Les données dans la musique

Enjeux et stratégies d'investissement

Par Julie Knibbe

Knibbe Julie

Julie Knibbe a fondé Music Tomorrow, une société de conseil spécialisée dans l'intelligence des données pour le secteur musical. Elle a plus de dix ans d'expérience dans le développement de solutions de données innovantes pour des acteurs comme Deezer, Soundcharts, la Sacem, Bandsintown, Revelator ou encore Believe, et a notamment lancé des fonctionnalités phares comme Flow, les radios interactives et la recommandation de playlists chez Deezer.

Introduction

Chaque jour, entre 60 000 et 100 000 nouvelles chansons sont distribuées sur les plateformes de streaming [1] : un chiffre minoré et voué à augmenter si l'on prend en compte le développement de réseaux comme TikTok, des plateformes de jeux vidéo ou encore des livestreams. Dans le même temps, le nombre de diffusions (radios, TV, téléchargements et streams) à analyser par les sociétés de collecte est passé de quelques milliards à presque 20 milliards de milliards depuis 2011. Dans son rapport « Drowning in Data », l'éditeur de solutions de gestion de royalties Synchtank estime que les volumes de diffusion continueront de croître à un taux minimum de 20 à 30 % par an au cours de la prochaine décennie.

Cette hausse, caractéristique de la transition numérique de la filière musicale, s'accompagne de la génération de vastes volumes de données. En effet les titres distribués sont accompagnés de métadonnées. En l'occurrence, les métadonnées musicales sont des informations qualitatives ou quantitatives sur une chanson ou une action numérique donnée. On en distingue généralement trois types, sans ordre particulier :

- la description d'un enregistrement : titre de la chanson, album, interprètes, label, date de sortie, genre musical (tel que déclaré par l'artiste ou le label), pochette de couverture, paroles, liens sur les différentes plateformes, etc. ;
- les métadonnées sur le signal sonore : durée, genre/sous-genres (tels que déduits par l'analyse audio), ambiance, tonalité, mode, volume, tempo, etc. ;
- les métadonnées relatives aux droits : artistes-interprètes, producteurs, auteurs, compositeurs, arrangeurs, éditeurs, sous-éditeurs et leurs clés de répartition respectives.

Par ailleurs, chaque action, comme un stream, un post sur les réseaux sociaux, ou encore l'achat d'un billet de concert en ligne, est associée à une empreinte numérique. Ces données sont collectées à partir de différentes sources (les plateformes de streaming musical, les portails de vente de billets en ligne, les réseaux sociaux, etc.) et prennent différentes formes :

- les données de consommation : nombre de streams, nombre d'ajouts à la bibliothèque, ajouts en playlists, taux de *skips* (les morceaux passés sans être écoutés), durée d'écoute moyenne, popularité relative, etc. ;
- les données démographiques sur les publics, telles que leur âge, leur genre, leur localisation géographique, leurs préférences musicales, leurs historiques d'achats, de vues, etc. ;
- les données relatives aux tournées et aux concerts, telles que le nombre de billets vendus, le calendrier des concerts, les salles, les recettes associées, etc. ;

- les données sur les paroles des chansons, telles que le nombre de mots, le nombre de vers, la fréquence des thèmes et des mots-clés, etc. ;
- les données sur les artistes, auteurs et compositeurs, telles que leur biographie, leur discographie, leur production musicale, leur activité en ligne, etc.

Suivre le rythme en matière de stockage, d'analyse et de monétisation de ces données s'avère compliqué et coûteux, ce qui peut entraver la croissance et la rentabilité de la filière. Les professionnels du secteur bénéficient malgré tout d'une opportunité unique : utiliser ces données pour mieux automatiser les traitements, comprendre leurs publics et personnaliser leur approche en vue de maximiser leur impact.

Les données, incontournables dans tous les métiers

Aujourd'hui, l'exploitation et l'analyse de ces données font partie intégrante de la majorité des métiers du secteur de la musique. La maîtrise de ces activités est devenue une compétence clé pour réussir : toute personne travaillant dans la filière musicale est irrémédiablement en contact avec des données.

Les données jouent un rôle crucial pour les professionnels dans leur prise de décision et leurs stratégies de commercialisation et de promotion. Grâce à ces outils, ils peuvent comprendre les habitudes d'écoute des auditeurs, personnaliser les stratégies, orienter les carrières des artistes. En effet, l'exploitation des données sert à suivre le rythme des tendances du marché et les attentes des publics. Ainsi, les labels utilisent des indicateurs d'engagement de communautés pour choisir les canaux de promotion de leurs artistes, à l'instar des promoteurs qui examinent l'historique de billets vendus dans une salle pour le spectacle d'un artiste avant de le programmer. Les artistes aussi sont concernés : ils s'appuient sur les métadonnées liées au signal sonore pour explorer de vastes banques de sons, et les plus curieux d'entre eux explorent l'usage d'algorithmes d'intelligence artificielle pour les accompagner dans la création (tableau 1).

