



HAL
open science

Contribution à la méthodologie du Data paper par la valorisation de données sur l'enseignement musical

Lorraine Roubertie Soliman, Odile Tripier-Mondancin, Adrien Bourg

► To cite this version:

Lorraine Roubertie Soliman, Odile Tripier-Mondancin, Adrien Bourg. Contribution à la méthodologie du Data paper par la valorisation de données sur l'enseignement musical. *Journal de Recherche en Education Musicale*, 2024, 15 1-2 (15), pp.95-110. 10.4000/134z5 . hal-04909779

HAL Id: hal-04909779

<https://hal.science/hal-04909779v1>

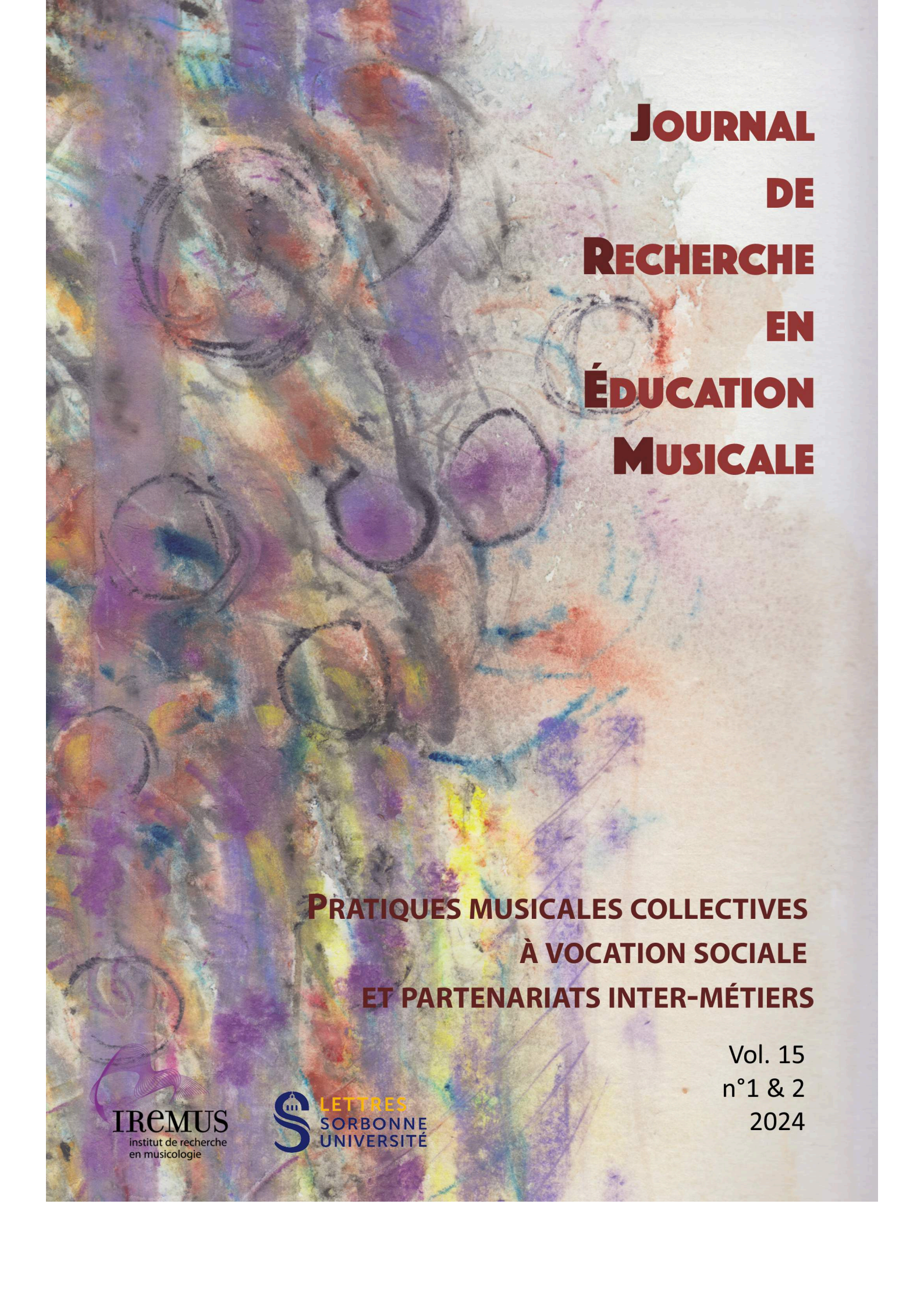
Submitted on 24 Jan 2025

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

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



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



**JOURNAL
DE
RECHERCHE
EN
ÉDUCATION
MUSICALE**

**PRATIQUES MUSICALES COLLECTIVES
À VOCATION SOCIALE
ET PARTENARIATS INTER-MÉTIERS**

Vol. 15
n°1 & 2
2024

**IREMUS**
institut de recherche
en musicologie

 **LETTRES**
SORBONNE
UNIVERSITÉ

Contribution à la méthodologie du *Data paper* par la valorisation de données sur l'enseignement musical

Le projet MusiPim, Musique et Partenariat Inter-Métiers

Lorraine Roubertie Soliman*, Odile Tripier-Mondançin**, Adrien Bourg***

* Université Toulouse-Jean Jaurès, LLA-CREATIS, LabEx SMS

** Université Toulouse-Jean Jaurès, UMR EFTS, LabEx SMS

*** Sorbonne Université, IReMus, UMR 8223

Introduction

Le présent article s'attache à poser les enjeux et décrire les étapes liées à la rédaction d'un *Data paper* en Sciences humaines et sociales (SHS), plus spécifiquement en sciences de l'éducation et de la formation musicale (SEFM).

Ces pratiques d'écriture et de publication sont récentes d'une manière générale, et nouvelles pour les auteurs de la présente contribution. Par ce partage, nous souhaitons participer au mouvement international de valorisation des données de recherche, dans le cadre de la science ouverte, en fournissant une aide aux chercheurs, notamment dans les disciplines où cette pratique est encore très peu répandue.

La première partie synthétise les principales finalités, caractéristiques et structurations d'un *Data paper* en SHS. La deuxième partie pose, pas à pas, les exigences méthodologiques en amont de la rédaction d'un *Data paper*. La troisième partie correspond à une brève discussion sur les difficultés rencontrées lors de l'expérience de rédaction d'un *Data paper* dans le cadre du projet MusiPim, Musique et Partenariat Inter-Métiers, qui regroupe des corpus témoignant de situations d'enseignement-apprentissage de la musique en orchestre.

Enjeux du *Data paper*

Caractéristiques et finalités du *Data paper*

Un *Data paper* est un article scientifique qui décrit un jeu de données scientifiques, à l'aide de métadonnées (informations ou données servant à définir ou à décrire d'autres données). Plus précisément, « Les *Data papers* sont des articles rédigés, révisés par des pairs et cités dans des revues académiques ou savantes, dont le contenu principal est une description des jeux de données de recherche publiés, ainsi que des renseignements contextuels sur la production et l'acquisition des données, dans le but de faciliter l'accessibilité, la disponibilité et la réutilisation des données de recherche. Ils sont intégrés à la gestion des données de recherche et liés aux entrepôts de données » (Schöpfel, Farace, Prost et Zane, 2020, p. 623).

Le *Data paper* vise à valoriser un ou plusieurs jeux de données, à leur accorder une seconde vie en permettant à l'équipe productrice ou à d'autres chercheurs de les utiliser pour mener de nouveaux projets, participer à des méta-analyses, ou encore dans un but de reproductibilité. Les *Data papers* participent de la science ouverte. Il faut souligner que lorsqu'elles correspondent à des données financées par des fonds publics, les données de recherche sont dorénavant soumises à un principe d'ouverture par défaut (Open Data), introduit par la Loi pour une République numérique (LRN, 2016).

