): 804-26. <https://doi.org/10.1108/JD-10-2017-0137>.
- Schneider, Ulrich Johannes. « Deutsche Nationalkataloge – Herausforderungen an Das Deutsche Bibliothekssystem: Was Aus Der Perspektive Der Digital Humanities Zu Tun Wäre ». *ABI Technik* 40, n° 1 (1 février 2020): 40-51. <https://doi.org/10.1515/abitech-2020-1005>.
- Wilson, Emma Annette. *Digital Humanities for Librarians*. Lanham, Maryland ; Rowman & Littlefield Publishing Group, 2020.

Wymer, Kathryn C. *Introduction to Digital Humanities: Enhancing Scholarship with the Use of Technology*. Routledge Focus on Literature Series. New York, New York ; Routledge, 2021.

Ying-Hsang Liu et Koraljka Golub. *Information and Knowledge Organisation in Digital Humanities: Global Perspectives*. Digital Research in the Arts and Humanities. Taylor and Francis, 2021.  
<https://doi.org/10.4324/9781003131816>.

## 6.5 Réutilisation des données

Avanço, Karla. « FAIR Principles and Linked Open Data ». Billet. *The Road to FAIR* (blog). Consulté le 23 juillet 2022. <https://roadtofair.hypotheses.org/288>.

Berners-Lee, Tim. « Linked Data ». W3C, 27 juillet 2006.  
<https://www.w3.org/DesignIssues/LinkedData.html>.

Candela, Gustavo, María Dolores Sáez, MPilar Escobar Esteban, et Manuel Marco-Such. « Reusing Digital Collections from GLAM Institutions ». *Journal of Information Science* 48, n° 2 (1 avril 2022): 251-67.  
<https://doi.org/10.1177/0165551520950246>.

Collins, Sandra, Jennifer McCarthy, Garfield Benjamin, et Lorna Hughes. « Using FAIR Data from Galleries, Libraries, Archives and Museums (GLAM) Sector ». Webinaire en ligne, 29 septembre 2020. <https://www.rd-alliance.org/using-fair-data-galleries-libraries-archives-and-museums-glam-sector>.

Cudré-Mauroux, Philippe, et Sebastian Sigloch. « Design Considerations on SWITCH's Connectome Vision », 17 février 2020.  
[https://www.switch.ch/export/sites/default/about/innovation/.galleries/files/SWITCHInnovationLab\\_eXascale-InfoLab\\_Results.pdf](https://www.switch.ch/export/sites/default/about/innovation/.galleries/files/SWITCHInnovationLab_eXascale-InfoLab_Results.pdf).

Go, Sugimoto. « Open Data Empowerment of Digital Humanities by Wikipedia/DBpedia Gamification and Crowd Curation –WiQiZi's Challenges with APIs and SPARQL ». Mexico City, 30 septembre 2019.  
<https://doi.org/10.5281/zenodo.3465654>.

Harrower, Natalie, Roy Stanley, Saskia Scheltjens, et Kathryn Cassidy. « Publishing GLAM Data as FAIR Data ». Webinaire en ligne, 29 octobre 2020. <https://www.rd-alliance.org/publishing-glam-data-fair-data>.

Hasnain, Ali, et Dietrich Rebholz-Schuhmann. « Assessing FAIR Data Principles against the 5-Star Open Data Principles ». Springer Verlag, 2018.  
[https://doi.org/10.1007/978-3-319-98192-5\\_60](https://doi.org/10.1007/978-3-319-98192-5_60).

Hauf, Nicolai, et Martin Jaekel. Data reuse in Social Science and Humanities. Entretien réalisé par SWITCH, 30 avril 2021.

<https://www.switch.ch/stories/Data-reuse-in-Social-Science-and-Humanities/>.

He, Lin, et Vinita Nahar. « Reuse of Scientific Data in Academic Publications: An Investigation of Dryad Digital Repository ». *Aslib Journal of Information Management* 68, n° 4 (18 juillet 2016): 478-94.

<https://doi.org/10.1108/AJIM-01-2016-0008>.

Hügi, Jasmin, Nicolas Prongué, et René (dir ) Schneider. « Les bibliothèques face aux Linked Open Data », 2014.

<https://sonar.ch/global/documents/314396>.

Kennisland, Maarten Zeinstra. « Report on Requirements for Usage and Reuse Statistics for GLAM Content », 2014. <https://www.kl.nl/wp-content/uploads/2014/04/report-on-requirements-for-usage-and-reuse-statistics-for-glam-content.pdf>.

Koster, Lukas, et Saskia Woutersen-Windhouver. « FAIR Principles for Library, Archive and Museum Collections: A proposal for standards for reusable collections ». *The Code4Lib Journal*, n° 40 (4 mai 2018).

<https://journal.code4lib.org/articles/13427>.

Montana State University-Bozeman, Sara Mannheimer, Leila Sterman, Montana State University-Bozeman, Susan Borda, et Montana State University-Bozeman. « Discovery and Reuse of Open Datasets: An Exploratory Study ». *Journal of eScience Librarianship* 5, n° 1 (19 juillet 2016): e1091.

<https://doi.org/10.7191/jeslib.2016.1091>.

Nicolai, Hauf, Fürholz Andreas, Klaas Christina Vanessa, Morger Jennifer, Šimukovič Elena, et Jaekel Martin. « Data Reuse in the Social Sciences and Humanities : Project Report of the SWITCH Innovation Lab “Repositories & Data Quality” ». 32,application/pdf, mars 2021.

<https://doi.org/10.21256/ZHAW-2404>.

Pinède, Nathalie, Matthieu Noucher, Françoise Gourmelon, et Karel Soumagnac-Colin. « De l’open data à l’open science : retour réflexif sur les méthodes et pratiques d’une recherche sur les données géographiques ». *Revue française des sciences de l’information et de la communication*, n° 11 (1 août 2017). <https://doi.org/10.4000/rfsic.3200>.

Rettig, Laura. De la source au connectome. Entretien réalisé par Cornelia Puhze, 2 juillet 2020. <https://www.switch.ch/fr/stories/source-to-connectome/>.

Robineau, Régis. « Adopter et utiliser les standards IIIF pour vos corpus d’images numériques ». *Biblissima - Campus Condorcet / EPHE-PSL*, 18 novembre 2020.

[https://projet.biblissima.fr/sites/default/files/atelier\\_iiif\\_dhnord\\_robineau\\_20201118.pdf](https://projet.biblissima.fr/sites/default/files/atelier_iiif_dhnord_robineau_20201118.pdf).

Shen, Qi, Shijun Wu, Yanzhen Zou, Zixiao Zhu, et Bing Xie. « From API to NLI: A New Interface for Library Reuse ». arXiv, 7 juillet 2020.

<https://doi.org/10.48550/arXiv.2007.03305>.

Tallerås, Kim. « Quality of Linked Bibliographic Data: The Models, Vocabularies, and Links of Data Sets Published by Four National Libraries ». *Journal of Library Metadata* 17, no 2 (3 avril 2017): 126-55.

<https://doi.org/10.1080/19386389.2017.1355166>.

Wallis, Jillian. « Data Producers Courting Data Reusers: Two Cases from Modeling Communities ». *International Journal of Digital Curation* 9, n° 1 (12 juin 2014): 98-109. <https://doi.org/10.2218/ijdc.v9i1.304>.

Wallis, Jillian C., Elizabeth Rolando, et Christine L. Borgman. « If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology ». Édité par Luís A. Nunes Amaral. *PLoS ONE* 8, n° 7 (23 juillet 2013): e67332. <https://doi.org/10.1371/journal.pone.0067332>.

Vancauwenbergh, Sadia. FAIR and Open Research Metadata as Leverage for Digital Libraries: The Flemish Case. *Digital Libraries - Advancing Open Science*. IntechOpen, 2021. <https://doi.org/10.5772/intechopen.97192>.

Wittmann, Rachel, Anna Neatrou, Rebekah Cummings, et Jeremy Myntti. « From Digital Library to Open Datasets »: *Information Technology and Libraries* 38, n° 4 (16 décembre 2019): 49-61.

<https://doi.org/10.6017/ital.v38i4.11101>.