Récemment, David Mahieux, directeur du pôle data de Warner Music France, expliquait au média en ligne FrenchWeb comment son équipe utilise les données pour faire des projections de streams : « Avec notre projet Hype Habits, on montre comment le choix de morceaux en concert dans une zone géographique va avoir des conséquences dans cette zone sur le streaming de certains titres après le show . » Autrement dit, le choix des chansons jouées en concert influence ensuite les écoutes sur les plateformes de streaming. Une vidéo interne illustre comment Ed Sheeran s'est montré « fasciné » par ce développement qui lui permet d'optimiser différemment les écoutes sur les plateformes après un concert, que ce soit à Paris ou à Porto Rico.

Persona	Exemples de cas d'usage	Exemples de données utilisées

Artiste	Utiliser efficacement les logiciels de production et les outils d'aide à la créativité	Métadonnées liées à l'analyse du signal sonore (explorer une banque de sons à l'aide de filtres sur les instruments, tempo, BPM…)
Manager d'artiste	Promouvoir son artiste auprès de futurs partenaires (label, distributeur, éditeur…)	Paliers de streams et *followers* atteints Progression des auditeurs mensuels sur les plateformes de streaming Progression des abonnements de fans sur les réseaux sociaux Taux d'engagement sur les fournisseurs de service numérique et les réseaux sociaux Mentions médias/radio
Direction artistique	Évaluer le profil d'un artiste avant signature pour déterminer son potentiel et ses objectifs de développement	S'ajoute aux indicateurs ci-dessus Vérification de l'absence de *fake streams* Comparaison de ces données avec des artistes aux trajectoires potentiellement similaires
Label/*label services*	Développer l'audience d'un artiste	S'ajoute aux indicateurs ci-dessus Taux de conversion des fans entre différentes plateformes Taux d'engagement des différents canaux de promotion (playlists, influenceurs, médias…)
Distributeur	Optimiser les revenus d'un catalogue	Progression et distribution du revenu par territoire, par offre et par plateforme Analyse des vecteurs de croissance Audit des métadonnées
Éditeur	Développer le catalogue d'un auteur/compositeur	Œuvres et contrats d'édition/sous-édition Rapports d'exploitation des œuvres (répartitions des OGC) pour analyse de revenus et audit (*tracking*)
Agent/ *booker*	Évaluer le profil d'un artiste pour un événement ou le *booking* dans une salle	Indicateurs communs aux managers, directions artistiques et labels, filtrés et croisés pour une ville et un pays donné : fans et *followers* dans une ville + analyse des passages radio et promotion locale Historique des billets vendus Progression des ventes Taux de remplissage

Promoteur	Organiser un événement	S'ajoute aux indicateurs ci-dessus Heures d'arrivée/départ Traçage des contacts (Covid) Achats durant l'événement (*catering*, *merch*… Reventes sur les marchés secondaires Suivi CRM

Tableau 1. Exemples et cas d'usages observés par persona dans la filière musicale.

Chaque année, Music Tomorrow mène une enquête à l'échelle internationale auprès des professionnels de la filière musicale pour rendre compte de l'état de l'art de ces pratiques autour de l'exploitation des données. Près de 80 entreprises du secteur ont répondu à un questionnaire d'environ 40 questions, faisant état de leurs pratiques autour de l'analyse et l'exploitation des données. En 2022, 68,6 % des professionnels de la musique interrogés ont déclaré utiliser des données au moins une fois par jour (en 2021, ce taux était de 48 % [2]). Les données sont non seulement le carburant qui fait tourner la filière de la musique numérique, mais elles constituent également le socle sur lequel des milliers de professionnels fondent leurs décisions opérationnelles et stratégiques.

Les professionnels de la musique ont ainsi accès à de nombreux outils et tableaux de bord, agrégeant parfois des dizaines voire des centaines de sources de données différentes. Savoir tirer parti au mieux de ces tableaux de bord fait partie des nouvelles compétences à maîtriser. Leur prise en main nécessite accompagnement et formation, et les instituts de formation spécialisés dans le secteur musical incluent désormais des modules consacrés à l'exploitation des données, comme en témoignent les programmes pédagogiques du Berklee College of Music, du Centre national de la musique (CNM) ou de l'École de management et des industries créatives (EMIC). De plus, de nombreuses entreprises ont cette année fait leurs premiers recrutements de *data analysts* et *data scientists* pour exploiter ces données et développer leurs propres solutions. 76,5 % des entreprises interrogées, tous secteurs confondus, ont de fait prévu d'augmenter leurs budgets de gestion et d'analyse des données en 2023 [3] .

L'attitude des professionnels par rapport aux données a très largement évolué ces dernières années. S'ils ont tout d'abord manifesté une certaine défiance devant ces nouvelles méthodes de travail, ils com- prennent de mieux en mieux la valeur qu'ils peuvent tirer de l'exploitation des données pour les assister. Les débats consistant à opposer des approches intuitives (où la qualité artistique prime) à des approches fondées avant tout sur les données (*data-driven*) sont de moins en moins nombreux, au fur et à mesure que la maturité sur le sujet se développe.