Les chercheurs ont été interpellés par la Ministre de l'Enseignement Supérieur de la Recherche et de l'innovation le 4 juillet 2018 (Premier plan national pour la science ouverte) à faire le nécessaire pour développer « *une démarche structurée de conservation de leurs données [...] pour que ces données obéissent au principe FAIR*'' [Findable, Accessible, Interoperable, Reusable¹] » (cit. in Tripier-Mondancin, 2022, p. 330). Une telle démarche implique de réfléchir 1) au cycle des données de la conception du projet de recherche jusqu'à son achèvement et 2) à la manière dont il sera le plus judicieux de les conserver pour rendre leur partage le plus efficace possible, à court, moyen et long termes. La rédaction du *Data paper* constitue une étape majeure de ce cycle. Ce type d'article a pour but de rendre disponibles ces données en indiquant les circonstances et méthodes de leur collecte, en montrant leur originalité, leur fiabilité ainsi que leur potentiel de réutilisation. Contrairement à l'article scientifique, l'enjeu ne consiste pas à relater les hypothèses, ni les conclusions de la recherche.

La préparation d'un *Data paper* participe au développement d'une culture de la constitution, de la gestion, de la collecte, de l'archivage et si possible du partage, de données fiables, dans un cadre éthique et responsable.

En Sciences de l'éducation et de la formation, tout comme dans les Sciences Humaines et Sociales (Gay, 2021 ; Dedieu, 2022) mais aussi en musicologie, la publication de *Data papers* est une pratique encore peu répandue. Comme n'importe quel article scientifique, le *Data paper* n'est publié qu'après expertise, incluant une vérification des possibilités d'accès aux données (Dedieu, 2022, p. 1).

Ces constats légitiment la rédaction du présent article qui vise 1) à informer la communauté des chercheurs, et en particulier ceux des deux disciplines précitées, 2) à exposer pas à pas les étapes nécessaires à la réalisation d'un *Data paper*. L'exemple proposé s'adosse de manière centrale à deux jeux de données collectées au moment où la démarche de la science ouverte commençait à se diffuser. Le premier correspond aux données collectées par le groupe de recherche collaborative EMIS (Éducation Musicale et Intégration Sociale), à propos des orchestres à l'école et au collège, à Toulouse, de 2017 à 2021. Le deuxième jeu de données correspond à la collecte dans le cadre du projet de recherche-action autour de Démos (Dispositif d'Éducation Musicale et Orchestrale à vocation Sociale), à Clermont-Ferrand, de 2019 à 2021. À ces deux corpus s'ajoute un troisième jeu de données collectées entre mai et décembre 2024, dans le cadre du projet MusiPim, projet qui constitue une extension des deux précédents.

Types de structurations attendues

D'une revue à l'autre, y compris dans un même champ disciplinaire, la longueur et le format des *Data papers* varient grandement, du plus « synthétique » à « un article très complet » (Dedieu, 2022, p. 1). Leur structuration est inspirée de celle, « éprouvée » (Gay, 2021, p. 2), des *Data papers* dans les sciences dites dures. Dans son article de 2014, Dedieu présente quatre types de structures de *Data papers* spécifiquement pour les sciences du vivant (p. 3). Dans celui de 2022, elle propose une synthèse. Dans le tableau 1, nous ajoutons à ces cinq modèles de structuration, la proposition de Gay, plus spécifiquement pensée pour les SHS (2021).

Généralement, un *Data paper* est structuré en trois parties (tableau 1) : 1) description du ou des jeu(x) de données, 2) description des méthodes d'obtention de ces données et 3) explication du potentiel de réutilisation (dans les exemples présentés, cette dernière catégorie n'est pas systématique). Moins fréquemment, le lecteur peut y trouver un résumé des résultats de recherche.

1 Faciles à trouver, accessibles, interopérables et réutilisables.

Champs	Exemples	Structuration							
Sciences du vivant	Ex. n° 1 (Dedieu, 2014)	Introduction (contexte, question de recherche, objectifs de l'étude)	Présentation des métadonnées Description du ou des jeu(x) de données			Description du format des fichiers	Description du statut, de la qualité et de l'accessibilité aux données	Contexte de l'étude	Méthodes et procédures
	Ex. n° 2 (Dedieu, 2014)	Introduction (idem)	Présentation des métadonnées						
		Couvertures taxonomique, spatiale et temporelle	Description du projet de recherche	Méthodes		Description du ou des jeu(x) de données			
	Ex. n° 3 (Dedieu, 2014)	Tableau de spécifications des données	Design expérimental, matériels et méthodes			Discussion mettant en valeur la portée du jeu de données			
	Ex. n° 4 (Dedieu, 2014)	Contexte scientifique, question de recherche, objectifs de l'étude, valeur et potentiel des données pour leur réutilisation	Méthodes, procédures et analyses, protocole expérimental (figure) et variables (tableau)	Description des données et de l'entrepôt, format des fichiers (tableau)	Présentation des métadonnées (tableaux) : processus expérimental, méthode d'échantillonnage, description du site, conditions d'expériences, protocoles suivis, données obtenues...		Validation technique (explication de la rigueur scientifique de l'étude et de la qualité techniques des données)		Notes d'usage (facultatives) pour faciliter l'usage des données par d'autres chercheurs
	Synthèse (Dedieu, 2022)	Description du jeu de données et de son contexte	Description des méthodes d'obtention		Explication du potentiel de réutilisation des données				
Sciences humaines et sociales	Ex. n° 1 (Gay, 2021)	Introduction (contexte et résumé)	Description des méthodes et du matériel utilisé pour produire les données	Description du ou des jeu(x) de données (variables, noms des fichiers, localisation, formats et poids numérique)	Description des métadonnées		Validation technique (procédures ayant permis de confirmer la validité des données)	Notes d'usage	

Tableau 1 : Exemples de structurations de *Data papers* en sciences du vivant et en SHS (d'après Dedieu, 2014, 2022 et Gay, 2021)

Le groupe de travail MusiPim a choisi pour le *Data paper* en cours d'écriture, la structuration suivante : 1) Introduction (contexte de l'étude, arrière-plan des deux études antérieures contributives à la nouvelle enquête, questions de recherche à l'origine des deux premières collectes de données, nouvelle collecte et nouvelle problématique, plus-value de la collecte), 2) Description des méthodes et du matériel dans les trois études, 3) Description des données et de l'entrepôt où elles sont déposées, 4) Description des métadonnées, 5) Information et discussion justifiant la fiabilité des données, 6) Potentiel de réutilisation, conseils, 7) Licence, 8) Références bibliographiques, 9), Figures, tableaux annexes, 10) Remerciements.

Exigences de la rédaction / préparer la rédaction du *Data paper*

Étape 0 — Protection des Données (DPO), comités éthiques

Toute nouvelle étude devrait prendre en considération les principes de protection des données en parallèle des choix méthodologiques. Ainsi, les formalités exigées pour être en conformité avec le Règlement Général de Protection des Données (RGPD²) de 2018 devraient être connues en amont de toute étude, et les démarches effectuées au fur et à mesure. Les exigences propres au DPO (Data Protection Officer, ou délégué à la protection des données³) de l'institution de rattachement devraient être prises en compte, par exemple la déclaration d'un traitement de données au démarrage de l'étude et, éventuellement, la soumission du protocole de recherche à un comité éthique CER ainsi que devant les Comités de Protection des Personnes (CPP) qui sont chargés d'émettre un avis préalable sur les conditions de validité de toute recherche impliquant la personne humaine, au regard des critères définis par l'article L 1123-7 du Code de la Santé Publique (CSP).