## 7 Table des abréviations

AMT	Automatic Music Transcription
APIs	Application Programming Interface
ARK	Archival Resource Key
BCUL	Bibliothèque Cantonale et Universitaire – Lausanne
BGE	Bibliothèque de GENève
BHL	Biodiversity Library Heritage
BN	Bibliothèque Nationale suisse
BnF	Bibliothèque Nationale de France
CC	Creative Commons
CCSA	Catalogue Collectif Suisse des Affiches

CQT	Constant Q-Transform
CWMN	Common Western Music Notation
DCML	Digital and Cognitive Musicology Lab
DDMAL	Distributed Digital Music Archives & Libraries Lab
DTW	Dynamic Time Warping
ELVIS	Electronic Locator of Vertical Interval Successions
DOI	Digital Object Identifier
EPFL	École Polytechnique Fédérale de Lausanne
FAIR	Findable Accessible Interoperable Re-usable
GLAM	Galleries, Libraries, Archives, Museums
HTR	Handwritten Text Recognition
IIIF	International Image Interoperability Framework
IMSLP	International Music Score Library Project
IRCAM	Institut de Recherche et Coordination Acoustique/Musique
ISAD(G)	International Standard Archival Description - General
MACLab	Music and Audio Computing Lab
MARC	MAchine-Readable Cataloging
MEI	Music Encoding Initiative
MIDI	Musical Instrument Digital Interface
MMC	Modèle de Markov Caché
MODS	Metadata Object Description Schema
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OCR	Optical Character Recognition
OMR	Optical Music Recognition
OPAC	Online Public Access Catalog
RiC-O	Records in Contexts-Ontology
RISM	Répertoire International des Sources Musicales
RDF	Resource Description Framework

SPARQL	SPARQL Protocol and RDF Query Language
SIGB	Système Intégré de Gestion de Bibliothèques
SIMSSA	Single Interface for Music Score Searching and Analysis
SONAR	Swiss Open Access Repository
SRU	Search/Retrieve via URL
URI	Uniform Resource Identifier
TEI	Text Encoding Initiative
UNIL	Université de Lausanne

## 8 Table des figures

Figure 1: Illustration des méthodes de conversion courantes entre audio, partitions, MIDI et MusicXML. Les conversions rouges (marquées d'un *) indiquent des conversions avec perte et difficiles ; les conversions vertes (†-marquées) indiquent des conversions sans perte et faciles ; les conversions bleues (non marquées) indiquent des conversions avec perte et faciles. Source : Lin Hao Lee, « Musical Score Following and Audio Alignment » (arXiv, 6 mai 2022), 27. ....	41
Figure 2: Diagramme du fonctionnement des APIs principales de IIF, Source : « How It Works - Introduction », IIF, <a href="https://iif.io/get-started/how-iif-works/">https://iif.io/get-started/how-iif-works/</a> . ....	42
Figure 3: MACLab de Kaist, démonstration web du projet "Score and Performance Alignment". Partition : Chopin, Etude op. 10 5 ; interprètes : Cortot, Horowitz, Paderewski et Vasary. La transparence des notes correspond à leur volume, il est possible de passer dynamiquement d'un interprète à l'autre. ....	48
Figure 4 : Etude en F mineur, Op. 10 n° 9, Chopin Frédéric. Démonstration des APIs IIF pour la mise en relation de la partition et d'une représentation musicale par Valentina Lisitsa. Source : <a href="https://ddmal.music.mcgill.ca/IIF-AV-player">https://ddmal.music.mcgill.ca/IIF-AV-player</a> . ....	49
Figure 5: Musiclibs Recognition & Indexing Process, Sources : Andrew Hankinson, « Optical Music Recognition » (Bodleian Libraries, University of Oxford, 2017), <a href="https://tm.web.ox.ac.uk/sites/default/files/tm/documents/media/hankinson_omr.pdf">https://tm.web.ox.ac.uk/sites/default/files/tm/documents/media/hankinson_omr.pdf</a> . ....	50

## 9 Annexes

### 9.1 Tableaux récapitulatifs

#### 9.1.1 Scriptorium, Patrinum et Alma digital

Nom de la Bibliothèque numérique	Patrinum	Scriptorium	Alma Digital
Institution responsable de la Bibliothèque numérique	BCU Lausanne (Renouvaud)	BCU Lausanne	Réseau Renouvaud
Année de mise en ligne	2018	2012	2019
Produit utilisé	Invenio	MediaINFO	Alma (Module du SIGB)
Développement et maintenance	Société externe : TIND	Société externe : Geneza	ExLibris
<b>Type de documents et format de fichier</b>			
<b>Type de documents acceptés</b>	<input checked="" type="checkbox"/> Texte	<input checked="" type="checkbox"/> Texte	<input checked="" type="checkbox"/> Texte
	<input checked="" type="checkbox"/> Image	<input checked="" type="checkbox"/> Image	<input checked="" type="checkbox"/> Image
	<input checked="" type="checkbox"/> Texte et Image	<input checked="" type="checkbox"/> Texte et Image	<input checked="" type="checkbox"/> Texte et Image
	<input checked="" type="checkbox"/> Son	<input checked="" type="checkbox"/> Son	<input checked="" type="checkbox"/> Son
	<input checked="" type="checkbox"/> Vidéo	<input checked="" type="checkbox"/> Vidéo	<input checked="" type="checkbox"/> Vidéo
	<input checked="" type="checkbox"/> Partition	<input type="checkbox"/> Partition	<input type="checkbox"/> Partition
	<input checked="" type="checkbox"/> Autre(s) : Tout fichier peut être accepté.	<input type="checkbox"/> Autre(s) :	<input type="checkbox"/> Autre(s) :
<b>Format de fichiers acceptés</b>	Tous les formats de fichiers sont acceptés par le système, mais celui-ci n'offre pas d'interface de consultation pour tous les fichiers. Il est par contre toujours possible d'offrir le téléchargement comme moyen de consultation.	PDF (documents texte/image), set JPEG2000/PDF/XML (documents texte/image), MP4 (documents vidéo), AVI (vidéo), MOV (vidéo), MTS (vidéo).	Tous, mais consultation possible sur l'une des visionneuses seulement dans certains format (téléchargement possible)
<b>Modalités d'import des documents et des métadonnées</b>			
<b>Métadonnées minimales exigées</b>	Des métadonnées minimales sont exigées pour la création de DOI : Identifier, Creator, Title, Publisher, PublicationYear, ResourceType.	Pas de métadonnées minimales exigées par le système.	Pas de métadonnées minimales exigées par le système.
	<input checked="" type="checkbox"/> marcxml	<input type="checkbox"/> marcxml	<input checked="" type="checkbox"/> marcxml

Nom de la Bibliothèque numérique	Patrinum	Scriptorium	Alma Digital
<b>Format de métadonnées accepté</b>	<input type="checkbox"/> dublin core	<input checked="" type="checkbox"/> dublin core	<input checked="" type="checkbox"/> dublin core
	<input checked="" type="checkbox"/> isad(g)/ead (les données sont stockées au format MARC mais affichées selon ISAD(G))	<input type="checkbox"/> isad(g)/ead	<input type="checkbox"/> isad(g)/ead
	<input checked="" type="checkbox"/> autre(s) : mrc et csv	<input type="checkbox"/> autre(s) :	<input checked="" type="checkbox"/> autre(s) : Excel, binaire, csv
<b>Remarques</b>	A côté d'une grille MARC classique, le système peut aussi proposer des formulaires de soumission, mais un développement est nécessaire pour gérer l'import au format MARC/ISAD(G)	Les champs de métadonnées dans Scriptorium sont entièrement personnalisables. En tant que tel, Scriptorium ne respecte aucun format standard. Scriptorium peut importer des métadonnées via OAI-PMH ou Z39.50. Pour le moment, seul Dublin Core est accepté (avec un mapping qui met en relation DC avec les champs de Scriptorium). Accepter d'autres formats XML ne devrait pas être un problème, mais la BCUL devrait probablement créer le mapping correspondant.	Alma permet de créer des profils d'import acceptant différents types de fichiers (XML, binaire, Excel et fichiers délimités) respectant différents formats MARC : MARC21, UNIMARC, etc.)
<b>Niveaux de préservation</b>			
<b>Niveau de préservation proposé</b>	Patrinum propose un module de conservation du numérique à long terme.	Scriptorium est un système de mise en ligne et ne dispose pas (de la totalité) des fonctionnalités standards que l'on exige d'un système de préservation du numérique à long terme.	Alma Digital n'est pas un système de conservation du numérique à long terme.
<b>Modalités de recherche</b>			
<b>Recherche dans les métadonnées</b>	Standard	Standard	Standard
	Fonctionnalité spécifique : Nombreux opérateurs disponibles, recherche par expressions régulières	Fonctionnalité spécifique : recherches complexes possibles, expressions régulières	Fonctionnalité spécifique : recherche avancée avec fonctionnalités habituelles