1. Des enjeux spécifiques à chaque secteur

La filière musicale est caractérisée par la fragmentation et la complexité, et cela se reflète souvent dans la gestion des données. Les différents acteurs de la filière gèrent leurs propres bases de données de manière indépendante, pour répondre aux enjeux qui leur sont propres. Ils utilisent des systèmes d'information et des méthodologies différentes pour collecter, stocker et analyser les données, ce qui peut rendre difficile la comparaison et l'utilisation conjointe des données.

La musique enregistrée

Le secteur de la musique enregistrée doit faire face à l'enjeu majeur de l'économie de l'attention et de la création : faire émerger une chanson ou un artiste parmi des milliers, voire des millions, de titres, dans un monde où l'attention des auditeurs est sollicitée de toutes parts. Les professionnels utilisent les données comme un outil pour réduire les risques et optimiser leurs investissements.

Aujourd'hui, les artistes et leurs équipes ont accès à des données clés sur la consommation de musique et le comportement des auditeurs de presque tous les acteurs de la chaîne de valeur de la musique, y compris, mais sans s'y limiter, les plateformes de streaming audio et vidéo, les réseaux sociaux, les stations de radio, la télévision et plus généralement l'audiovisuel. Les plateformes de streaming fournissent des outils pour suivre les écoutes en temps réel, comme Deezer Backstage, Spotify for Artists ou encore Apple Music for Artists. Les distributeurs fournissent quant à eux une vue agrégée de ces données, et des plateformes de veille de marché comme Chartmetric ou Soundcharts permettent de consolider ces informations avec des statistiques complémentaires sur la performance des artistes sur les réseaux sociaux ou en radio.

Le secteur de la musique enregistrée dispose d'un avantage de taille par rapport au spectacle vivant ou l'édition musicale : les données générées par les plateformes numériques concernent directement les enregistrements. Un stream sur une plateforme de streaming est directement identifié par un ISRC (*international standard recording code*), associé à un titre de chanson et ses artistes-interprètes. Bien que ce standard soit encore imparfait et qu'un enregistrement soit encore souvent associé à de multiples ISRC [4] , le suivi de la performance des titres est facilité grâce aux différents outils cités précédemment.

Le développement de carrières d'artistes implique de nombreux processus et enjeux liés aux données, comme pour la détection de talents : face à l'impossibilité d'écouter tous les nouveaux titres, les directions artistiques ont recours à des outils d'intelligence de marché pour identifier et évaluer les artistes avec qui elles envisagent de collaborer. De même, pour optimiser la performance des opérations de promotion et de marketing numérique, les professionnels s'appuient sur des données de ciblage, qu'ils adaptent en temps réel en fonction des résultats de leurs opérations. Cependant, les problèmes d'attribution, c'est-à-dire la difficulté à relier une écoute sur une plateforme de streaming à une provenance externe, comme un post sur un réseau social, empêchent le secteur d'égaler les performances publicitaires d'autres industries [5] .

Pour faire face à l'augmentation constante de titres distribués, les plateformes ont mis en place puis perfectionné des algorithmes de recommandation, en vue de proposer le bon titre à la bonne audience. Des pages d'accueil aux playlists, toutes les fonctionnalités phares de ces plateformes sont adaptées selon les profils. Sur Spotify, par exemple, plus d'un tiers de toutes les découvertes de nouveaux artistes se produisent grâce à des fonctionnalités personnalisées comme les playlists *made for you*, selon leur rapport publié en 2022 [6] .

Réussir à être bien identifié et classifié par ces algorithmes pour être présenté aux cibles appropriées, autrement dit développer sa découvrabilité en ligne, est un enjeu majeur pour le secteur de la musique enregistrée. Les professionnels comptent bien souvent sur ces algorithmes pour augmenter la visibilité de leurs artistes et pour optimiser les investissements réalisés sur les plateformes en ligne. Cette problématique de découvrabilité est autant liée au déploiement de nouveaux titres qu'au développement des *back catalogues*. Utiliser au mieux les données pour exploiter les catalogues existants est clé pour le secteur. Pour cela, les professionnels s'appliquent à comprendre les bases du fonctionnement de ces algorithmes et à adapter leurs stratégies, tant en distribution qu'en marketing, pour en tirer parti. Cependant, un rapport récent du gouvernement britannique sur l'impact des recommandations algorithmiques sur la filière souligne à quel point le manque d'informations et de formation sur ces sujets se fait ressentir [7] : 83 % des créateurs interrogés n'étaient pas d'accord ou fortement en désaccord avec le fait que les plateformes de streaming fournissent suffisamment d'information sur la façon dont les recommandations sont faites aux consommateurs, et 89 % ont exprimé le désir d'avoir plus de détails sur la façon dont les recommandations sont faites.