La protection des données est généralement relative à leur mise en ligne sur un entrepôt public, ce qui implique de recourir à au moins deux types de documents, à destination des enquêtés : 1/ une demande d'autorisation d'enregistrement et de cession de l'image et de la voix, et 2/ un formulaire d'information et de consentement éclairé.⁴ Ces formulaires doivent être préparés le plus tôt possible, en lien étroit avec le DPO. Dans le cas où plusieurs instances (universités, laboratoires, académies...) sont impliquées dans l'étude, il convient également de prévoir un Accord de responsabilité conjointe signé par les différentes directions.

D'une manière générale, nous conseillons de recourir à l'arbre de décision proposé par le Cirad sur le site <https://www.loginos.net/base/ylxxVO> : « L'arbre de décision proposé vise à guider les scientifiques dans le processus de diffusion des données de recherche. Il permet d'identifier les points de vigilance et, le cas échéant, les actions à mettre en œuvre avant la diffusion des données ». Selon les informations fournies par les chercheurs, des préconisations sont données à la fin du questionnaire.

Étape 1 — Gestion des Données

La tenue d'un Plan de Gestion des Données (PGD), ou Data Management Plan (DMP), est une condition *sine qua non* de la démarche de valorisation des données selon le principe FAIR. L'objectif du PGD est d'assurer au mieux un bon suivi des données jusqu'à l'étape du partage. La rédaction du PGD est une première étape de travail qui s'actualisera tout au long du projet de recherche : le PGD est évolutif en fonction des apports de données (anciennes et nouvelles). Il existera donc plusieurs versions du PGD.

Le PGD expose : 1/ l'ensemble des données collectées, la manière dont elles ont été produites et collectées ; 2/ leur organisation et stockage à court et long termes, les traitements et analyses effectués ; 3/ leur documentation pour les rendre intelligibles ; 4/ en vue d'un archivage visant si possible le partage de celles-ci ainsi qu'une réutilisation.

2 <https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-on> « Le RGPD encadre le traitement des données personnelles sur le territoire de l'Union européenne. Le contexte juridique s'adapte pour suivre les évolutions des technologies et de nos sociétés (usages accrus du numérique, développement du commerce en ligne...). Ce nouveau règlement européen s'inscrit dans la continuité de la Loi française Informatique et Libertés de 1978 et renforce le contrôle par les citoyens de l'utilisation qui peut être faite des données les concernant. Il harmonise les règles en Europe en offrant un cadre juridique unique aux professionnels. Il permet de développer leurs activités numériques au sein de l'UE en se fondant sur la confiance des utilisateurs. » (<https://doranum.fr/> s.p.).

3 Liste disponible sur La plateforme DoRANum, conçue pour offrir aux chercheurs des ressources pour la gestion et le partage des données : <https://doranum.fr/>

4 Modèles-types : <https://ldi.hypotheses.org>

La plateforme DoRANum (cf. *supra*) propose notamment des tutoriels concernant l'élaboration de PGD, ainsi qu'un guide (téléchargeable) intitulé *Comment choisir un modèle de PGD ?*

Étape 2 — Choix d'un entrepôt

Les données décrites dans un *Data paper* doivent être accessibles « soit sous forme de fichiers annexés, soit plus généralement par un lien pérenne (URL, DOI) vers "l'entrepôt de données" en ligne (*data repository*, ou *repository of research data*) où elles sont déposées et correctement formatées », explique Dedieu (2014, p. 1).

Un entrepôt est un espace d'archivage des données sécurisé et fiable. Pour choisir l'entrepôt le plus adapté aux données, il existe la base de données des entrepôts Cat OPIDoR⁵. Cette base référence 213 entrepôts et leurs principales caractéristiques : domaine scientifique, thématique/mots-clés, structure de rattachement, type de données, conditions d'usage (accès libre ou restreint, type de licence...). En Sciences humaines et sociales, si l'on sélectionne les deux sous-catégories « SH3 Le monde social et sa diversité » et « SH5 Cultures et production culturelle », Cat OPIDoR recense douze entrepôts ; il n'en repère que cinq dans la seule sous-catégorie « SH5 Cultures et production culturelle ». Les entrepôts sont tous rattachés à une structure : par exemple, BeQuali est rattaché au Centre de Données Socio-Politiques, Cybergeog Dataverse à l'UMR Géographie-Cités ou encore Nakala à l'Infrastructure de Recherche des Humanités Numériques Huma-Num⁶. La plupart de ces entrepôts sont gratuits. Il existe des entrepôts privés et des entrepôts publics. Les entrepôts institutionnels sont de plus en plus nombreux, mais ils ne sont pas tous certifiés. La certification (ex. ISO 27001, IBSA, CoreTrustSeal) est vivement recommandée. C'est un critère de qualité qui « atteste qu'un entrepôt est "de confiance" pour les personnes qui utilisent les données qu'il met à disposition, mais aussi pour ceux [les chercheurs] dont il gère les données »⁷.

Étape 3 — Choix d'une revue

La *Data paper* peut être abrité dans deux types de revues : les revues classiques qui disposent d'une rubrique dédiée au *Data paper*, les revues dont le contenu est exclusivement dirigé vers la publication de *Data papers*. Certaines sont disciplinaires, d'autres multidisciplinaires. La manière d'organiser et de présenter les données et métadonnées est plus ou moins liée au choix de la revue. Certaines revues ont un facteur d'impact (IF) d'autres non, certaines sont en libre accès quand d'autres ont un accès payant, certaines ont un coût de publication quand d'autres n'en ont pas.

Étape 4 — Choix d'une licence de diffusion

Le choix de la licence de diffusion des données est une étape cruciale puisque celle-ci fixe les conditions de l'usage des données. Elle encadre les « droits d'utilisation et de modification de la donnée, [les] droits de réutilisation commerciale et non commerciale, [les] obligations éventuelles comme la mention de la source des données ou le partage à l'identique » (Cirad)⁸.

Les licences Creative Commons sont les plus répandues mais il en existe d'autres⁹. En cas de doute sur le choix de la licence à appliquer aux données, il est recommandé de demander conseil auprès du

5 https://cat.opidor.fr/index.php/Entrep%C3%B4t_de_donn%C3%A9es

6 <https://www.huma-num.fr/>

7 <https://www.ouvrirlascience.fr/certification-des-entrepots-et-services-de-donnees/>

8 <https://coop-ist.cirad.fr/gerer-des-donnees/ouvrir-ses-jeux-de-donnees-scientifiques/6-choisir-une-licence-de-diffusion-pour-la-reutilisation-des-donnees>

9 Liste accessible sur le site du Cirad : <https://coop-ist.cirad.fr/etre-auteur/utiliser-les-licences-creative-commons/1-droits-d-auteurs-et-droit-des-bases-de-donnees>

service juridique de l'institution impliquée. La licence de diffusion concerne à la fois le *Data paper* en lui-même et les données présentes dans l'entrepôt. La licence de diffusion appliquée aux *Data papers* dépend de la revue. Ainsi, certaines revues peuvent appliquer la licence Creative Commons CC-BY (obligation de créditer les auteurs), tout comme la licence CC0, sans obligation de créditer les producteurs de données (Dedieu, 2022).

Étape 5 — Préparation des données, réalisation des métadonnées

Qu'est-ce qu'une donnée ?

Une fois l'entrepôt choisi, il convient de s'assurer de ce que ses concepteurs entendent par *donnée*. Plusieurs conceptions sont possibles, selon les entrepôts, mais aussi selon le choix des équipes de recherche :

- un à n fichiers numériques de formats différents, liés par une même thématique,
- et/ou un groupement de fichier.s de mêmes formats, liés par une même thématique,
- et/ou un à n fichier.s liés provenant d'un même auteur de données.

D'une manière générale une donnée de un à n fichiers est complétée par une notice descriptive, les métadonnées.