Nom de la Bibliothèque numérique	Patrinum	Scriptorium	Alma Digital
<b>Recherche dans le texte</b>	<input checked="" type="checkbox"/> Oui (dans totalités du corpus)	<input checked="" type="checkbox"/> Oui (dans totalités du corpus)	<input type="checkbox"/> Oui (dans totalités du corpus)
	<input type="checkbox"/> Oui (dans un document à la fois)	<input checked="" type="checkbox"/> Oui (dans un document à la fois)	<input checked="" type="checkbox"/> Oui (dans un document à la fois)
<b>Remarque</b>	La recherche plein texte n'est pas activée par défaut. Le texte recherché n'est pas mis en évidence dans l'affichage des résultats. Nombreuses facettes disponibles pour raffiner la recherche.	Le texte recherché est mis en évidence dans l'affichage des résultats. Fonctionnalité de Timeline disponible (représentation chronologique du nombre de documents en fonction de la date de publication). Possibilité de filtrer les résultats par revue.	Les usagers font leurs recherches dans Renouvaud/Primo qui est un outil de découverte.
<b>Modalités de consultation</b>			
<b>Modalités de consultation disponibles</b>	<input checked="" type="checkbox"/> En ligne (pdf, jpg)	<input checked="" type="checkbox"/> En ligne	<input checked="" type="checkbox"/> En ligne
	<input checked="" type="checkbox"/> Par téléchargement	<input checked="" type="checkbox"/> Par téléchargement	<input checked="" type="checkbox"/> Par téléchargement
<b>Fonctionnalités de consultation en ligne disponibles</b>	Aucune avec le paramétrage actuel pour les fichiers PDF. Zoom et rotation pour fichiers JPG (usage d'Universal Viewer).	Zoom et affichage des pages selon différents modes. Navigation aisée entre les pages d'un document. Fonctionnalité de sélection, d'export texte et de traduction automatique pour les personnes connectées.	Le visionneur d'Alma peut afficher des images au format jpg, gif et png, des documents PDF, des documents Microsoft Word, et des formats audio/vidéo pris en charge par HTML5 (la liste peut varier en fonction des navigateur).

Nom de la Bibliothèque numérique	Patrinum	Scriptorium	Alma Digital
<b>Modalités de téléchargement disponibles</b>	<p>Document complet. Pour les documents pdf dont le texte est reconnu, possibilité de téléchargement au format txt. Pour les images possibilités de téléchargement en pdf. Possibilités de téléchargement par lot si plusieurs documents sont rattachés à une même notice.</p> <p>Les notices peuvent être téléchargées dans les formats suivants : BibTeX, MARC, MARCXML, DublinCore, EndNote, NLM, RefWorks, RIS</p>	<p>Document complet au format pdf. Plus d'options disponibles pour les personnes connectées : export possible d'une sélection de pages ou d'une portion de texte au format texte ou image.</p>	<p>Document complet</p>
<b>Comptes utilisateurs et professionnels</b>			
<b>Création de compte possible</b>	<p>Oui, pour l'instant uniquement les professionnels disposent d'un compte, mais un développement est en cours pour permettre aux usagers d'utiliser leurs identifiants Renouvaud.</p>	<p>Oui</p>	<p>Oui</p>
<b>Remarques</b>	<p>Les comptes permettent de télécharger les métadonnées à partir du résultat d'une recherche ou d'une sélection (fonction panier) dans différents formats dont MARC, MARCXML, MODS (Metadata Object Description Schema), Dublin Core, HTML brief, BibTeX, EndNote, NLM, RefWorks.</p>	<p>Comptes utilisateurs disponibles et indispensables pour accéder aux outils de sélection de texte et de traduction.</p>	<p>Possibilité de donner un accès à certaines collections en fonction des permissions accordées aux usagers.</p>

Nom de la Bibliothèque numérique	Patrinum	Scriptorium	Alma Digital
<b>Modalités de création/gestion des comptes usagers</b>	L'interface administrateur permet de gérer les comptes.	L'interface administrateur permet de gérer les comptes. La création des comptes peut aussi se faire à travers un formulaire à remplir par les usagers intéressés.	L'interface administrateur permet de gérer les comptes. La création des comptes peut aussi se faire à travers un formulaire à remplir par les usagers intéressés.
<b>Gestion des droits sur les documents</b>			
<b>Possibilité d'indiquer les conditions d'utilisation pour chaque collection</b>	Oui	Oui	Oui, mais pas lors de la saisie manuelle : usage d'un job Alma ou des APIs requis
<b>Possibilité d'indiquer les conditions d'utilisation pour chaque document</b>	Oui	Oui	Oui
<b>Possibilité de limiter l'accès en fonction du document et du type de compte</b>	Oui	Oui	Oui
<b>Possibilité de limiter l'accès en fonction du document et de l'adresse IP</b>	Oui	Oui	Oui
<b>Remarques</b>			Possibilité d'indiquer les conditions d'utilisation au niveau de la notice bibliographique ou au niveau de la représentation.
<b>Crowdsourcing</b>			
<b>Description des possibilités de crowdsourcing et des conditions associées</b>	Possibilité de permettre aux usagers de rédiger des commentaires, mais la fonction est désactivée.	Des options existent pour permettre aux usagers de corriger l'OCR, de transcrire les textes non reconnus ou de rendre des notes publiques, mais ces fonctionnalités ne sont pas encore activées.	Inexistantes
<b>Modalités de partage des métadonnées et documents</b>			

<b>Nom de la Bibliothèque numérique</b>	<b>Patrinum</b>	<b>Scriptorium</b>	<b>Alma Digital</b>
<b>Possibilité de mise à disposition des métadonnées</b>	OAI-PMH (publication active des données bibliothèques uniquement, mais non documentée publiquement), APIs (disponibles mais non documentées)	OAI-PMH (pas de publication active), APIs (possible, mais autorisations à demander à Geneza)	OAI-PMH, APIs disponibles mais pas de volonté de promouvoir cette option.
<b>Possibilités de mise à disposition des documents</b>	Développement en cours pour inclure le standard IIIF	Mise en conformité requise pour utiliser le standard IIIF	APIs IIIF pour le contenu hébergé sur Alma Digital.
<b>Signalement dans Renouvaud</b>	Oui, en partie (notices bibliographiques du côté bibliothèque uniquement) via OAI-PMH	N'est pas envisageable pour le moment. Il est possible de moissonner Scriptorium via OAI-PMH, mais cette solution ne serait pas optimale puisque cela créerait une notice pour chaque parution de chaque journal.	Oui, obligatoire : Alma digital fait partie du catalogue Renouvaud.
<b>Identifiants pérennes disponibles</b>	Un identifiant DOI est créé pour chaque notice. L'Universal Viewer permet aussi de créer des urls personnalisés, de même que des « YouTube-style embedding with deep links to specific pages/zoom regions ».	Scriptorium permet d'obtenir une url pour chaque liste de résultats et chaque affichage particulier d'une page. Cette fonctionnalité peut en particulier être utilisée pour faire des liens vers une collection de Scriptorium.	Non. Mais possibilité d'avoir des liens pérennes vers les notices bibliographiques
<b>Modalités d'export des données (si migration vers un autre système)</b>	Les données se trouvent sur des serveurs de Switch. Les métadonnées peuvent être exportées via un fichier csv, de même qu'en MARCXML, marc, mods ou Dublin Core.	Les données sont sur des serveurs du centre informatique de l'UNIL. Il est aussi possible de récupérer les métadonnées (Dublin Core) via OAI-PMH.	Les données se trouvent sur des serveurs Amazon. Les métadonnées peuvent être publiées via OAI-PMH ou MARCXML, Dublin Core, ETD-MS, Bibframe et RDA/RDF

