French Gaia-X Hub

POSITION PAPER

GT Industries Culturelles et Créatives

v.1.2 – 01.06.2022

gaia-x

POSITION PAPER

GT Industries Culturelles et Créatives

Préfigurateur du groupe de travail :

Louis-Cyrille Trébuchet, Directeur de l'IT Transverse, France Télévisions

Rapporteur :

Véronique Demilly, Chargée de missions, France Télévisions
Veronique.Demilly@francetv.fr

Table des matières

1 LISTE DES PARTICIPANTS 4

2 LE SECTEUR DES INDUSTRIES CULTURELLES ET CREATIVES 5

Définition Générale 5

Les secteurs des ICC 5

Hétérogénéité du secteur 6

3 IMPORTANCE DE LA DATA POUR LE SECTEUR 8

Contexte 8

1. Numérisation de l'économie des ICC 8
2. Plateformisation de l'accès aux œuvres 9
3. Une stratégie Européenne 9
4. Un soutien français 10

Les différents types de données 11
1. Contenus 11
2. Métadonnées 11
3. Les données de consommation 12
4. Les données personnelles 13
5. Les autres données 13
Assurer la découvrabilité grâce aux métadonnées 13
Affiner les propositions grâce aux données d'usage 13
Garantir la rémunération des ayants droit 14
Rationaliser les ressources 14
Intérêt de l'initiative Gaia-X 15

4 CAS D'USAGE 16
Caractéristiques des projets Gaia-X 16
Organisation des cas d'usage selon la chaîne de valeur 16
Cas d'usage pour la création, la production ou la fabrication 16
1. Dépôt et référentiel des œuvres - Mise en valeur de la création 16
2. Faciliter et documenter la fabrication 20
3. Intelligence artificielle et autorégulation des contenus publicitaires – Projet Invenio 24
Cas d'usage pour l'édition, la distribution ou la diffusion 26
1. Gestion unifiée du consentement des utilisateurs 26
2. Accélérer la rémunération des ayants droit 28
3. Plateforme collaborative de création et de diffusion de contenus éditoriaux – projet CREA 30
Cas d'usage pour la vente de biens culturels 32
1. Agrégation des données de billetterie 32
2. Sécurisation de la billetterie 33
Cas d'usage Transverses 33
1. Alliance Culture Data : projet de plateforme d'échange de données de la filière des Industries Culturelles et Créatives 34

3 BESOINS TRANSVERSES 37
Définition d'une ontologie des Métadonnées liées aux contenus culturels et créatifs 37
Gouvernance d'un Data Space ICC 37

5 POURSUITE DES TRAVAUX 38

6 ANNEXE 39

1 LISTE DES PARTICIPANTS

En date du 17 mai 2022, les participants au groupe de travail sont issus des organisations suivantes :

- Arenamétrix
- ARPP
- BnF
- BnF Partenariats
- Cap Digital
- Comité Colbert
- CST
- Dag Conseil
- Deezer
- Epsilon – groupe Publicis
- France Télévisions
- France Télévisions Publicité
- Frédéric Goldsmith
- Hachette
- INA
- Institut Français
- Médiamétrie
- Minalogic
- Ministère de la Culture
- Monark
- Le Monde
- Nomalab
- One Point
- Panodyssey
- Plurimédia
- Procirep
- Radio France
- RMN – Grand Palais
- Sacem
- Salto
- Singulart
- SNPTV
- UER

2 LE SECTEUR DES INDUSTRIES CULTURELLES ET CREATIVES

DEFINITION GENERALE

Les industries culturelles et créatives (ICC) sont les secteurs d'activité ayant comme objet principal la création, le développement, la production, la reproduction, la promotion, la diffusion ou la commercialisation de biens, de services et activités qui ont un contenu culturel, artistique et/ou patrimonial.

On les reconnaît à plusieurs caractéristiques :

- L'intersection entre l'économie et la culture ;
- La créativité au cœur de l'activité ;
- Le contenu artistique, culturel ou inspiré de la création du passé ;
- La production de biens et de services fréquemment protégés par la propriété intellectuelle — droit d'auteur et droits voisins ;
- La double nature : économique (génération de richesse et d'emploi) et culturelle (génération de valeurs, de sens et d'identité) ;
- L'innovation et le renouvellement créatif ;
- Une demande et des comportements du public difficiles à anticiper ;
- Un secteur marqué par la diversité des modes de rémunération du travail, notamment le salariat et le versement de droits de propriété littéraire et artistique, ainsi que par la prédominance de micro-entreprises.