Une donnée (ou un jeu de données, cf. *infra*) déposée dans un entrepôt est automatiquement dotée d'un DOI¹⁰.

Définir un jeu de données

Dans le même ordre d'idées, il convient de définir précisément ce qui est entendu par « jeu de données » ou encore « collection de données », en lien étroit avec le vocabulaire de l'entrepôt.

L'arborescence, dans l'espace de stockage initial des données, est à concevoir en fonction de l'entrepôt choisi, en lien avec la cohérence de ce qui peut, ou non, être qualifié de donnée.

Le site recherche.data.gouv parle d'un jeu de données comme d'un « ensemble cohérent de données brutes ou dérivées, auquel sont associées des métadonnées qui permettent de décrire ces données (présentation, date de publication, identifiant pérenne, mots-clefs, couverture géographique, couverture temporelle, etc.) »¹¹

Partage des droits

Si on observe les pratiques de l'entrepôt Nakala, il convient de faire, en amont du dépôt des données, un partage des droits entre les auteurs et/ou collecteurs des données. Une liste d'utilisateurs doit être créée pour attribuer les différents rôles. Chaque rôle correspond à un ensemble de droits prédéfinis. Le gestionnaire peut attribuer les droits à d'autres personnes. Il peut, le cas échéant, accorder le rôle de gestionnaire à quelqu'un d'autre, auquel cas il n'aura plus la main sur le partage des droits. Le déposant (souvent également gestionnaire) peut déposer les données dans l'entrepôt. L'administrateur peut modifier ou supprimer la liste des rôles. Le simple membre, ou lecteur, n'a qu'un droit de consultation.

10 Le DOI (Digital Object Identifier, identifiant numérique d'objet) est un identifiant unique permettant d'identifier, de référencer et de citer différentes ressources par un lien durable.

11 <https://recherche.data.gouv.fr/fr>

Plan de nommage et abréviations

Dans le prolongement de ce travail, il faut impérativement convenir d'un plan de nommage des fichiers. Pour établir cette nomenclature il est possible, par exemple, de recourir au *Guide des bonnes pratiques – Gestion et valorisation des données de recherche* (OTELo-CNRS, 2017, p. 6). Ses recommandations sont les suivantes :

- « commencer par une lettre ;
- maximum 30 caractères¹² ;
- pas de caractères spéciaux, d'accents, d'espaces ou de points ;
- utilisation des lettres de a à z (majuscules et minuscules), les chiffres (0-9), le tiret et le tiret bas (souligné), l'acronyme du projet ;
- la date de création sous la forme yyyyymmdd ou yyyy-mm-dd ;
- les initiales du nom du créateur de la donnée ;
- une description brève du contenu (ex. établissement où s'est fait la collecte, nature du document, niveau de classe concerné) ;
- le numéro de version si nécessaire ;
- le format du fichier. »

Par exemple, le verbatim d'un entretien avec le directeur du Conservatoire à Rayonnement Régional de Toulouse est renommé ainsi : EMIS_2019-08-08_VERB_dir_CRR-Tlse_LA_docx ; soit : nom du projet de recherche_date_fonction de l'interviewé_lieu d'exercice_initiales auteur de la donnée_format. Nous avons ajouté, quand nous en avons eu l'autorisation, les initiales de l'auteur de la donnée (auteur au sens de celui qui construit la donnée pour les données secondaires).

Il peut être nécessaire de réaliser un tableau des correspondances entre le nommage initial et le nouveau nommage.

Il est indispensable d'établir une liste des abréviations utilisées dans le plan de nommage, et de la rendre disponible, dans le readme (cf. *infra*), pour les futurs chercheurs qui souhaiteraient travailler à partir de ce corpus.

Les métadonnées

Les abréviations utilisées dans le plan de nommage serviront également pour l'élaboration des métadonnées. Les métadonnées visent à décrire précisément les données, dans un vocabulaire idoine, de préférence le vocabulaire contrôlé de la discipline. D'une manière générale, « les métadonnées détaillent pourquoi, par qui et comment ces données ont été collectées, qui en est propriétaire, sous quel format elles sont stockées, etc. » (s.p.), explique Dedieu (2014). Selon les entrepôts, le nombre de métadonnées obligatoires par donnée peut varier. Plus une donnée est finement décrite (et donc, dépasse le nombre de métadonnées obligatoires), plus elle sera réutilisable facilement.

Les métadonnées sont élaborées selon un schéma spécifique (ou format, ou standard) imposé par l'entrepôt et/ou par la revue. Ce standard est soit spécialisé pour un type de données, soit généraliste et ouvert à toutes sortes de données. Par exemple, pour les SHS, tel entrepôt peut fonctionner avec le schéma DDI (Data Documentation Initiative, 1995) spécialisé dans les données d'enquêtes et

12 « Afin d'éviter d'éventuels problèmes de transcription du nom en cas de sauvegarde automatique sur un serveur ou de changement de système d'exploitation informatique » (OTELo-CNRS, 2017, p. 6).

notamment les données statistiques, tandis que d'autres entrepôts recourent au schéma généraliste et interdisciplinaire, Dublin Core (DC, 1995) ou Dublin Core Qualifié (depuis 2000).

L'anonymisation des données

Certaines données comportent des informations à caractère personnel (ex. nom, prénom, moyennes scolaires dans certaines matières...). Même si la personne a signé une demande d'autorisation d'enregistrement et de cession de l'image et de la voix en bonne et due forme, le chercheur se doit de rendre anonyme la ou les donnée(s) s'il souhaite les rendre publique.

De plus, les données qui comportent des informations sensibles ne peuvent être rendues publiques sans que la ou les personne(s) concernée(s) aient donné son ou leur consentement expressément par écrit. Une donnée sensible est une « donnée à caractère personnel qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que les données génétiques, les données biométriques aux fins d'identifier une personne physique de manière unique, les données concernant la santé ou les données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique » (Cirad, Couperin.org, Loginos¹³, s.p.).

Le readme

Le readme fournit des informations complémentaires et accessibles qui ne sont pas nécessairement mises à disposition dans les métadonnées, comme la description des sources et méthodes utilisées pour collecter et générer les données, ou encore les procédés d'anonymisation ou de pseudonymisation¹⁴. Il explique aussi la charte de nommage et l'arborescence ou le plan de classement des fichiers. Il est fortement recommandé d'associer un readme aux données sur l'entrepôt. Celui-ci est réalisé pour un ensemble de données.

Rédiger le *Data paper* : une entreprise très exigeante

Les paragraphes qui suivent illustrent par des exemples empiriques, les différentes notions abordées dans les parties précédentes, à partir d'une expérience : le projet MusiPim, qui a débuté en septembre 2023, porté par le LabEx SMS (Université Toulouse-Jean Jaurès). Ce projet réunit deux corpus existants auxquels s'ajoute un recueil de données complémentaires. MusiPim, en effet, « vise à repérer et à catégoriser au sein de ces corpus augmentés, grâce à une nouvelle enquête par entretiens, les tensions, dilemmes et conflits inhérents à la situation partenariale dans les projets OAE [corpus Orchestre à l'école collecté par le groupe de recherche collaborative EMIS, Éducation Musicale et Intégration Sociale, entre 2017 et 2021] et Démos [corpus de la recherche-action sur Démos Clermont-Ferrand entre 2019 et 2021] » (Tripier-Mondancin et al. 2024, en cours).