### 9.1.2 Tableau récapitulatif des collections patrimoniales numériques BCUL

Nom des collections	Archives musicales	Iconopôle	Manuscrits et archives	Documentation vaudoise	Dépôt légal numérique	Réserve précieuse
<b>Responsable(s) de la collection</b>	Verena Monnier	Sophie Donche Gay	Ramona Fritschi	Myriam Jouhar	Héloïse Hahling	Marc Tiefenauer
<b>Thématique</b>	Création musicale et activité musicographique	Documents iconographiques vaudois	Fonds d'archives, pièces isolées, manuscrits médiévaux et cartes géographiques anciennes	Documents relatifs au canton de Vaud sur tous les supports	Publications imprimées ou numériques éditées dans le canton de Vaud depuis 1938	Documents imprimés au 15 <sup>e</sup> -16 <sup>e</sup> siècles, publications en lien avec le canton d'avant 1800, livres rares et livres modernes à tirage limité
<b>Plateforme BCUL</b>	<input checked="" type="checkbox"/> Patrinum	<input checked="" type="checkbox"/> Patrinum	<input checked="" type="checkbox"/> Patrinum	<input checked="" type="checkbox"/> Patrinum	<input checked="" type="checkbox"/> Patrinum	<input type="checkbox"/> Patrinum
	<input type="checkbox"/> Scriptorium	<input type="checkbox"/> Scriptorium	<input type="checkbox"/> Scriptorium	<input type="checkbox"/> Scriptorium	<input type="checkbox"/> Scriptorium	<input checked="" type="checkbox"/> Scriptorium
	<input type="checkbox"/> Renouvaud	<input type="checkbox"/> Renouvaud	<input type="checkbox"/> Renouvaud	<input checked="" type="checkbox"/> Renouvaud	<input checked="" type="checkbox"/> Renouvaud	<input checked="" type="checkbox"/> Renouvaud
<b>Plateformes externes</b>	RISM	*CCSA *Bientôt Memobase pour les photos et films fixes *Peut-être notreHistoire.ch	*e-codices *ego-documents	Wikipédia (800 articles à partir de notices d'autorités).	-	*e-rara *Google books
<b>Type de document et format de fichier</b>						
<b>Type de documents</b>	<input checked="" type="checkbox"/> Texte	<input type="checkbox"/> Texte	<input checked="" type="checkbox"/> Texte	<input checked="" type="checkbox"/> Texte	<input checked="" type="checkbox"/> Texte	<input checked="" type="checkbox"/> Texte
	<input checked="" type="checkbox"/> Image	<input checked="" type="checkbox"/> Image	<input checked="" type="checkbox"/> Image	<input checked="" type="checkbox"/> Image	<input type="checkbox"/> Image	<input checked="" type="checkbox"/> Image
	<input checked="" type="checkbox"/> Texte et Image	<input type="checkbox"/> Texte et Image	<input checked="" type="checkbox"/> Texte et Image	<input checked="" type="checkbox"/> Texte et Image	<input type="checkbox"/> Texte et Image	<input checked="" type="checkbox"/> Texte et Image

Nom des collections	Archives musicales	Iconopôle	Manuscrits et archives	Documentation vaudoise	Dépôt légal numérique	Réserve précieuse
	<input type="checkbox"/> Son	<input type="checkbox"/> Son	<input checked="" type="checkbox"/> Son	<input checked="" type="checkbox"/> Son	<input type="checkbox"/> Son	<input type="checkbox"/> Son
	<input type="checkbox"/> Vidéo	<input type="checkbox"/> Vidéo	<input type="checkbox"/> Vidéo	<input checked="" type="checkbox"/> Vidéo	<input type="checkbox"/> Vidéo	<input type="checkbox"/> Vidéo
	<input checked="" type="checkbox"/> Partition	<input type="checkbox"/> Partition	<input type="checkbox"/> Partition	<input type="checkbox"/> Partition	<input checked="" type="checkbox"/> Partition	<input type="checkbox"/> Partition
	<input type="checkbox"/> Autre(s) :	<input type="checkbox"/> Autre(s) :	<input checked="" type="checkbox"/> Autre(s) : Objet	<input type="checkbox"/> Autre(s) :	<input type="checkbox"/> Autre(s) :	<input type="checkbox"/> Autre(s) :
<b>Format des fichiers</b>	Pdf, tiff, musicxml, divers formats propriétaires (partitions)	Jpg, tiff, pdf	Jpg, tiff	Jpg, mp3, mpg	Pdf ; epub, peut-être format warc dans le futur (publications web)	Pdf
<b>Format des métadonnées et lien avec les documents</b>						
	<input type="checkbox"/> Oui	<input type="checkbox"/> Oui	<input type="checkbox"/> Oui	<input type="checkbox"/> Oui	<input checked="" type="checkbox"/> Oui	<input type="checkbox"/> Oui
Métadonnées disponibles	<input checked="" type="checkbox"/> Partiellement	<input checked="" type="checkbox"/> Partiellement	<input checked="" type="checkbox"/> Partiellement : depuis 2019 seulement	<input type="checkbox"/> Partiellement	<input type="checkbox"/> Partiellement	<input type="checkbox"/> Partiellement
	<input type="checkbox"/> Non	<input type="checkbox"/> Non	<input type="checkbox"/> Non	<input type="checkbox"/> Non	<input type="checkbox"/> Non	<input type="checkbox"/> Non
Format des métadonnées	<input type="checkbox"/> marcxml	<input type="checkbox"/> marcxml	<input type="checkbox"/> marcxml	<input checked="" type="checkbox"/> marcxml	<input checked="" type="checkbox"/> marcxml	<input checked="" type="checkbox"/> marcxml
	<input checked="" type="checkbox"/> mods	<input type="checkbox"/> mods	<input type="checkbox"/> mods	<input type="checkbox"/> mods	<input type="checkbox"/> mods	<input type="checkbox"/> mods
	<input type="checkbox"/> dublin core	<input type="checkbox"/> dublin core	<input type="checkbox"/> dublin core	<input type="checkbox"/> dublin core	<input type="checkbox"/> dublin core	<input checked="" type="checkbox"/> dublin core
	<input checked="" type="checkbox"/> isad(g)/ead	<input checked="" type="checkbox"/> isad(g)/ead	<input checked="" type="checkbox"/> isad(g)/ead	<input type="checkbox"/> isad(g)/ead	<input type="checkbox"/> isad(g)/ead	<input type="checkbox"/> isad(g)/ead
	<input type="checkbox"/> autre(s) :	<input type="checkbox"/> autre(s) :	<input type="checkbox"/> autre(s) :	<input type="checkbox"/> autre(s) :	<input type="checkbox"/> autre(s) :	<input type="checkbox"/> autre(s) :

Nom des collections	Archives musicales	Iconopôle	Manuscrits et archives	Documentation vaudoise	Dépôt légal numérique	Réserve précieuse	
<b>Remarques</b>	Le format mods est sauvegardé en marcxml, de même que le format isad(g)	Métadonnées embarquées à fabriquer : pour la réutilisation de l'image en sachant d'où elle vient			Les métadonnées sont produites par la BCUL. Dans certains cas, les métadonnées sont introduites par l'éditeur, de manière préalable.		
<b>Droits</b>							
<b>Statut des documents</b>	<input checked="" type="checkbox"/> Domaine public	<input checked="" type="checkbox"/> Domaine public	<input checked="" type="checkbox"/> Domaine public	<input type="checkbox"/> Domaine public	<input type="checkbox"/> Domaine public	<input checked="" type="checkbox"/> Domaine public	
	<input type="checkbox"/> Licence CC (préciser laquelle) :	<input type="checkbox"/> Licence CC (préciser laquelle) :					
	<input checked="" type="checkbox"/> Autre statut : Autorisation de publication réglée dans une convention	<input checked="" type="checkbox"/> Autre statut : Autorisation de publication réglée dans une convention	<input checked="" type="checkbox"/> Autre statut : Autorisation de publication réglée dans une convention	<input checked="" type="checkbox"/> Autre statut : Autorisation de publication réglée dans une convention	<input checked="" type="checkbox"/> Autre statut : Autorisation de publication réglée dans une convention	<input checked="" type="checkbox"/> Autre statut : Sous droit	<input type="checkbox"/> Autre statut
	<input checked="" type="checkbox"/> Pas de diffusion en ligne autorisée	<input type="checkbox"/> Pas de diffusion en ligne autorisée	<input type="checkbox"/> Pas de diffusion en ligne autorisée	<input type="checkbox"/> Pas de diffusion en ligne autorisée	<input type="checkbox"/> Pas de diffusion en ligne autorisée	<input checked="" type="checkbox"/> Pas de diffusion en ligne autorisée	<input type="checkbox"/> Pas de diffusion en ligne autorisée
<b>Existence d'une convention/contrat entre la BCUL et l'ayant-droit</b>	<input checked="" type="checkbox"/> Oui (pour une grande partie des fonds)	<input checked="" type="checkbox"/> Oui (souvent pour les auteurs contemporains)	<input checked="" type="checkbox"/> Oui (pour une grande partie des fonds)	<input type="checkbox"/> Oui	<input type="checkbox"/> Oui	<input checked="" type="checkbox"/> Oui (pour les livres récents)	
	<input type="checkbox"/> Non	<input type="checkbox"/> Non	<input type="checkbox"/> Non	<input checked="" type="checkbox"/> Non	<input checked="" type="checkbox"/> Non	<input type="checkbox"/> Non	

Nom des collections	Archives musicales	Iconopôle	Manuscrits et archives	Documentation vaudoise	Dépôt légal numérique	Réserve précieuse
<b>Remarques</b>	Approximativement, seule 40% de la collection peut être rendue publique.	Tous les cas de figure. Vignette basse définition quel que soit le droit.	Très différent selon les fonds. Les fonds sont toujours sélectionnés en fonction des droits. Certains fonds ne peuvent même pas être numérisés et même certains fonds anciens ne peuvent pas être publiés à cause des conventions.	Fonds prêtables dans le cadre standard des bibliothèques. Ce cadre n'autorise pas de mettre à disposition des versions numériques documents. Réflexion en cours pour rendre disponible la copie numérique de quelques 170 cassettes VHS.	Pas de convention bilatérale directe entre les éditeurs et le DLN. Les choses pourraient évoluer et permettre la publication en open access si les éditeurs l'acceptent.	Beaucoup d'ouvrages anciens dans le domaine public.
					Les documents numériques sont uniquement consultables sur place via un ordinateur dédié.	
<b>Niveaux de préservation</b>						
<b>Code PAC</b>	PAC 1	PAC 1	PAC 1	PAC 2	PAC 1	PAC 1