Ces industries ont été définies comme « toute industrie qui a pour origine la créativité individuelle, l'habileté et le talent et qui a le potentiel de produire de la richesse et de l'emploi à travers la création et l'exploitation de la propriété intellectuelle ».

LES SECTEURS DES ICC

La définition générale des Industries Culturelles et Créatives établit clairement qu'il s'agit d'un groupement de secteurs variés. Selon les études ou les sources, les groupements comprennent plus ou moins de secteurs d'activité.

Ainsi, l'étude EY[1] pour France Créative, dont les chiffres sont fréquemment cités, définit le secteur ICC comme regroupant les 10 secteurs d'activités suivants :

- Les arts visuels ;
- La musique ;
- Le spectacle vivant ;
- Le cinéma ;
- La télévision ;
- La radio ;
- Le jeu vidéo ;
- Le livre
- La presse

[1] « *L'économie mosaïque - 3e Panorama des Industries Culturelles et Créatives en France* » - rapport EY et France Créative – Novembre 2019

- La publicité et la communication.

La commission européenne, dans sa proposition de règlement établissant le programme Europe Créative pour 2021 à 2027, estime que le secteur ICC comprend tous les secteurs dont les activités sont fondées sur des valeurs culturelles ou sur des expressions artistiques et autres expressions créatrices individuelles ou collectives.

Ces activités peuvent inclure le développement, la création, la production, la diffusion et la conservation de biens et services incarnant une expression culturelle, artistique ou toute autre expression créative, ainsi que les tâches qui s'y rapportent, comme l'éducation ou la gestion. Elles auront le potentiel de créer de l'innovation et de l'emploi, en particulier grâce à la propriété intellectuelle. Ces secteurs comprennent ainsi l'architecture, les archives, les bibliothèques et les musées, l'artisanat d'art, l'audiovisuel (y compris le cinéma, la télévision, les jeux vidéo, le multimédia) et la publicité, le patrimoine culturel matériel et immatériel, le design (y compris la mode), les festivals, la musique, le théâtre, les arts du spectacle, la littérature, l'édition, la radio et les arts visuels.

Dans la perspective d'un développement à l'échelle européenne, le groupe de travail Industries Culturelles et Créatives du Hub France Gaia-X est susceptible d'accueillir toute entreprise répondant à la définition européenne du secteur des ICC.

HETEROGENEITE DU SECTEUR

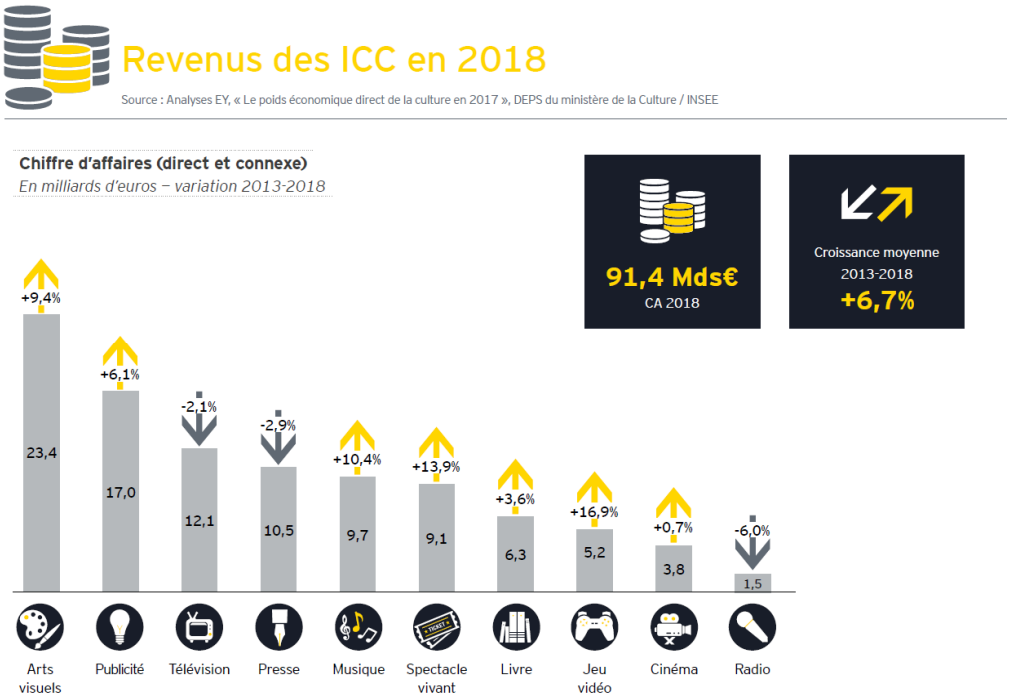
Contrairement à d'autres secteurs d'activité, celui des ICC est composé de structures de taille et de mode de fonctionnement très divers. A côté de grandes entreprises de droit privé ou des établissements publics de grande envergure évoluent de nombreuses très petites entreprises ou des structures fondées sur des modèles associatifs, avec très peu de personnel permanent, des indépendants et des salariés intermittents du spectacle. Dans certains pans du secteur ICC, la majorité des structures sont des micro-entreprises.