Une volonté de partager assez récente

En amont de MusiPim, le groupe de recherche collaborative porteur de l'étude sur l'Orchestre à l'école (OAE) à Toulouse¹⁵ s'était posé, dès sa création en 2017, la question du codage des données à l'aide de métadonnées, dans l'idée d'aller vers la science partagée. Cette préoccupation a toutefois été différée au profit de la publication de plusieurs articles scientifiques. La volonté de rédiger un *Data paper* et si

13 <https://www.loginos.net/base/ylxxVO>

14 D'après l'*Avis 05/2014 sur les techniques d'anonymisation*, « pour rendre des données anonymes il faut en retirer suffisamment d'éléments pour que la personne concernée ne puisse plus être identifiée » (2014, p. 6). Les données pseudonymisées ne sont pas équivalentes aux données anonymisées : « les données pseudonymisées ne peuvent être assimilées à des informations anonymisées puisqu'elles continuent à permettre l'individualisation d'une personne concernée et la corrélation entre différents ensembles de données. Le pseudonymat n'est pas de nature à empêcher qu'une personne concernée soit identifiable et reste donc dans le champ d'application du régime juridique de la protection des données » (Avis, 2014, p. 11).

15 <https://ldi.hypotheses.org/227>

possible un article sur la méthodologie d'élaboration d'un *Data paper*, remonte à la fin du projet EMIS. Tripiier-Mondancin (2022) écrivait alors : « Même si les données ne sont pas [encore] partageables, la méthodologie en vue de l'écriture d'un *Data paper* serait instructive pour le groupe ou tout chercheur engagé ; elle contribuerait à solidifier la méthodologie des recherches engagées. L'élaboration d'un plan de gestion des données participerait également à clarifier les méthodologies construites dans les travaux menés » (p. 330-331).

La collecte en 2024 de nouvelles données complémentaires aux précédentes, dans le cadre du projet MusiPim, est l'occasion de réfléchir à nouveau à la méthodologie d'un *Data paper*, tout en mettant en œuvre parallèlement son écriture.

Passage à l'acte et difficultés repérées

Mise en conformité rétroactive avec les deux délégués des deux universités (étape 0)

Cette étape de mise en conformité a été entièrement revue en 2024 à l'aune des préconisations des Délégués à la Protection des Données ou DPO des deux universités responsables des deux projets (Toulouse et Clermont-Ferrand). En effet, la mise en conformité des données déjà collectées ainsi que le traitement des futures données de la nouvelle enquête sont des étapes complexes. Le DPO de l'Université Toulouse-Jean Jaurès et ses associées ont répondu rapidement à nos sollicitations. La prise de contact avec le DPO de l'Université Clermont Auvergne a également été rapide et fructueuse.

Pour le premier corpus (Orchestre à l'École à Toulouse), si des « autorisations de droits à l'image » avaient été signées par les parents, le principal du collège, les enseignants d'éducation musicale et de chant choral, en revanche, les droits de cession pour publication en ligne ne l'avaient pas été.

Pour le deuxième corpus (Demos à Clermont-Ferrand), l'autorisation de filmer n'avait été demandée qu'oralement lors des séminaires et les autorisations de droit à l'image concernant les enfants filmés n'avaient pas été prévues pour être valides au-delà des trois années du projet Demos. Dans les deux cas, Toulouse et Clermont-Ferrand, aucune « Notice d'information et formulaire de consentement éclairé » n'avait été prévue pour les enquêtés. Ainsi, pour les deux corpus, les formulaires 1) de consentement éclairé, et 2) de droit à l'enregistrement et à la cession de l'image et de la voix, ont rétroactivement été distribués aux enquêtés qui ont eu la possibilité de les signer ou non. À l'heure où nous écrivons ces lignes, environ 50 % de réponses ont été recueillies.

Enfin, pour le projet MusiPim (3^e corpus), les démarches pour les signatures des deux formulaires ont été réalisées en parallèle de la nouvelle enquête.

Il est à noter que les démarches faites rétroactivement, une fois le recueil des données effectuées, sont très chronophages et incertaines : le chercheur n'est ni assuré de retrouver les contacts des enquêtés, ni d'obtenir des réponses favorables au partage. En effet, il peut être nécessaire de solliciter à plusieurs reprises les personnes, ou encore de tenter de convaincre les personnes réticentes. En outre, lorsque le nombre de personnes impliquées est grand (ce qui est le cas, par exemple, lors des répétitions d'orchestre, dans les deux projets), il est difficile de faire circuler les formulaires à la totalité des enquêtés dans les temps impartis. Le floutage des visages et la suppression des passages audio concernés sont alors des solutions envisageables. Le processus est en cours à l'heure où nous écrivons ces lignes. Mais ce travail est à nouveau particulièrement chronophage. S'il ne peut être accompli en bonne et due forme, il faut alors renoncer à publier ces données audio-visuelles.

La mise en conformité RGPD des entretiens audios est plus facilement envisageable dans la mesure où un nombre restreint d'individus est concerné. Il a donc été possible, rétroactivement, de leur faire

parvenir un formulaire de consentement éclairé ainsi qu'une demande d'autorisation de captation et de cession pour publication rétroactive.

Concernant les verbatims des entretiens, ils ont été systématiquement anonymisés.

En outre, une déclaration de traitement des données a dû être faite rétroactivement pour les corpus toulousain et clermontois. Pour le corpus MusiPim, elle a été réalisée au démarrage de l'enquête.

Gestion des données prévue dès le démarrage (étape 1)

Concernant l'écriture et l'actualisation du Plan de Gestion des Données (étape 1), elles ont été prévues dès l'écriture du projet MusiPim (corpus 3) soumis au LabEx SMS. La version initiale a été saisie sur la plateforme DoRANum¹⁶ à l'aide de *DMP OPIDoR* ; il est fondé sur le modèle « *Science Europe : modèle structuré* » fourni par Science Europe.

Ci-dessous (figure 1), un extrait de la version initiale du PGD de la recherche MusiPim :

DMP du projet "Le partenariat inter-métiers au sein de deux dispositifs de pratique musicale collective à vocation sociale, Orchestre à l'école et Démon."

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "Science Europe : modèle structuré" fourni par Science Europe.

Renseignements sur le plan

Titre du plan DMP du projet "Le partenariat inter-métiers au sein de deux dispositifs de pratique musicale collective à vocation sociale, Orchestre à l'école et Démon."
Livrable Livrable PGD version 1
Version Version initiale

Objet/périmètre du plan

Les travaux de MusiPim seront fondés sur l'exploitation des données produites en grande partie par deux enquêtes. La première a été menée par le groupe de recherche collaborative EMIS dans le cadre de la classe orchestre du collège Bellefontaine, à Toulouse. L'ensemble des données est déjà archivé sur la plateforme Huma-Num, mais n'a pas fait l'objet d'élaboration des métadonnées, faute de temps. Cette enquête a donné jour à deux articles publiés (Roubertie Soliman et al., 2019 ; Abello et al., 2022). Le deuxième article fait partie du numéro 47 des *Dossiers des Sciences de l'Éducation* co-dirigé par deux des fondateurs d'EMIS (Canguilhem et Tripiet-Mondancin, 2022). La deuxième enquête a été menée par L. Roubertie Soliman au sein de l'orchestre Démon de Clermont-Ferrand autour de la question des modalités de coordination inter-professionnelles (Roubertie Soliman et al., 2021). Ces dernières données ne sont pas archivées sur Huma-Num. Un premier travail consistera à le faire et à élaborer les métadonnées. Le projet MusiPim reposera donc sur ce double corpus, complété par de nouveaux entretiens qui feront l'objet d'un traitement similaire : organisation, archivage, élaboration des métadonnées.