Nom des collections	Archives musicales	Iconopôle	Manuscrits et archives	Documentation vaudoise	Dépôt légal numérique	Réserve précieuse
<b>Remarques</b>	Les documents nés numériques sont donc des originaux. Les numérisations sont des copies. Il peut s'agir de copie de sauvegarde ou de copie de diffusion. Dans les deux cas, les standards de numérisation sont les mêmes et les documents sont toujours numérisés en entier.			Préservation du contenu obligatoire, mais pas du support en tant que tel.		
<b>Modalités de recherche</b>						
<b>Recherche dans les métadonnées</b>	<input checked="" type="checkbox"/> Standard	<input checked="" type="checkbox"/> Standard	<input checked="" type="checkbox"/> Standard	<input checked="" type="checkbox"/> Standard	<input checked="" type="checkbox"/> Standard	<input checked="" type="checkbox"/> Standard
	<input checked="" type="checkbox"/> Besoin spécifique : Voir Remarques ci-dessous.	<input checked="" type="checkbox"/> Besoin spécifique : voir inventaire. Rubrique description, inscription.	<input checked="" type="checkbox"/> Besoin spécifique : cibler selon le rôle de la personne : auteur, enlumineur, etc.	<input type="checkbox"/> Besoin spécifique :	<input type="checkbox"/> Besoin spécifique :	<input type="checkbox"/> Besoin spécifique :
<b>Recherche dans le texte</b>	<input type="checkbox"/> Exigé (niveau corpus)	<input type="checkbox"/> Exigé (niveau corpus)	<input type="checkbox"/> Exigé (niveau corpus)	<input type="checkbox"/> Exigé (niveau corpus)	<input type="checkbox"/> Exigé (niveau corpus)	<input type="checkbox"/> Exigé (niveau corpus)
	<input type="checkbox"/> Exigé (niveau document)	<input type="checkbox"/> Exigé (niveau document)	<input type="checkbox"/> Exigé (niveau document)	<input type="checkbox"/> Exigé (niveau document)	<input type="checkbox"/> Exigé (niveau document)	<input type="checkbox"/> Exigé (niveau document)

Nom des collections	Archives musicales	Iconopôle	Manuscrits et archives	Documentation vaudoise	Dépôt légal numérique	Réserve précieuse
	<input type="checkbox"/> Souhaité (niveau corpus)	<input type="checkbox"/> Souhaité (niveau corpus) oui	<input type="checkbox"/> Souhaité (niveau corpus) : oui	<input checked="" type="checkbox"/> Souhaité (niveau corpus)	<input checked="" type="checkbox"/> Souhaité (niveau corpus)	<input checked="" type="checkbox"/> Souhaité (niveau corpus)
	<input type="checkbox"/> Souhaité (niveau document)	<input type="checkbox"/> Souhaité (niveau document) oui	<input type="checkbox"/> Souhaité (niveau document) : oui	<input checked="" type="checkbox"/> Souhaité (niveau document)	<input checked="" type="checkbox"/> Souhaité (niveau document)	<input checked="" type="checkbox"/> Souhaité (niveau document)
	<input checked="" type="checkbox"/> Pas nécessaire (niveau corpus)	<input checked="" type="checkbox"/> Pas nécessaire (niveau corpus)	<input type="checkbox"/> Pas nécessaire (niveau corpus)	<input type="checkbox"/> Pas nécessaire (niveau corpus)	<input type="checkbox"/> Pas nécessaire (niveau corpus)	<input type="checkbox"/> Pas nécessaire (niveau corpus)
	<input checked="" type="checkbox"/> Pas nécessaire (niveau document)	<input checked="" type="checkbox"/> Pas nécessaire (niveau document)	<input type="checkbox"/> Pas nécessaire (niveau document)	<input type="checkbox"/> Pas nécessaire (niveau document)	<input type="checkbox"/> Pas nécessaire (niveau document)	<input type="checkbox"/> Pas nécessaire (niveau document)
	<input type="checkbox"/> Non pertinent	<input type="checkbox"/> Non pertinent	<input type="checkbox"/> Non pertinent	<input type="checkbox"/> Non pertinent	<input type="checkbox"/> Non pertinent	<input type="checkbox"/> Non pertinent
<b>Remarques</b>	Il faut pouvoir faire des recherches sur le genre musical et la formation détaillée (instruments) pour les partitions. La recherche texte n'est nécessaire que pour les programmes et les affiches de concerts.		Non pertinent tant que des transcriptions ne sont pas disponibles sur Patrinum. Plusieurs projets en cours pour obtenir des transcriptions.			Nécessaire pour les textes avec reconnaissance OCR
<b>Modalités de consultation</b>						

Nom des collections	Archives musicales	Iconopôle	Manuscrits et archives	Documentation vaudoise	Dépôt légal numérique	Réserve précieuse
<b>Fonctionnalités de consultation en ligne nécessaires / souhaitées</b>	La navigation à l'aide des signets est nécessaire pour les fichiers PDF.	Consultation en salle Consultation sur place sur des ordinateurs dédiés.	Consultation sur place sur des ordinateurs dédiés.	Consultation sur place sur des ordinateurs dédiés.	Consultation sur place sur des ordinateurs dédiés.	Consultation sur place sur des ordinateurs dédiés.
<b>Comptes usagers</b>						
<b>Comptes usagers</b>	<input type="checkbox"/> Nécessaires et obligatoires pour tous les usagers	<input type="checkbox"/> Nécessaires et obligatoires pour tous les usagers	<input type="checkbox"/> Nécessaires et obligatoires pour tous les usagers	<input type="checkbox"/> Nécessaires et obligatoires pour tous les usagers	<input checked="" type="checkbox"/> Nécessaires et obligatoires pour tous les usagers	<input type="checkbox"/> Nécessaires et obligatoires pour tous les usagers
	<input checked="" type="checkbox"/> Nécessaires et facultatifs pour les usagers	<input type="checkbox"/> Nécessaires et facultatifs pour les usagers	<input checked="" type="checkbox"/> Nécessaires et facultatifs pour les usagers	<input type="checkbox"/> Nécessaires et facultatifs pour les usagers	<input type="checkbox"/> Nécessaires et facultatifs pour les usagers	<input type="checkbox"/> Nécessaires et facultatifs pour les usagers
	<input type="checkbox"/> Non nécessaires	<input checked="" type="checkbox"/> Non nécessaires	<input type="checkbox"/> Non nécessaires	<input checked="" type="checkbox"/> Non nécessaires	<input type="checkbox"/> Non nécessaires	<input checked="" type="checkbox"/> Non nécessaires
<b>Fonctionnalités nécessaires liées aux comptes usagers</b>	Il faut pouvoir donner un accès ponctuel à certains documents pour les usagers qui en font la demande.		Il faut pouvoir donner un accès ponctuel à certains documents pour les usagers qui en font la demande.		Consultation uniquement, sans possibilité de télécharger.	
<b>Crowdsourcing</b>						
<b>Description des possibilités de crowdsourcing</b>	Non nécessaire	Serait un plus	Serait un plus pour certains fonds de photos	Serait un plus pour la base des personnalités	Non nécessaire	Non nécessaire

\*Le patrimoine sonore n'est pas inclus à ce tableau, car il n'inclut pas de documents numériques, seules les notices d'autorité étant disponibles sur Patrinum.