La multiplication de faux streams a également fait l'objet d'une enquête récente du Centre national de la musique [8] . Des plateformes telles que Deezer, Qobuz et Spotify ont coopéré en fournissant des données sur les faux streams détectés sur le marché français. Cette première enquête en la matière a permis de montrer l'importance de la lutte contre les faux streams pour garantir la rémunération juste des ayants droit. Il reste cependant des défis à relever afin de trouver des solutions adéquates. Le rapport montre la nécessité de la coopération complète des acteurs du secteur et le besoin d'apporter des solutions aux professionnels pour mieux comprendre et contrôler leur visibilité sur les plateformes en ligne.

L'édition

Le secteur de l'édition quant à lui fait face à des défis similaires pour développer les carrières des auteurs-compositeurs, et doit, en outre, s'attaquer aux enjeux liés aux données, notamment aux métadonnées, qui lui sont propres.

La plupart des entreprises qui monétisent ou licencient des enregistrements musicaux et des œuvres à grande échelle – plateformes de streaming, réseaux sociaux, salles de concert, etc. – s'appuient sur les métadonnées relatives aux droits pour rémunérer les principaux ayants droit. Les acteurs de la musique enregistrée concluent généralement des accords directs avec les plateformes de streaming : celles-ci déclarent ensuite les ventes et

rémunèrent directement les ayants droit sur une base mensuelle.

Le secteur de l'édition, cependant, subit des délais bien plus longs entre le moment où les auteurs-compositeurs sortent un titre et le moment où ils reçoivent leur premier paiement – environ six à douze mois après la première diffusion d'une chanson. Non seulement les accords directs entre les éditeurs et les plateformes de streaming sont rares, mais la gestion des droits d'édition varie également considérablement selon les territoires. Selon le pays, une seule société peut gérer à la fois les droits mécaniques et les droits d'exécution (par exemple, la Sacem en France ou l'Apra/Amcos en Australie), ou plusieurs sociétés différentes peuvent être chacune responsable de la collecte d'un type de droit distinct. En outre, les répartitions exactes entre les droits mécaniques et les droits d'exécution, et entre les parts de l'auteur-compositeur et de l'éditeur, diffèrent d'un pays à l'autre, souvent en fonction des réglementations gouvernementales.

Toutes ces entités ont besoin de savoir qui est impliqué dans une œuvre et un enregistrement donnés, pour répartir correctement les revenus. Il est important de noter que ces plateformes reposent en grande partie sur les métadonnées que les artistes et les ayants droit saisissent eux-mêmes. Si les plateformes reçoivent des informations incomplètes ou inexactes de la part de ces partenaires – ou si les ayants droit ne soumettent tout simplement pas les informations sur leurs œuvres –, elles ne peuvent pas retracer les revenus jusqu'à leurs propriétaires légitimes.

Le nombre d'acteurs impliqués dans la gestion des métadonnées liées aux droits rend ces métadonnées extrêmement difficiles à gérer. Le niveau de savoir-faire technique que les entreprises doivent mettre en œuvre pour trier jusqu'à la partie la plus simple des métadonnées comme l'identifiant d'une œuvre est tout simplement considérable.

Selon la start-up Blokur, qui fournit un service automatisé de détection et de réconciliation des conflits de métadonnées, seulement 30 % des chansons dans les bases de données des éditeurs ont un identifiant associé à une œuvre (ISWC [9] , pour *international standard work code*), ce qui est une moyenne optimiste.

Bien que des solutions de gestion des métadonnées, comme Synchtank ou Blokur pour ne citer qu'elles, visent à résoudre ce problème à grande échelle, les défis liés aux métadonnées demeurent au premier plan des activités d'édition [10] . Pour la majorité des sociétés d'édition interrogées dans notre enquête, la gestion des métadonnées est la priorité pour les investissements futurs liés aux données [11] .

À ce jour, il n'existe pas de base de données qui fasse consensus pour les informations sur la propriété des œuvres. Jusqu'à présent, les tentatives pour construire un tel outil se sont soldées par un échec. L'exemple le plus notoire est certainement la Global Repertoire Database (GRD), une initiative à l'échelle de l'industrie fondée en 2008 et soutenue par les investissements d'une coalition d'éditeurs et de sociétés de collecte, dans le but de fournir « une représentation unique, complète et faisant autorité de la propriété, de l'administration et du contrôle des œuvres musicales à l'échelle mondiale », pour reprendre les propos de l'organisme de gestion collective PRS for Music. Après six ans de va-et-vient et des

financements significatifs, la GRD a finalement été mise de côté, en raison de la difficulté et de la réticence des sociétés de collecte à partager les données relatives aux droits ouvertement. La principale leçon à tirer des retombées de la GRD est que les retards historiques dans l'innovation financière dans la filière musicale ont été autant politiques et contractuels que techniques.

L'ambition de cette base de données référente revient aujourd'hui sur le devant de la scène avec l'initiative Verifi Rights Data Alliance, lancée début 2022 [12] . Ses quatre premiers membres, Warner Music Group, Deezer, FUGA et Unison, ont prévu de créer un référentiel sur les droits associés aux enregistrements et aux œuvres, dans lequel chaque acteur a la responsabilité de ses propres données.