Dès lors, MusiPim propose la création et la tenue d'un Plan initial de Gestion des Données (PGD), dans l'idée d'assurer, si possible, un bon partage de celles-ci entre chercheur.e.s et acteurs.trices. MusiPim s'engage à ce que ces données « obéissent au principe FAIR¹¹ » dans le cadre de la science ouverte (CNRS, 2019, p. 5). Le PGD évoluera au cours de ces 12 mois, intégrant à la fois les anciennes données et les nouvelles. Deux versions seront livrées au laboratoire SMS (6 mois, 12 mois). Ces deux versions récapituleront : 1/ l'ensemble des données collectées dans les deux enquêtes ainsi que les nouvelles, la manière dont elles ont été produites et collectées ; 2/ leur organisation et stockage à court et long termes, les traitements et analyses effectués ; 3/ leur documentation pour les rendre intelligibles ; des métadonnées viseront à décrire précisément ces données ; 4/ en vue d'un archivage visant si possible le partage de celles-ci, ainsi qu'une réutilisation, comme ce sera le cas avec MusiPim. Ce projet est donc l'occasion de réfléchir au cycle des données en développant une culture de la constitution, de la gestion, de la collecte, de l'archivage et si possible du partage de données fiables, dans un cadre éthique et responsable.

¹¹ FAIR : « Faciles à trouver, accessibles, interopérables et réutilisables » (CNRS, 2019, p. 5).

Domaines de recherche (selon classification de l'OCDE)

Art (arts, history of arts, performing arts, music), Educational sciences, Political science, Social and economic geography

Langue fra

Date de création 2023-01-29

Date de dernière modification 2023-02-01

Documents (publications, rapports, brevets, plan expérimental....), sites web associés

- Roubertie Soliman L., Rix-Lièvre G. et Wierre-Gore G. (2021). L'orchestre Démon Clermont-Ferrand, l'œuvre d'un travail interprofessionnel : Que pouvons-nous apprendre des modalités de coordination entre les acteurs ? Rapport d'étude, ACTÉ, Université Clermont Auvergne. : https://demos.philharmoniedeparis.fr/evaluation.aspx?_lg=fr-FR
- Roubertie Soliman L., Tripiet-Mondancin O., Martinez E. et Bach N. (2019). Projet orchestre dans un collège REP+, à Toulouse : caractériser les difficultés déclarées par des professeurs chargés de l'enseignement musical, Education et socialisation - Les cahiers du CERFEE, Presses Universitaires de la Méditerranée : <https://doi.org/10.4000/edso.8090>
- Abello L., Tripiet-Mondancin, O., Roubertie Soliman, L., Guy, D. (2022). L'impact des pratiques musicales collectives à vocation sociale en milieu scolaire. Le cas de deux projets en éducation prioritaire. Les Dossiers des Sciences de l'Éducation n°47, 11-32. : https://www.researchgate.net/publication/366864075_L'impact_des_pratiques_musicales_collectives_a_vocation_sociale_en_milieu_scolaire-version_auteur
- Canguilhem, P. et Tripiet-Mondancin, O. (dir.). (2022). Pratiques musicales collectives à vocation sociale : un état de la recherche. (Vol. 47). Toulouse : PUM. : https://www.researchgate.net/publication/366864168_Pratiques_musicales_collectives_a_vocation_sociale_un_etat_de_la_recherche_Version_auteur
- L'CEIL DU LINQS Les carnets du Laboratoire des Idées – (Nouvelles) Questions SocialesCatégorie : EMIS Publié le 28/11/2017 Pour une recherche collaborative : observer des pratiques musicales collectives en REP+ à Toulouse : <https://ldi.hypotheses.org/category/rencontres>

Figure 1 : Extrait de la version initiale du PGD ou DMP du projet MusiPim (capture écran)

16 <https://doranum.fr/>

Un entrepôt en SHS, fiable et gratuit (étape 2)

Lorsque l'on choisit un entrepôt, il est intéressant de s'assurer que l'on peut bénéficier de conseils, de soutien, d'assistance technique, de la part de l'équipe qui l'administre. Par ailleurs, il est important de se renseigner sur les fonctionnalités de l'entrepôt : facilité du dépôt, de l'authentification, gestion, manipulation des données (versionnage, suppressions...), préservation des données, stockage, sécurité. La dimension nationale ou internationale de l'entrepôt est un autre point à prendre en considération, selon la portée estimée des données.

Différents exemples d'entrepôts (français et internationaux) ont été comparés (fonctionnement, points forts et points faibles). Diverses questions se sont alors posées. Y répondre nous a permis de décider de téléverser, lorsqu'elles seraient prêtes, les données déjà archivées sur Huma-Num¹⁷ vers l'entrepôt Nakala¹⁸ qui lui est associé (tableau 2).

Questions	Informations	Entrepôt Nakala	Adéquation de l'entrepôt Nakala au projet MusiPim
Langue de publication des métadonnées ?	Les données collectées à Toulouse comme celles collectées à Clermont-Ferrand sont destinées en premier lieu à un lectorat francophone.	Français, avec possibilité de rédiger les métadonnées dans d'autres langues, en différé si nécessaire.	Nakala est adapté au projet MusiPim
Standard de métadonnées ?	Dublin Core est un format ouvert et peu spécialisé, adapté aux données et métadonnées ; Data Documentation Initiative (DDI) est spécialisé SHS mais réservé aux données d'enquête (pas adapté à d'autres formes de données ni aux métadonnées).	Dublin Core qualifié	Nakala semble plus adapté (d'autant que rien n'empêche d'aller vers d'autres standards par la suite et de les relier).
Pérennité de la conservation des données ?	--	Illimitée	Nakala est adapté
Coût éventuel ?	Certains entrepôts sont payants.	Gratuit	Nakala est adapté
Certification de l'entrepôt ?	Une grande partie des entrepôts ne sont pas certifiés ou en cours de certification.	Demande en cours (Core Trust Seal)	Nakala est adapté

Tableau 2 : Questions, informations et évaluation de l'adéquation de l'entrepôt Nakala au projet MusiPim

Nakala est gratuit, national et public. En outre, cet entrepôt a effectué une demande de certification CoreTrustSeal¹⁹ qui est en cours de traitement. Par ailleurs, les auteurs ont pu bénéficier de plusieurs jours de formations à la maîtrise des techniques de dépôt, dont la procédure de téléversement et les choix possibles des formats des fichiers. Ce type de formation ne peut que contribuer à la réussite du projet de partage.

Le repérage des revues spécialisées en *Data paper* versus non spécialisées (étape 3)

Dans un premier temps, un repérage des revues acceptant les *Data papers* a été réalisé, notamment à l'aide de l'article de Dedieu, *Revue publiant des Data papers* (Cirad, 2017, <http://coop-ist.cirad.fr/>), et de sa conférence en ligne « Publier un *Data paper* : enjeux de bonnes pratiques, d'intégrité scientifique et de science ouverte » (Dedieu, 2022). Ce repérage a permis d'établir une sélection de

17 <https://www.huma-num.fr/>

18 <https://www.nakala.fr/>

19 « Core TrustSeal est une organisation internationale, communautaire, non gouvernementale et à but non lucratif qui promeut des infrastructures de données durables et fiables. » <https://www.coretrustseal.org/about/>

revues spécialisées (*data journals*) et non spécialisées mais acceptant les *Data papers*. Trois revues ont été retenues pour leur adéquation possible avec le projet : *Data in Brief* (Elsevier), multidisciplinaire, *Research Data Journal for Humanities and Social Sciences* (Brill) et *Rio Journal* (Research and Ideas Outcomes : <http://www.pensoft.net/about.php>).

Le colloque « Publier, partager, réutiliser les données de la recherche : les *Data papers* et leurs enjeux », organisé par la MESH de Lille (8 et 9 novembre 2021) suivi dans le cadre de la réflexion sur le cycle des données par le groupe de recherche collaborative EMIS sur les OAE à Toulouse, a permis de repérer une quatrième revue susceptible d'accueillir notre futur *Data paper*, le *Journal of Open Humanities Data* (JOHD) (tableau 2).