### 9.1.3 Tableau récapitulatif des plateformes externes BCUL et BGE

Nom catalogue Numérique	Description	Nb. docs	URL	Langue	Licence	Partage données	Identifiant pérenne	Lien vers notice catalogue source	Compte usagers	Crowdsourcing	Publication sur autres catalogues	Collaboration projets humanités numériques	Modalités de recherches	Téléchargement	Type visionneuse	Outils de présentations grand public	BCUL BGE	Nb. docs BCUL BGE	Versement unique/régulier (U/R)
<b>BHL (Biodiversity Heritage Library)</b>	La biodiversité (relative à l'histoire naturelle et à la botanique)	280'000	<a href="https://biodiversitylibrary.org/">https://biodiversitylibrary.org/</a>	en	Varie selon notice, souvent : CC BY-NC-SA	*APIs *OAI-PMH *Open-Sci *Macaw	✓	Non, mais lien page d'accueil institution	Non	Non	Internet Archives	Oui	Recherche plein texte, avancée (métadonnées), par noms scientifiques et sujets	✓	Recherche dans le texte du document, affichage texte reconnu format texte	Non	BCUL	133	U
<b>e-codices</b>	Les manuscrits suisses (points forts sont l'antiquité tardive et la période allant du moyen-âge jusqu'au temps modernes)	2'600	<a href="https://www.e-codices.unifr.ch/fr">https://www.e-codices.unifr.ch/fr</a>	de, en, fr, it	Varie selon notice, Souvent CC BY-NC	*IIIF *OAI-PMH	✓	Non	Oui mais non activé	Possible pour personnes connectées (ajout notes et références)	Biblissima, Europeana, Gallica, SLSP	Oui	Facettes métadonnées, options de tri, index auteurs, recherche par annotations	✓	Fonctions standard, intégration de Mirador (outil de comparaison entre images)	Non	BCUL	15	U
<b>e-rara</b>	Les imprimés suisses (dont documents iconographiques et cartes) du 15e au 20e siècle	92'000	<a href="https://www.e-rara.ch/">https://www.e-rara.ch/</a>	de, en, fr, it	Domaine public ou CC BY-SA	*IIIF *OAI-PMH	✓	Oui	Non	Non	Gallica	Oui	Recherche plein texte et avancée. Facettes disponibles	✓	Recherche texte dans document. Réglage luminosité, contraste, négatif, saturation, etc.	Sélection thématique	BGE	6'300	R
<b>RISM Suisse</b>	Partitions manuscrites et imprimées et les écrits sur la musique conservés dans les bibliothèques et archives suisses	101'000 dont 10'300 documents numérisés	<a href="https://rism.digital/rism-ch.html">https://rism.digital/rism-ch.html</a>	de, en, fr, it	CC BY	*IIIF *SPAQL *SRU	✓	Oui	Non	Non	-	Oui	Recherche simple et application de filtres possibles	✓	Fonction standard, n'est disponibles que pour les documents accessibles via IIIF	Non	BCUL	1300	U
																	BGE	1700	?

Nom catalogue Numérique	Description	Nb. docs	URL	Langue	Licence	Partage données	Identifiant pérenne	Lien vers notice catalogue source	Compte usagers	Crowdsourcing	Publication sur autres catalogues	Collaboration projets humanités numériques	Modalités de recherches	Téléchargement	Type visionneuse	Outils de présentations grand public	BCUL BGE	Nb. docs BCUL BGE	Versement unique/régulier (U/R)
<b>Onstage</b>	Contient des programmes de concerts, conférences, cours, etc. depuis 1826 en lien avec la musique.	12'500	<a href="https://onstage.rism.digital">https://onstage.rism.digital</a>	de, en, fr, it	Bibliothèque détentrice	fichiers TEI (XML) accessibles via requêtes XPath	?	Non	Non	Non	Non	Non	Recherche simple avec possibilité de choisir le champ.	Non	Basique	Non	BGE	2'300	U
<b>notreHistoire.ch</b>	Les documents ayant trait à l'histoire commune des Romands (images, vidéos, sons, textes documentaires)	107'000	<a href="https://notrehistoire.ch/">https://notrehistoire.ch/</a>	fr	Documents sous droit	Non	X	Non	Oui	Oui, la plateforme repose sur ce principe (ajout contenu, créer collections, commentaire, favoris, préféré)	Non	Non	Recherche plein texte. Recherche par texte, membre, galerie, lieu, date	Non	Fonctions basiques	Collections thématiques des usagers	BGE	16'000	-
<b>Memobase</b>	Le patrimoine audiovisuel suisse	400'000 dont 323'000 consultables sur le web	<a href="https://memobase.ch">https://memobase.ch</a>	de, fr, it	Varie selon document, majorité avec droit d'auteur inconnu	*OAI-PMH *IIIF *API *SRU *SPARQL	✓	Oui	Non	Lien pour envoyer par e-mail un feedback depuis chaque objet	Europeana, **Connectome sera utilisé pour ajouter des infos géographiques.	Oui	Recherche simple et avancée, facettes,	Non	Fonctions basiques	Vitrines (exposition virtuelle)	BCUL	À venir (photos et films fixes)	-
<b>Connectome</b>	Mise à disposition des données de la	Projet en cours	<a href="https://www.swit.ch/c">https://www.swit.ch/c</a>	en	?	*APIs *Sparql	✓	?	Oui	Pas prévu à priori	?	Oui, en lien	Facettes disponible, paramétra	?	A priori pas	A priori pas	BCUL	sélection de manuscrits	-

Nom catalogue Numérique	Description	Nb. docs	URL	Langue	Licence	Partage données	Identifiant pérenne	Lien vers notice catalogue source	Compte usagers	Crowdsourcing	Publication sur autres catalogues	Collaboration projets humanités numériques	Modalités de recherches	Téléchargement	Type visionneuse	Outils de présentations grand public	BCUL BGE	Nb. docs BCUL BGE	Versement unique/régulier (U/R)
	recherche et celles des institutions GLAM au niveau de la Suisse		connectome									avec SATW	ge de l'algorithme				BGE	Probablement ?	
Wikipédia	Encyclopédie libre et participative en ligne.	Près de 2.5 millions d'articles en français	<a href="https://fr.wikipedia.org">https://fr.wikipedia.org</a>	Plus de 300 langues	CC BY-SA	*Media Wiki API *Wikidata API	X	Varie selon l'article	Oui	Oui, ajout de contenu, modifications, participation à la discussion.	*DBpedia *Wikidata	Oui	Recherche simple	✓	-	-	BCUL	Plus de 800	R
																	BGE	?	R
Google Books	Livres numérisés dans le monde entier	Plus de 25 millions	<a href="https://books.google.ch">https://books.google.ch</a>	400 langues	Varie en fonction du document et du pays de consultation	*APIs	X	Non	Oui	Possibilité de laisser des commentaires et donner un avis	-	Oui	Recherche simple et avancée, mais peu de critères disponibles dans la recherche avancée	Non	Adaptée à la lecture, possibilité de recherche dans le texte, permet de sélectionner et partager un extrait. Texte reconnu automatiquement	-	BCUL	100'000	U
réro doc	Ancienne plateforme du REseau ROmand RERO pour mettre à disposition les documents numériques des bibliothèques membre	82'000	<a href="http://doc.rero.ch">http://doc.rero.ch</a>	de, en, fr, it	Non précisée	*OAI-PMH	?	lien vers page d'accueil du site	Oui	Non	?	?	Recherche simple et avancée	Oui	Basique	Non	BCUL	-	-
																	BGE	2'400	-

Nom catalogue Numérique	Description	Nb. docs	URL	Langue	Licence	Partage données	Identifiant pérenne	Lien vers notice catalogue source	Compte usagers	Crowdsourcing	Publication sur autres catalogues	Collaboration projets humanités numériques	Modalités de recherches	Téléchargement	Type visionneuse	Outils de présentations grand public	BCUL BGE	Nb. docs BCUL BGE	Versement unique/régulier (U/R)
egodocuments	Liste des écrits personnels entre 1500 et 1820 conservés en Suisse	1'700	<a href="https://wp.unil.ch/egodocuments">https://wp.unil.ch/egodocuments</a>	de, en, fr, it	Non précisée	Non	X	Oui, si disponible	Non	Non	Non	Non	Recherche simple et avancée	-	-	-	BCUL	33	?
																	BGE	125	?
Lumières.Lausanne	Plateforme mettant à disposition des transcriptions de sources inédites, de la littérature secondaire et des fiches biographiques en lien avec Lausanne ou le pays de Vaud	?	<a href="https://lumières.unil.ch">https://lumières.unil.ch</a>	fr	Non précisée	Non	X	-	Oui	Non	Non	?	Recherche simple et avancée	Oui, quand disponible	-	Pages de présentation sur le site web	BCUL	230	?
																	BGE	120	?
From the page	Plateforme collaborative pour la transcription de documents manuscrits	145'000	<a href="https://fromthepage.com">https://fromthepage.com</a>	en, es, pt	Non précisée	*IIIF *API	X	Non	Oui	Oui : transcription, commentaires	?	Oui	Recherche simple	Non	-	Non	-	-	-
CCSA	Affiches suisses dès 1950	74'000	<a href="https://www.posters.nb.admin.ch">https://www.posters.nb.admin.ch</a>	de, en, fr, it	Réutilisation libre pour les métadonnées et vignettes	*OAI-PMH *Z39.50	✓	Pas de lien url mais numéro de cote	Non	Possibilité de laisser un commentaire	Non	Non	Recherche simple et avancée, facettes disponibles	Oui	Pas de visionneuse, affichage simple de l'image	Non	BCUL	1'020	?
																	BGE	26'000	R
Daguerreobase	Catalogue collectif pour les daguerréotypes	16'700	<a href="http://www.daguerreobase.org">http://www.daguerreobase.org</a>	en	Spécifiée pour chaque image	*RDF	?	Oui	Non	Possibilité de laisser un commentaire	Europeana	Non	Recherche simple avec facettes	Vignette uniquement	Avancée avec nombreux réglages dont luminosité, saturation, etc.	Non	BGE	90	U