Un ensemble de 600 métiers, 300 000 entreprises et 1,3 million de personnes contribuent à la vivacité et à la croissance du secteur des ICC, en France (1, 2 millions d'entreprises et plus de 8 millions de personnes y contribuent en Europe[2] pour un poids économique de 3,95% de la valeur ajoutée en Europe soit 477 Mds€).

Les revenus du secteur des ICC proviennent de sources variées. Il peut s'agir de revenus issus d'une activité commerciale comme de subventions publiques. Le secteur des ICC fait l'objet de régulations spécifiques à ses différentes composantes, qui répondent à l'objectif de préservation de la diversité culturelle comme à d'autres objectifs d'intérêt public tels que le renforcement du lien social, le développement des activités éducatives, les effets de levier sur l'activité touristique ou l'attractivité territoriale.

Compte-tenu des différences de structures au sein des industries Culturelles et Créatives, nombre d'entre elles disposent de peu de moyens pour développer leurs compétences dans le domaine numérique, en particulier lorsqu'il ne s'agit pas de leur cœur de métier. On note d'importants écarts de maturité sur les sujets relatifs à la Data et aux technologies de l'information au sein du secteur.

[2] Commission Staff Working Document 2010 SWD (2021) 351final, pp 98-104 (https://ec.europa.eu/info/sites/default/files/swd-annual-single-market-report-2021_en.pdf)


Revenus des ICC en 2018
Source : Analyses EY, « Le poids économique direct de la culture en 2017 », DEPS du ministère de la Culture / INSEE
Chiffre d'affaires (direct et connexe)
En milliards d'euros – variation 2013-2018
91,4 Mds€
CA 2018
Croissance moyenne
2013-2018
+6,7%
+9,4%
23,4
Arts visuels
+6,1%
17,0
Publicité
-2,1%
12,1
Télévision
-2,9%
10,5
Presse
+10,4%
9,7
Musique
+13,9%
9,1
Spectacle vivant
+3,6%
6,3
Livre
+16,9%
5,2
Jeu vidéo
+0,7%
3,8
Cinéma
-6,0%
1,5
Radio

3 IMPORTANCE DE LA DATA POUR LE SECTEUR

CONTEXTE

1. Numérisation de l'économie des ICC

La généralisation de la numérisation de l'industrie est un des éléments clés de la conquête de l'autonomie stratégique de l'Union Européenne. Le secteur des Industries Culturelles et Créatives n'échappe pas à cette ambition.

Tous les secteurs des ICC sont, aujourd'hui, concernés par cette transformation.

Si, pour les secteurs de la musique enregistrée, du jeu vidéo ou de l'audiovisuel, les nouvelles créations sont depuis plusieurs années numériques de bout en bout, de leur création jusqu'à leur exploitation sous toutes ses formes, pour certaines autres industries, comme le cinéma, le spectacle vivant ou le livre, des usages essentiels sont encore dépendants du déplacement des membres du public dans des lieux physiques pour l'accès aux œuvres, comme la projection en salle, les représentations vivantes dans des lieux publics ou la vente de livres en librairie. La numérisation affecte toutefois également ces modes d'accès à travers la numérisation des outils de projection en salle, le développement du « live streaming » ou encore le livre électronique ou la vente de livres physiques sur les plateformes de vente et/ou de référencement à distance.