Dans un second temps de ce repérage, il nous a semblé pertinent de faire apparaître le *Data paper* dans une revue en sciences de l'éducation pour atteindre un public plus ciblé. Notre enquête nous a amené à observer une absence de ce type de publication dans les revues francophones (c'est le même constat concernant les revues en musicologie). Au plan international, les *Data papers* se situent du côté des sciences dites dures, ils sont encore peu présents en SHS. Citons, parmi les revues qui publient exclusivement des *Data papers* en SHS, le *Journal of Open Humanities Data* et le *Research Data Journal for the Humanities and Social Sciences*. En sociologie, *Frontiers in Sociology* publie, en plus des articles traditionnels, des *Data papers* dans la section des articles courts (sans qu'il existe une rubrique spécifique dédiée aux *Data papers*). Les revues spécialisées en éducation musicale (par exemple, le *British Journal of Music Education*, *l'International Journal of Music Education*, le *Bulletin of the Council for Research in Music Education*) ne proposent pas ce type de papier. L'offre est assez réduite, ce qui implique de s'adresser à un lectorat plus large que son propre champ de recherche. Dans notre investigation nous avons sollicité plusieurs revues francophones en sciences de l'éducation (nos jeux de données étant en langue française et étant susceptibles d'intéresser avant tout des chercheurs se situant dans le champ éducatif). Certaines, comme *Recherches en didactiques*, acceptent d'accueillir ce type de production, d'autres y réfléchissent actuellement au sein de leur comité éditorial. Ajoutons que la *Revue française des sciences de l'information et de la communication* propose une rubrique dédiée aux *Data papers*, et qu'un groupe de travail s'est constitué à la suite d'une journée d'études à la MSH Lorraine afin de créer, en France, un data journal SHS qui sera publié en ligne, selon un calendrier amenant à une ouverture fin 2025. Les tableaux 2 et 3 résument les principales caractéristiques des 9 revues envisagées.

Nom de la revue, éditeur (Dedieu, 2017)	Data journal	Pluri-disciplinaire	Date de création de la revue	Impact Factor	Coût
<i>Data in Brief</i> (Elsevier)	Oui + articles de synthèse et de perspective sur les pratiques de partage des données	Oui (sciences dures et SHS)	2014	IF 1.2	Frais de publication : 840 \$
<i>Journal of Open Humanities Data</i> (JOHD)	Non	SHS	2015	IF 0.84	Frais de publication : 485 £
<i>Research and Ideas Outcomes Journal</i> (Pensoft)	Non	Oui, mais aucun DP publié en SHS	2015	--	Gratuit pour les manuscrits en ARPHA Writing Tool, sinon payant
<i>Research Data Journal for the Humanities and Social Sciences</i> (Brill)	Oui	SHS	2016	IF 0.143 [Dedieu indique « pas d'IF », 2022]	Gratuit

Tableau 3 : Repérage de revues anglophones publiant des *Data papers* en SHS (d'après Dedieu 2017, 2022)

Nom de la revue, éditeur	Data journal	Discipline	Date de création de la revue	Impact Factor	Coût	Réponse de la revue
<i>Bulletin de psychologie</i>	Publie des articles scientifiques et des <i>Data papers</i>	Psychologie	1948	--	gratuit	Intéressée par les <i>Data papers</i>
<i>Revue Française de Pédagogie</i> (ENS Éditions)	N'a jamais publié de <i>Data paper</i>	Sciences de l'éducation et de la formation	1967	IF 0.11	gratuit	Pas intéressée par les <i>Data papers</i>
<i>Recherches en didactiques – Les Cahiers de Théodile</i> (Presses Universitaires du Septentrion)	N'a jamais publié de <i>Data paper</i>	Sciences de l'éducation et de la formation	2000	--	gratuit	Intéressée par un <i>Data paper</i>
<i>Recherches en éducation</i> (Nantes Université)	N'a jamais publié de <i>Data paper</i>	Sciences de l'éducation et de la formation	2006	--	gratuit	En cours de discussion au sein du comité éditorial
<i>Éducation & Didactique</i> (Presses Universitaires de Rennes)	N'a jamais publié de <i>Data paper</i>	Sciences de l'éducation et de la formation	2007	--	gratuit	En cours de discussion au sein du comité éditorial
<i>La pensée d'Ailleurs</i> (Pareo)	N'a jamais publié de <i>Data paper</i>	Sciences de l'éducation et de la formation	2019	--	gratuit	En cours de discussion au sein du comité éditorial
<i>Data Journal SHS</i> (projet Université de Lorraine)		SHS	En cours de création	--		Première publication envisagée fin 2025, soit hors délais pour MusiPim

Tableau 4 : Repérage de revues francophones en SHS, sollicitées par MusiPim

Le choix de la revue s'est finalement fait en fonction du domaine scientifique, de la langue du lectorat potentiel, des modalités d'accès (libre accès) et du coût de publication qui peut aller de la gratuité jusqu'à environ 3000 €. Les auteurs du présent article ont privilégié les revues dans les sciences humaines, en langue française (puisque les projets musicaux sont mis en œuvre en France) et la gratuité. Aucune des revues intéressées n'était dotée d'un facteur d'impact (Impact Factor).

Une licence ouverte qui impose la citation (étape 4)

En ce qui concerne le choix d'une licence de diffusion, nous avons opté pour la CC-BY-NC, c'est-à-dire la « Licence Creative Commons Attribution — Pas d'Utilisation Commerciale — Partage dans les Mêmes Conditions 4.0 International ». Selon les bonnes pratiques scientifiques, le crédit approprié sera mentionné dans le *Data paper*²⁰, par le biais d'une citation.

Rappelons qu'il faut se préoccuper non seulement de la licence qui concerne l'article à publier, mais aussi de celle qui concerne les données de l'entrepôt. Selon la revue, la licence de diffusion des données du *Data paper* est imposée ou non ; le choix des licences est plus ou moins ouvert.

Une préparation exigeante et chronophage (étape 5)

En ce qui concerne la préparation des données, nous nous sommes alignés sur les préconisations de l'entrepôt Nakala. Ainsi, une donnée regroupe des fichiers de même nature selon les catégories

20 <https://coop-ist.cirad.fr/gerer-des-donnees/publier-un-data-paper/1-qu-est-ce-qu-un-data-paper>

proposées par Nakala, par exemple : texte, audio, vidéo, poster, article de journal, cours, partition, présentation.

Il a été décidé qu'un jeu de données correspond à toutes les données recueillies sur un projet. Nous avons donc trois jeux de données : OAE Toulouse, Démonstrations Clermont-Ferrand et MusiPim.

Le partage des droits a été décidé entre les trois auteurs du présent article, selon trois rôles principaux : gestionnaire (un seul ; il peut attribuer le rôle de gestionnaire à un autre utilisateur, modifier et supprimer la liste, déposer des données), administrateur (il peut modifier et supprimer la liste, déposer des données), éditeur (il peut déposer des données) et simple lecteur (simple consultation des données).

Le plan de nommage a été évoqué ci-dessus (cf. *supra*).

Quant aux métadonnées, le format Dublin Core qualifié a été adopté car c'est celui préconisé par l'entrepôt Nakala qui a été choisi. (cf. tableau 2).

La figure 2 (cf. *infra*) résume sous forme d'une flèche du temps les différentes étapes qui ont participé à l'écriture du *Data paper*.