Le secteur de l'édition est pleinement concentré sur cette problématique : il a en effet mis la priorité sur la gestion et le traitement de métadonnées liées aux droits pour accélérer les flux financiers. À cela s'ajoutent les enjeux liés à la traçabilité des diffusions et à l'amélioration de la visibilité sur la performance des œuvres. Les éditeurs ont besoin d'avoir une vue complète de la performance de leurs œuvres musicales pour pouvoir planifier et développer leurs stratégies de marketing. Cependant, la fragmentation des données rend difficile leur collecte et leur analyse pour obtenir cette synthèse. En effet, les données de performance sont associées aux enregistrements, et non aux œuvres. Actuellement, la chaîne de valeur globale de la musique favorise la création de valeur (et la création de données) au niveau de l'artiste-interprète, et non au niveau de l'auteur-compositeur. Ainsi, pour mettre en place ce suivi de performance, il faut être capable de rattacher les enregistrements aux œuvres concernées, ce qui pose à nouveau la question des métadonnées et de leur partage à grande échelle.

Le spectacle vivant

De la même manière, le secteur du spectacle vivant fait face à des défis uniques, également liés à la propriété et au partage de données entre acteurs, et aux problématiques techniques, politiques et réglementaires qui y sont associées, comme la protection des données personnelles. En effet, les données spécifiques aux métiers du spectacle vivant comme la vente de billets, la capacité d'accueil des salles, les historiques de vente, etc., sont souvent fragmentées entre de multiples acteurs et plateformes (plusieurs plateformes de billetterie, plusieurs salles, différents outils CRM, etc.) [13] .

Lorsque ces données sont collectées, les professionnels du secteur rencontrent beaucoup de difficultés pour y accéder et les interpréter. Réunir ces informations dans des tableaux de bord relève d'un parcours du combattant, et de plus en plus de solutions comme Arenametrix ou encore RealCount visent à résoudre ce problème. De plus, en raison des silos existants, sécuriser un accès aux données provenant des artistes et des labels est souvent laborieux pour les professionnels du spectacle vivant, ce qui les prive de sources de données de valeur pour ce qui concerne la programmation ou le ciblage d'audiences. Ceux-ci ont ainsi souvent recours à des tableaux de bord ou solutions tierces comme Chartmetric ou Soundcharts pour exploiter les statistiques des artistes alors même qu'ils collaborent

avec leurs équipes.

Par ailleurs, la nature non numérique de l'expérience d'un spectacle vivant rend plus complexes la collecte et l'exploitation d'autres types de données. Une opportunité majeure qui reste inexploitée dans l'analyse des données sur les événements concerne par exemple les données qualitatives plutôt que purement quantitatives. Quels étaient l'ambiance et le sentiment du public ? Comment mesurer de façon fiable les retombées ?

Selon Live Nation, 68 % des fans qui ont assisté à un événement ont partagé leur expérience sur les médias sociaux. Cela dit, bien que l'évaluation des ventes de merchandising soit aisée, mesurer l'impact de ce qui est publié sur les médias sociaux est plus difficile, car les artistes ou les lieux ne sont pas toujours identifiés dans de tels contenus. En dehors du moteur d'attribution de Pex, très peu d'acteurs sont capables de surveiller ces contenus générés par les utilisateurs à grande échelle, et presque aucune option n'existe pour relier ces données avec un CRM.

Les livestreams, et plus généralement les expériences proposées dans des environnements virtuels, présentent également une opportunité dans la mesure où ceux-ci ont eu un impact significatif sur la façon dont les artistes, les agents et les promoteurs intègrent et partagent les données avant, pendant et après un spectacle. Étant donné que tout ce qui se passe en ligne est mesuré, les données sur les événements et les spectacles en ligne deviennent plus accessibles que jamais. « L'une des principales différences maintenant, c'est que l'artiste est vraiment le promoteur », indique Diana Gremore [14] , fondatrice de RealCount et anciennement analyste chez Paradigm. Elle ajoute : « Vous ne pouvez pas vraiment compter sur un promoteur autant que vous le feriez normalement pour vendre des billets. Le public potentiel est maintenant là où se trouvent vos fans, et pas seulement le marché géographique habituel du promoteur. » De plus, la signification du marketing « local » prend une forme différente en ligne, c'est-à-dire que, selon Diana Gremore, « ces événements ne sont pas géographiquement locaux, mais locaux au niveau des communautés en ligne [15] ». Les artistes et leurs managers peuvent par ce biais disposer de plus de données sur leurs événements. Plus précisément, ils peuvent interagir directement avec les fans, obtenir des statistiques sur leurs comportements et éventuellement inclure leurs contacts dans le CRM de l'artiste – des activités qui sont généralement difficiles à réaliser dans le domaine des événements physiques, où le promoteur n'a pas toujours l'occasion ni la possibilité de partager ces informations avec l'artiste. Cela signifie qu'il existe dorénavant des conversations plus directes entre les managers, les labels, les agents et les promoteurs au sujet de la propriété des données sur les événements. Des transactions récentes, comme l'acquisition par BMG du promoteur de concerts indépendant Undercover et l'acquisition récente par Eventbrite de ToneDen, un outil publicitaire populaire pour les artistes et les labels, témoignent de la façon dont cette quête de données supplémentaires conduit des entreprises auparavant cloisonnées à converger plus rapidement qu'auparavant.