### 9.1.4 Tableau récapitulatif des méta-catalogues internationaux

Nom catalogue Numérique	Description	Nb. docs	URL	Licence	Partage données	Identifiant pérenne	Lien vers notice catalogue source	Compte usagers	Crowdsourcing	Publication sur autres catalogues	Collaboration projets humanités numériques	Modalités de recherches	Format des métadonnées	Téléchargement	Visionneuse	Outils de présentation grand public
<b>Wikimedia Commons</b>	Plateforme hébergeant le contenu multimédia de l'ensemble des projets de la Fondation Wikimedia	87,5 mio de fichiers dont 56 mio d'images	<a href="https://commons.wikimedia.org">https://commons.wikimedia.org</a>	Domaine public ou CC BY-SA	*APIs *IIIF	-	Oui	Oui	Oui	*Wikidata *DBpedia	*Wikidata visualizations *Wikipedia visualizations *App gallery	Recherche simple	Format spécifique	✓	Oui, basique	Non
<b>Wikisource</b>	Bibliothèque numérique multilingue des ouvrages sous licence libre	Environ 345 K textes en français (560 K en anglais, 51 K en allemand)	<a href="https://fr.wikisource.org">https://fr.wikisource.org</a>	Essentiellement domaine public (parfois sous licence libre)	*APIs	-	Oui	Oui	Oui	*Wikidata *DBpedia	Oui, <a href="https://de.wikisource.org/wiki/Wikisource:Digital_Humanities">https://de.wikisource.org/wiki/Wikisource:Digital_Humanities</a>	Recherche simple	Format spécifique	✓	Non	Non
<b>Europeana</b>	Le patrimoine culturel européen	30 Mio images ; 22,5 Mio textes ; 771 K audios ; 338 K vidéos ; 5 K objets 3D	<a href="https://www.europeana.eu">https://www.europeana.eu</a>	CC0 ou domaine public	*APIs REST *IIIF *SPARQL	URI	Oui	Oui	Oui, <a href="https://crowdheritage.eu">https://crowdheritage.eu</a>	-	Oui, <a href="https://pro.europeana.eu/about-us/projects">https://pro.europeana.eu/about-us/projects</a>	Recherche simple, filtres disponibles	<a href="#">Europeana Data Model</a> (EDM)	✓	Oui, pour tous les types de documents	Oui
<b>Internet Archives</b>	Donner un accès universel à toutes les connaissances	726 Mia pages web ; 35 Mio textes ; 8,4 Mio vidéo ; 14,5 Mio audio ; 877 K logiciels ; 4,4 Mio images	<a href="https://archive.org">https://archive.org</a>	Spécifiée pour chaque doc	*APIs *IIIF	-	Non, mais collection relative à l'institution	Oui	Oui	-	Oui. Exemple : <a href="https://digital.humanities.ox.ac.uk/project/internet-archive-live-music-archive-linked-data">https://digital.humanities.ox.ac.uk/project/internet-archive-live-music-archive-linked-data</a>	Recherche simple, filtres disponibles	Non standard	✓	Oui, pour tous les types de documents	Non
<b>Google Arts and Culture</b>	Mettre en ligne l'art et la culture au niveau mondial	Collaboration avec plus de 2'000 institutions	<a href="https://artsandculture.google.com">https://artsandculture.google.com</a>	<a href="#">Licence libre ou « cleared »</a>	*APIs	-	Possible	Oui	Non	-	Oui, <a href="https://experiments.withgoogle.com">https://experiments.withgoogle.com</a>	Recherche basique	Non standard	Possible	Oui	Oui, beaucoup d'outils
<b>Gallica</b>	Bibliothèque numérique de la BnF et de ses partenaires	9 Mio	<a href="https://gallica.bnf.fr">https://gallica.bnf.fr</a>	CC BY-NC	*APIs *IIIF *OAI-PMH *Sparql	ARK	Oui	Oui	Non	*Europeana *Internet Archives *Wikisource *Wikimedia Commons	Oui, <a href="https://odhn.enspsl.eu/liste/etiquettes/gallica">https://odhn.enspsl.eu/liste/etiquettes/gallica</a>	Recherche simple et avancée	*MPEG-21 DIDL (périodiques et lots d'images) *Dublin core (pour le reste)	✓	Oui, pour tous les types de documents	Oui : blog et « à la loupe »

Nom catalogue Numérique	Description	Nb. docs	URL	Licence	Partage données	Identifiant pérenne	Lien vers notice catalogue source	Compte usagers	Crowdsourcing	Publication sur autres catalogues	Collaboration projets humanités numériques	Modalités de recherches	Format des métadonnées	Téléchargement	Visionneuse	Outils de présentation grand public
Hathi Trust	Corpus pour la recherche universitaire issus de bibliothèques universitaires et de recherche	17,6 Mio dont 7 Mio dans le domaine public	<a href="https://www.hathitrust.org">https://www.hathitrust.org</a>	Spécifiée pour chaque doc	*Outils <a href="#">HTRC</a> *APIs *OAI- <a href="#">PMH</a>	-	Mention de la source, pas de lien	Oui	Non	-	Outils disponibles, mais difficulté à trouver des projets accessibles en ligne	Recherche simple et avancée	Marc	Restreint	Oui, fonctions avancées	Non

## 9.2 Questions posées aux responsables de collections lors des entretiens

- Description des différentes collections numériques (ce qui inclut les collections nées numériques, les collections numérisées et les bases de données)
  - Types des documents
  - Formats de fichiers
  - Format de métadonnées
  - Licence
  - Plateforme interne utilisée pour la mise à disposition
  - Accessibilité des données pour la réutilisation
  - Modalités de recherche nécessaire
    - Utilité de la recherche plein texte
  - Nécessité d'un compte usager
    - Fonctionnalités requises
    - Crowdsourcing envisagés ?
- Parmi tous les documents de vos collections :
  - Quelle est la proportion des documents chargés sur une plateforme BCUL ?
  - Quelle est la proportion des documents disposants d'une notice sur une plateforme BCUL ?
  - Est-ce que le nombre de documents et de notices disponibles sur la plateforme BCUL sont suffisants pour pouvoir élaborer des corpus de données porteurs de sens ?
- Quelles sont les collections sur lesquelles un effort de catalogage et/ou de numérisation particulier est fourni ?
- Est-ce que tout ou partie de vos collections ont déjà été versées dans le cadre de projets externes (par exemple méta-catalogue externes) ?
- Quels sont vos challenges actuels par rapport à l'exploitation des données numériques ?

### 9.3 Politique des bibliothèques numériques de la BCUL



## Politique des bibliothèques numériques BCU Lausanne – Solution A

Dernière mise à jour : TN, le 28 janvier 2019

Version 3-A



**Bibliothèque cantonale et universitaire – Lausanne**  
Site Unithèque - CH-1015 Lausanne  
[www.bcu-lausanne.ch](http://www.bcu-lausanne.ch)

## Table des matières

1	Introduction.....	3
2	Terminologie.....	3
3	Politique des bibliothèques numériques de la BCU Lausanne .....	4
3.1	Bibliothèques numériques .....	4
3.1.1	Bibliothèques numériques externes .....	4
3.1.2	Serval .....	4
3.1.3	Scriptorium .....	4
3.1.4	Patrinum.....	5
3.1.5	Alma digital.....	5
3.2	Diffusion .....	5
4	Responsabilité et processus de décision.....	5
4.1	Coordination .....	5
4.2	Collections .....	5
4.3	Bibliothèques numériques .....	6
4.4	Décisions .....	6
Annexe .....		6
	Listes des bibliothèques numériques actuellement utilisées par la BCU Lausanne.	6

## 1 Introduction

D'une part, la BCU Lausanne développe et participe à un nombre croissant de bibliothèques numériques. D'autre part, elle poursuit ses efforts en matière de numérisation de collections « papier » et d'acquisition de ressources nées numériques.