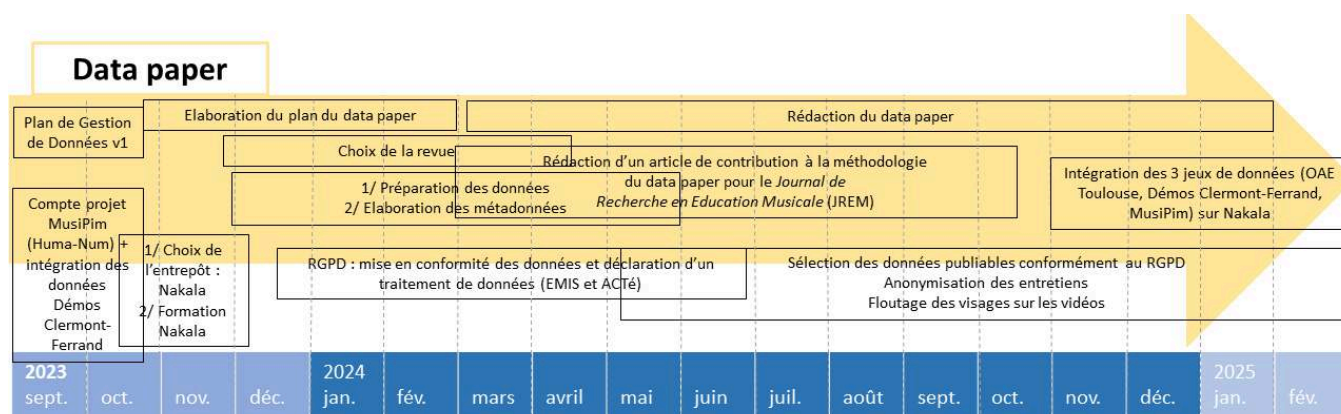


Figure 2 : Principales étapes de préparation du Data paper (projet MusiPim)

Conclusion

La relative nouveauté des principes de la Science ouverte appliqués aux SHS, en particulier dans le champ des sciences de l'éducation et de la formation musicale, et la rareté des *Data papers* dans ce domaine, nous ont amenés à un travail d'adaptation constant sur les méthodes de collecte, d'archivage, de classement. Peu à peu l'idée s'est imposée de tout mettre en œuvre pour publier un *Data paper*. L'objectif de cet article était de rendre compte de ce qui fait la spécificité d'un article de type *Data paper* et de partager notre expérience, dans la perspective souhaitée de faire gagner du temps aux chercheurs nouvellement intéressés par ces questions.

Le potentiel de réutilisation des données, dans le cadre de nouveaux travaux, nous semble être au cœur d'un questionnement sur le partage des données, tout comme sur les enjeux écologiques et éthiques de production des données de la recherche.

L'expérience a montré que dans le cas où la volonté d'un groupe de chercheurs est d'aller vers une mise en ligne publique des données, un temps considérable est gagné lorsque les démarches nécessaires à la publication des données sont faites au démarrage du projet de recherche et que les travaux s'inscrivent dans le respect du RGPD (Règlement Général sur la Protection des Données), si la

collecte et le traitement des données sont réalisés dans le respect de la loi de 2018. Les demandes d'autorisation et les formulaires de consentement doivent être préparés le plus tôt possible. En outre, la déclaration des traitements de données doit être réalisée auprès des DPO en préliminaire à toute enquête.

Références bibliographiques

- Arnoud, P.-Y. et Jacquemot-Perbal, M.-C. (2017). *Guide de bonnes pratiques – Gestion et valorisation des données de recherche*, OTELo-CNRS. <https://hal.science/hal-01275841>
- Charles, L.A. et Acquier, F. (2022, 22 juillet). Retour d'expériences de publication de données : l'entrepôt de données SHS Nakala. *Lab & doc*. <https://labeledoc.hypotheses.org/?p=10424>
- CINES (2017). Les formats de fichier. <https://www.cines.fr/archivage/des-expertises/les-formats-de-fichier/>
- CNIL (2020). L'anonymisation des données personnelles. <https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles>
- CNRS (2021). Guide de bonnes pratiques sur la gestion des données de recherche. Publier un Datapaper pour valoriser et expliciter les données. <https://mi-gt-donnees.pages.math.unistra.fr/guide/07-publier.html>
- CNRS (2020). Groupe de travail *Atelier Données* : Foire aux questions. <https://mi-gt-donnees.pages.math.unistra.fr/site/FAQ.html>
- Core TrustSeal. <https://www.coretrustseal.org/about/>
- Dedieu, L. (2014). Rédiger et publier un *Data paper* dans une revue scientifique en 5 points. CIRAD, 7 p. <https://collaboratif.cirad.fr/alfresco/s/d/workspace/SpacesStore/75735aae-dacb-4052-833c-9017a2b2bba4/rediger-et-publier-un-data-paper-octobre2014.pdf>
- Dedieu, L. (2022). Rédiger et publier un *Data paper* en 5 points. CIRAD, 5 p. <http://doi.org/10.18167/coopist/0057>
- Dedieu, L. (2017). Revues publiant des *Data papers*. CIRAD, 23 p. <http://coop-ist.cirad.fr/>
- Dedieu, L. et Barale, M. (2020). Déposer des données dans un entrepôt, en 6 points. CIRAD, 4 p. <http://doi.org/10.18167/coopist/0070>
- Dedieu, L. (2022, 5 juillet). « Publier un *Data paper* : enjeux de bonnes pratiques, d'intégrité scientifique et de science ouverte ». Conférence en ligne. <https://www.canal-u.tv/chaines/callisto/publier-un-data-paper-laurence-dedieu>
- Directive européenne CELEX 32008R1205 FR. Règlement (CE) n° 1205/2008 de la Commission du 3 décembre 2008. Journal Officiel de l'Union Européenne, 4.12.2008, 19 p.
- DoRANum (2018). La minute *Publier un Data paper*. <https://doi.org/10.13143/4mhn-mq42>
- DoRANum (2020). *Format ouvert ou fermé ?* <https://doi.org/10.13143/mcwq-qs64>
- DoRANum (2020). Webinaire *Data paper – Une incitation à la qualification et à la réutilisation des jeux de données*. <https://doranum.fr/2020/11/20/supports-du-webinaire-data-paper-une-incitation-a-la-qualification-et-a-la-reutilisation-des-jeux-de-donnees/>
- Gay, V. (2021). Un *Data paper* en SHS : pourquoi ? pour qui ? comment ? #dhnord2021 — Publier, partager, réutiliser les données de la recherche : les *Data papers* et leurs enjeux. MESHs, Nov. 2021. [hal-03434216](https://hal.archives-ouvertes.fr/hal-03434216)
- LRN (2016). Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746/>

MESHS Lille (2021). Publier, partager, réutiliser les données de la recherche : les *Data papers* et leurs enjeux. Colloque, 8 – 9 novembre 2021. https://www.meshs.fr/page/accéder_en_ligne_a_dhnord2021_informations_pratiques

Open Data Soft (2020). *Métadonnées : ce qu'il faut savoir avant de publier vos données*. ODS Blog, 29/07/20. <https://www.opendatasoft.com/fr/blog/metadonnees-ce-quil-faut-savoir-avant-de-publier-vos-donnees/>

RGPD : de quoi parle-t-on ? (2018, 10 avril). <https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-on?>

Schöpfel, J., Farace, D. J., Prost, H. et Zane, A. (2020). Data papers as a new form of knowledge organization in the field of research data. *Knowledge Organization*, 46(8), 622-638.

Tripier-Mondancin, O. (2022). *Vers une modélisation du rapport aux valeurs et aux choix en éducation musicale*. [Habilitation à Diriger des Recherches, Université Toulouse - Jean Jaurès].

Tripier-Mondancin, O., Roubertie Soliman, L., Bourg, A. (article accepté). Analyse comparée de recherches en sciences de l'éducation et de la formation musicales fondées sur des corpus de films de classe (2004-2024). Vers le partage de la science. Presses Sorbonne Nouvelle.

Références complémentaires

Ministère de l'Enseignement Supérieur et de la Recherche. *Ouvrir la science !* <https://www.ouvrirlascience.fr/accueil/>

Vasile, A. et Capelli, L. (2021). *Tuto39_Nakala_Capelli-Vasile*. <https://mate-shs.cnrs.fr/actions/tutomate/tuto39-nakala-capelli-vasile/>