Jusqu'à récemment, les données collectées lors des événements servaient principalement à des fins de rentabilité et de marketing : il s'agissait de s'assurer que la logistique et les opérations se déroulent bien et que l'événement soit rentable, de mieux comprendre le

comportement des fans pendant l'événement, afin d'optimiser les spectacles suivants et de développer des campagnes marketing ciblées post-événement. Depuis 2020, la nécessité de connaître les déplacements des individus lors de la pandémie du Covid-19 a permis le développement rapide d'une nouvelle génération d'outils de suivi des événements (Aloompa, Appmiral ou encore atVenu), qui permettent de tracer les actions et mouvements des participants aux événements, à un rythme beaucoup plus soutenu que prévu [16].

2. Innovation et stratégies d'investissement

Si le panorama des enjeux propres à chaque secteur dessine des différences de rapport à la donnée au sein de l'écosystème musical, les innovations et les pistes d'investissements sur la donnée font envisager des perspectives communes à l'ensemble des acteurs.

En premier lieu, pour la plupart des indépendants qui continuent à travailler avec des feuilles de calcul – 47 % d'entreprises concernées dans notre enquête [17] –, la priorité va consister à s'outiller, recruter et se former pour intégrer l'analyse de données dans leurs processus quotidiens.

À l'inverse, les pratiques les plus avancées en matière d'innovation autour des données musicales se trouvent employées par les plus grandes entreprises, et en particulier celles du secteur de la musique enregistrée. En effet, celles-ci ont un accès facilité aux plus vastes jeux de données, ce qui leur confère un avantage de taille.

Ainsi, les plateformes de streaming, les majors, les distributeurs et les sociétés de gestion collective sont les figures de proue de l'adoption de technologies émergentes de mégadonnées, comme l'analyse prescriptive ou l'intelligence artificielle [18] . Ces types d'algorithmes nécessitant *a minima* des dizaines de milliers de points de données pour fonctionner, la majorité des indépendants ne disposent pas de jeux de données ni de ressources techniques et financières suffisantes pour les développer.

Une telle disparité en termes d'accès, d'outillage et de compétences accentue ainsi l'écart entre les plus grands acteurs et les autres. Les entreprises disposant des ressources et des compétences nécessaires pour développer ces algorithmes de pointe ont un avantage concurrentiel sur les autres, ce qui pourrait accroître les inégalités économiques à plus long terme. De plus, des algorithmes entraînés avec des jeux de données incomplètes ou biaisées peuvent refléter et perpétuer les biais existants dans la filière. Par ailleurs, les problématiques liées au partage et à la confidentialité des données empêchent souvent l'écosystème des start-up de mettre la main sur de précieux jeux de données privés, favorisant leur acquisition ou leurs partenariats avec des acteurs plus importants.

Améliorer la qualité et fluidifier les échanges de données

Comme nous l'avons vu dans la section précédente, beaucoup d'investissements consistent à améliorer la qualité et fluidifier les échanges de données. La collaboration entre différents acteurs, et le partage de données, est clé pour l'innovation dans la filière. Les silos, les systèmes d'information historiques, la confidentialité des données et le manque de personnel qualifié pour structurer la collecte d'informations empêchent souvent le partage de données clés. En 2022, plusieurs initiatives de *hubs* de données (*data hubs*), comme Gaia-X [19] ou encore Verifi Rights Data Alliance, ont ainsi vu le jour pour pallier cette problématique, mais elles devront faire face à des enjeux tant politiques que technologiques pour réussir, à l'instar de la GRD. Par ailleurs, de nombreuses initiatives d'amélioration des standards [20] et de formalisation du partage de données s'efforcent de combler les difficultés mentionnées ci-dessus [21].

Les outils issus du mouvement Web3 [22], visant à construire une nouvelle génération d'Internet décentralisée, transparente, sécurisée et participative, présentent également des opportunités. De nouveaux outils permettent par exemple d'enregistrer les données liées aux droits d'auteur sur une base de données blockchain pour fiabiliser les paiements, comme l'Artist Wallet de Revelator, ou encore de créer des preuves d'antériorité pour les créateurs comme le projet Musicstart de la Sacem. Toujours dans le domaine des droits d'auteur, la Sacem a récemment annoncé un partenariat avec Pianity [23], start-up qui crée des jetons non fongibles (NFT) musicaux. Ces projets, reposant sur l'expertise et le savoir-faire « Web2 » d'acteurs établis et intégrant l'innovation permise par le Web3, montrent le développement des solutions basées sur la décentralisation et la blockchain. Cependant, ces expérimentations sont isolées, et il n'existe pas encore de standards de métadonnées pour les NFT [24].