A titre d'exemple, notons

le développement de

- Scriptorium ([scriptorium.bcu-lausanne.ch](http://scriptorium.bcu-lausanne.ch)) qui permet la diffusion de la presse vaudoise
- Patrinum ([patrinum.ch](http://patrinum.ch)) qui permet la conservation et la diffusion du patrimoine documentaire vaudois

la participation à diverses bibliothèques numériques collaboratives

- e-codices ([e-codices.ch](http://e-codices.ch))
- e-rara ([e-rara.ch](http://e-rara.ch))
- biodiversity heritage library ([biodiversitylibrary.org](http://biodiversitylibrary.org))
- etc.

l'existence d'importants projets et programmes de numérisation passés et présents

- Googlos (en collaboration avec Google)
- Presse vaudoise (Scriptorium)
- Collection iconographique (fête des vigneronns)
- etc.

l'acquisition de nombreuses ressources nées numériques à travers le dépôt légal numérique et d'autres sources.

La conséquence est que chaque document (ou collection) numérique peut potentiellement être diffusé à travers un choix de bibliothèques numériques<sup>1</sup>. De manière à faciliter ces choix et à assurer une cohérence nécessaire pour un travail rationnel et un accès facilité pour les usagers, le présent document donne les lignes directrices pour une bonne politique des bibliothèques numériques à la BCU Lausanne.

Cette politique sera mise à jour en fonction des changements qui auront lieu dans le futur.

## 2 Terminologie

**Bibliothèque numérique** : site web mettant à disposition une collection de documents numériques.

**Collection** : un ensemble défini de documents.

---

<sup>1</sup> Voir l'annexe du présent document qui liste les bibliothèques numériques utilisées par la BCU Lausanne.

Sauvegarder (un fichier) : action consistant à copier un fichier sur un support de stockage adapté.

Conserver (un fichier) : action consistant à prendre toutes les mesures nécessaires pour conserver un fichier à long terme.

### **3 Politique des bibliothèques numériques de la BCU Lausanne**

#### **3.1 Bibliothèques numériques**

Chaque document numérique est sauvegardé dans une unique bibliothèque numérique par souci de bonne gestion des ressources disponibles.

##### **3.1.1 Bibliothèques numériques externes**

Lorsqu'il existe une bibliothèque numérique externe qui

- est pertinente pour la collection concernée
- est une bibliothèque numérique faisant partie des systèmes incontournables dans le domaine
- est satisfaisante du point de vue des fonctionnalités offertes
- est prometteuse en termes de durabilité et de développements futurs
- offre des conditions financières supportables par la BCUL
- ne fait pas double-emploi avec une bibliothèque numérique gérée par la BCUL

alors elle est utilisée pour la collection en question.

Exemples : e-codices, biodiversity heritage library, etc.

##### **3.1.2 Serval**

Le SERveur Académique Lausannois (Serval) contient des collections numériques dont la principale est le dépôt institutionnel de l'Université de Lausanne et du CHUV. Les publications des chercheurs de ces deux institutions y sont signalées et parfois déposées. Les publications déposées y sont préservées et disponibles en libre accès pour certaines.

Il s'agit donc d'un outil de l'UNIL et la BCUL a un rôle de soutien en participant

- aux travaux de publication des thèses (saisie des notices)
- à la validation des notices<sup>2</sup>
- aux tâches du Help desk
- au comité exécutif de Serval.

##### **3.1.3 Scriptorium**

Scriptorium est utilisé pour la presse vaudoise et quelques collections complémentaires à définir au cas par cas.

---

<sup>2</sup> Cette tâche disparaîtra lors de la prochaine mise à jour de Serval.

Par exemple, les annuaires vaudois et le bulletin du grand conseil sont de telles collections complémentaires.

#### **3.1.4 Patrinum**

Patrinum est utilisé pour les collections patrimoniales (hors presse vaudoise numérisée) et comme archive ouverte/dépôt institutionnel pour la HEP (mémoires des étudiants), certaines HES (mémoires des étudiants) et la Bibliothèque du Musée et jardins botaniques cantonaux (articles).

#### **3.1.5 Alma digital**

Alma digital (un module d'Alma) est utilisé pour la littérature grise et les ressources académiques non patrimoniales. Seules les ressources textuelles sont acceptées.

### **3.2 Diffusion**

Pour assurer une bonne diffusion des ressources numériques, la BCUL exploite diverses techniques permettant de diffuser ces ressources à travers les bibliothèques numériques et outils de découverte pertinents.

Les techniques utilisées et envisagées comprennent le protocole OAI-PMH et les APIs définis par le standard IIIF.

En particulier, un effort est fourni pour assurer la meilleure compatibilité possible entre les systèmes gérés par la BCU Lausanne et Renouvaud.

D'autres travaux en ce sens sont entrepris au cas par cas en fonction des collections concernées.

## **4 Responsabilité et processus de décision**

### **4.1 Coordination**

Le coordinateur ingest est responsable de la mise à jour de la présente politique. Les mises à jour sont soumises à la direction pour validation.

Le coordinateur ingest tient à jour un inventaire des bibliothèques numériques utilisées par la BCUL et un inventaire des ressources numériques que la BCUL possède.

Il développe également une vue d'ensemble et il maintient une cohérence de l'offre numérique de la BCUL.

Note : ceci n'inclut pas les ressources numériques pour lesquelles la BCUL achète un droit d'accès.

### **4.2 Collections**

Chaque nouvelle collection de ressources numériques est évaluée de concert par le responsable de collection et le coordinateur ingest à l'aide de la grille ad hoc<sup>3</sup>. La

---

<sup>3</sup> Cette grille est un document annexe du présent document.

bibliothèque numérique pertinente pour accueillir cette collection est choisie en fonction de cette évaluation.

### 4.3 Bibliothèques numériques

L'introduction d'une bibliothèque numérique dans l'inventaire des bibliothèques numériques de la BCUL est évaluée par le coordinateur ingest et le(s) responsable(s) de collection concerné(s) à l'aide de la grille ad hoc<sup>3</sup>.

### 4.4 Décisions

La mise en ligne d'une nouvelle collection et l'introduction d'une nouvelle bibliothèque numérique sont soumises à la direction pour validation.

## Annexe

### Listes des bibliothèques numériques actuellement utilisées par la BCU Lausanne

Bibliothèques numériques gérées par la BCU Lausanne ou Renouvaud		
Nr	Nom	Description (état à la date de mise à jour de la présente politique)
01	<i>Scriptorium</i>	<i>Presse vaudoise et collections complémentaires. Le système utilisé est MediaINFO, développé et mis à jour par la société Geneza<sup>i</sup>. Ce système offre des fonctionnalités de consultation avancées.</i>
02	<i>Patrinum</i>	<i>Collections patrimoniales (BCUL) et archive ouverte (mémoires d'étudiants de certaines HES et de la HEP et articles de la Bibliothèque du Musée et jardins botaniques cantonaux). Le système utilisé est basé sur Invenio, et est implémenté et développé par la société TIND<sup>ii</sup>. Il s'agit d'un système complet qui offre également des fonctionnalités de gestion de collection physique et de conservation du numérique.</i>
03	<i>eLectures</i>	<i>Système de prêt pour ebooks grand public. Développé et mis à jour par Cantook. La BCUL peut également intégrer ses propres fichiers epub qui seraient alors en libre téléchargement.</i>
04	<i>Alma digital</i>	<i>Ressources académiques et littérature grise. Alma digital est un module d'Alma. Alma digital devrait entrer en production au printemps</i>

		2019.
Bibliothèques numériques gérées par une institution tierce		
05	Google Books	Collaboration avec Google, qui s'est chargé de la numérisation et de la mise en ligne. Système propriété de Google.
06	<i>BHL</i>	<i>Publie des documents numérisés ayant pour thème la biodiversité. Bibliothèque numérique gérée et coordonnée par le réseau de bibliothèques du Smithsonian Institution.</i>
07	<i>e-rara.ch</i>	<i>Contient des imprimés numérisés des bibliothèques suisses. Bibliothèque numérique gérée et coordonnée par la bibliothèque de l'EPFZ.</i>
08	<i>e-codices.ch</i>	<i>Contient des manuscrits numérisés suisses. Bibliothèque numérique gérée et coordonnée par l'Université de Fribourg.</i>
09	<i>Rero doc</i>	<i>Bibliothèque numérique gérée et coordonnée par Rero. Le canton de Vaud compte s'en retirer au profit de Patrinum durant la période 2018-19.</i>
10	<i>Serval</i>	<i>Serveur académique lausannois : Bibliothèque numérique développée et mise à jour par l'Université de Lausanne. La BCUL participe à la gestion de Serval.</i>

<sup>i</sup> Geneza est une société, basée en Slovénie et aux Emirats arabes unis, qui développe et propose le système de bibliothèque numérique MediaINFO.

<sup>ii</sup> TIND est une société issue du CERN (spin-off). Elle propose ses services pour l'implémentation du/de la SIGB/Bibliothèque numérique open source Invenio.