Des solutions pour trouver l'information et faciliter la prise de décision

Une fois les données collectées et partagées, une autre stratégie d'investissement consiste à développer des solutions pour trouver l'information pertinente parmi ce volume grandissant de données : le nombre croissant de créateurs et de données générées ne permettent plus le traitement manuel. Les processus de découverte d'artistes, de signatures ou encore de suivi doivent maintenant intégrer un certain niveau de fouille de données pour suivre le rythme. De la détection d'anomalie dans les rapports de vente pour optimiser la performance d'un catalogue, à des services de découverte et de filtrage avancés pour les directions artistiques, les professionnels ont de plus en plus recours à des technologies et services de mégadonnées pour les accompagner dans leurs tâches quotidiennes.

Ceux-ci ont désormais pris l'habitude de consulter des tableaux de bord pour suivre leurs activités. Ce premier palier étant franchi, une nouvelle génération d'outils personnalisés voit le jour, fondés sur des algorithmes d'analyse prédictive et prescriptive. L'analyse prescriptive est une forme d'analyse qui utilise des algorithmes et des techniques de modélisation pour déterminer la meilleure action à prendre dans une situation donnée. Elle se concentre sur le futur et se repose sur les données historiques pour prédire les résultats possibles d'une action donnée et recommander la meilleure solution. L'analyse prescriptive se différencie de l'analyse descriptive et de l'analyse prédictive dans le sens où elle ne se

contente pas de décrire ou de prédire les tendances dans les données, mais fournit également des solutions concrètes et des recommandations pour atteindre des objectifs précis.

Des solutions d'analyse descriptive et généraliste comme Soundcharts ou Chartmetric continuent de se développer, mais en complément de solutions d'analyse prescriptive centrées sur des tâches spécifiques. Les acteurs de la filière investissent dorénavant dans l'industrialisation d'algorithmes d'intelligence artificielle spécifiques à leurs opérations afin d'améliorer et d'automatiser certaines prises de décisions, comme la prédiction de ventes de billets, la prédiction des flux financiers sur un catalogue, ou encore la prédiction de performance de campagnes marketing, etc. Du développement de technologies d'automatisation des campagnes marketing comme Ingrooves à l'optimisation de la présence algorithmique proposée par Music Tomorrow, un nouvel écosystème de méthodologies et d'outils prend forme [25].

L'usage d'intelligences artificielles génératives se remarque également pour assister, non seulement la création, mais aussi la prise de décision. La fin d'année 2022 aura été marquée par la sortie de ChatGPT, intelligence artificielle sous forme de *chat* capable de converser et générer des connaissances structurées sur de nombreux sujets. Les intelligences artificielles sont non seulement capables de générer des chansons, des paroles, mais aussi de créer des images à la demande (DALL.E) pouvant servir de couverture d'album ainsi que des articles de référence. Les professionnels les plus avertis s'adaptent pour utiliser ces nouveaux outils, dont les possibilités s'étendent bien au-delà de la création artistique. Par exemple, un label utilise ChatGPT pour générer ses modèles de communication externe, un agent pour éditer un dossier de presse, etc.

Conclusion

À une époque où des milliards de nouveaux points de données sont générés chaque jour, la filière musicale se déploie pour assurer son bon fonctionnement et tire parti de ce potentiel dans la conception de nouveaux services. Jusqu'à présent retardés dans l'adoption de technologies de mégadonnées et d'IA en raison de la difficulté à collecter, nettoyer et exploiter de larges jeux de données, les professionnels affichent désormais une volonté de développement.

Les temps sont passionnants : l'intelligence artificielle, qui avait perdu un peu de son attrait lors de l'engouement récent pour le Web3 et les NFT, est à nouveau sous les projecteurs depuis la sortie par OpenAI de ChatGPT fin 2022. Cette annonce a remis sur le devant de la scène cet écosystème, dont les progrès ne cessent de s'accélérer.

La compréhension des grands principes et du fonctionnement de ces algorithmes va cependant demeurer essentielle pour en tirer pleinement parti. Le traitement manuel des données n'est plus une option envisageable. Le développement d'une nouvelle génération d'outils d'exploration et d'aide à la décision fondés sur des algorithmes d'intelligence

artificielle s'intensifie, pour accompagner la croissance rapide des volumes de données dans la filière : le nombre grandissant d'ingénieurs recrutés en témoigne. Malgré tout, il est crucial de continuer à investir dans l'ensemble des métiers entretenant la connaissance de ces technologies, notamment les professions moins techniques qui seront à même d'effectuer ce nécessaire travail de vulgarisation et de formation, afin que celles-ci soient adoptées à bon escient par tout un chacun.

1. Ingham T., « It's happened, 100,000 tracks are now uploaded to streaming services like Spotify each day », Music Business Worldwide, octobre 2022, en ligne : www.musicbusinessworldwide.com/its-happened-100000-tracks-are-now-being-uploaded.
2. Knibbe J. et Pastukhov D., « State of data in the music industry », Music Tomorrow, novembre 2021, en ligne : www.music-tomorrow.com/report.
3. Ibid.
4. KnIbbe J., « Understanding music rights data. The challenges of delivering timely royalty payments to artists », Water and Music, avril 2021, en ligne : www.waterandmusic.com/understanding-music-rights-data-the-challenges-of-delivering-timely-royalty-payments-to-artists.
5. Buger A., « Music advertising has a streaming attribution problem », Musically, avril 2022, en ligne : musically.com/2022/04/25/music-advertising-streaming-attribution-problem-guest-post.
6. « Made to be found », Spotify for Artists,2022, en ligne : found.byspotify.com.
7. « The impact of recommendation algorithms on the UK's music industry », UK Government, Centre for Data Ethics and Innovation, février 2023, en ligne : www.gov.uk/government/publications/research-into-the-impact-of-streaming-services-algorithms-on-music-consumption/the-impact-of-recommendation-algorithms-on-the-uks-music-industry.
8. « Faux streams, vrai phénomène : le CNM, avec les professionnels pour lutter contre la fraude », Centre national de la musique, 16 janvier 2023, en ligne : cnm.fr/faux-streams-vrai-phenomene-le-cnm-avec-les-professionnels-pour-lutter-contre-la-fraude.
9. Barry P., « Leaving Beta. Three things for music publishers to look out for in the new Blokur », Medium, 23 octobre 2019, en ligne : medium.com/blokur/leavingbeta-three-things-for-music-publishersto-look-out-for-in-the-new-blokur-61d8ac709c16.
10. Pastukhov D., « How broken metadata affects the music industry (and what we can do about it) ? », Soundcharts Blog, juin 2019, en ligne : soundcharts.com/blog/music-metadata.
11. Knibbe J. et Pastukhov D., « State of data in the music industry », art. cité.
12. « Verifi rights data alliance launches with inaugural members Warner music group, Unison, Deezer, and Fuga », Verifi Media, février 2022, en ligne : www.verifi.media/press/vrda.
13. Knibbe J., « How the touring industry will use data in 2021 – even in a pandemic », Water and Music, décembre 2020, en ligne : https://www.waterandmusic.com/how-the-touring-industry-will-use-data-in-2021-even-in-a-pandemic.
14. Ibid.
15. Cela va de pair avec la façon dont l'ancien dirigeant d'Apple Music, Ian Rogers, a qualifié le « local » à l'ère d'Internet d'« intérêts communs, peu importe la géographie ».
16. Knibbe J., « How the touring industry will use data in 2021 – even in a pandemic », art. cité.
17. Knibbe J. et Pastukhov D., « State of data in the music industry », art. cité.
18. Pastukhov D., « Under the hood. How revelator uses data to accelerate the streaming economy », Music Tomorrow, décembre 2021, en ligne : www.music-tomorrow.com/blog/under-the-hood-streaming-royalty-processing.
19. Demilly V., « Position paper, GT Industries culturelles et créatives, French Gaia-x Hub », Cigref, juin 2022, en ligne : www.cigref.fr/wp/wp-content/uploads/2022/06/French-Gaia-X-Hub-position-paper-GTICC_v1.2.pdf.
20. Isherwood M., « Early test of MEAD at Amazon Music and Universal Music Group leads to an increase in active and passive streaming behavior», DDEX, juin 2022, en ligne : ddex.net/ddexs-the-big-summit-2022-conference-and-plenary-meeting.
21. Hardjono T., Howard G., Sace E. et al., « Towards an Open and Scalable Music Metadata Layer », Research Gate, 20 novembre 2019, en ligne : www.researchgate.net/publication/337386847_Towards_an_Open_and_Scalable_Music_Metadata_Layer.
22. Hu C., « Starter Packs : Music/Web3 data », Water and Music, 23 janvier 2023, en ligne : www.waterandmusic.com/starter-packs-understanding-music-web3-data.
23. Loignon S., « La Sacem adapte le droit d'auteur aux NFT», Les Échos, novembre 2022, en ligne : www.lesechos.fr/tech-medias/medias/la-sacem-adapte-le-droit-dauteur-aux-nft-1878725.
24. Guichardaz R., Bach L. et Schenk E., « Blockchain et NFT. Quel futur pour l'industrie musicale ? », CNMlab, juin 2022, en ligne : cnmlab.fr/onde-courte/blockchain-et-nft.
25. Pastukhov D., « Towards recommender system optimization. Part 1 : our vision and our data tool for algorithmic optimization on streaming platforms », Music Tomorrow, septembre 2022, en ligne : www.music-tomorrow.com/blog/towards-recommender-system-optimization-data-tool-for-algorithmic-optimization-on-streaming-platforms.