La numérisation n'est ainsi pas nécessairement synonyme de dématérialisation totale ou de disparition des modes physiques d'accès aux œuvres. En revanche, elle se conçoit de façon systémique comme une transformation des processus industriels afin de tirer parti des données, des réseaux et des puissances de calcul disponibles pour leur création, leur exploitation, leur découverte. Elle bouleverse la relation avec le public.

Les modèles d'affaires des industries culturelles et créatives se diversifient grâce aux possibilités offertes par la combinaison du numérique et des communications électroniques. Le développement de l'intelligence artificielle, des métavers, des NFT, tout comme celui de la réalité immersive, offrent de nouvelles perspectives dont la vitesse de réalisation oblige à une innovation constante mettant au défi les industries culturelles de s'adapter en permanence et, mieux encore, d'anticiper sur les bouleversements à venir.

Parmi les grands modèles économiques disponibles, tous fondés sur les droits de propriété intellectuelle et les services proposés au public, se trouvent :

- la vente de produits comme des supports matériels (livres, CD ou vinyles, DVD ou Blu-Rays, etc.) ou virtuels (NFT, …) ;
- la vente de droits d'accès à un spectacle, une œuvre, un objet patrimonial, etc…, dans un lieu physique ou à distance, généralement fondés sur une « billetterie » de plus en plus numérisée et riche de données d'usage ;
- les revenus liés à des abonnements à des services d'agrégation de contenus, linéaires ou non linéaires, de façon centralisée ou non ;
- les revenus issus de l'exposition du public à de la publicité, dont le ciblage est un enjeu croissant, qu'il s'agisse de services linéaires (radio, télévision, …) ou non, parfois combinés avec le modèle par abonnement (freemium).

2. Plateformisation de l'accès aux œuvres

La « plateformisation » de l'accès aux œuvres consiste en l'agrégation, par un tiers, de créations sous forme de fichiers numériques qui sont ensuite mis à disposition du public par son intermédiaire, via ses outils d'indexation et de recommandation.

Le développement des réseaux de communication, leur démocratisation et l'émergence de nouvelles technologies et de nouveaux matériels ont permis l'apparition de multiples nouveaux usages, pilotés en grande partie par des acteurs mondiaux disposant de ressources considérables et représentant de véritables « gardiens d'accès » aux œuvres. Ces bouleversements ne doivent pas laisser échapper la valeur des œuvres à leurs créateurs et leurs partenaires économiques. Des équilibres économiques adaptés doivent ainsi être constamment recherchés afin de répartir équitablement la valeur entre les parties prenantes et les nouveaux intermédiaires sur les réseaux de communications électroniques.

Une particularité essentielle de cette « plateformisation » est l'automatisation d'une partie des modalités d'accès aux œuvres et donc de la capacité du public à les connaître, à travers les algorithmes, dont les données numériques représentent le « carburant ». En tant que biens d'expérience, les œuvres doivent être accessibles sous forme d'extraits ou, en tout cas, d'informations référencées, pour que le public puisse s'y intéresser. Tout goulet d'étranglement de l'accès aux œuvres a des effets de distorsion considérable si des alternatives ne sont pas disponibles avec la même facilité.

C'est la raison pour laquelle la valeur des usages tend à rester chez les intermédiaires qui, par leur puissance, peuvent créer un déséquilibre en leur faveur, ce qui oblige à une régulation.

La législation a pour vocation de rétablir les équilibres nécessaires à une dynamique de circulation des œuvres compatible avec celle de la diversité de leur création et de leur distribution.

L'un des enjeux est la disponibilité de données sur les œuvres pour les opérateurs qui les mettent à disposition du public ou les référencent afin de favoriser l'accès du public à celles-ci, ainsi que des données sur les usages de ces œuvres au bénéfice des filières de création pour guider leur activité et développer leur dynamisme.

3. Une stratégie Européenne

La stratégie européenne en matière de données doit permettre à l'Union européenne de devenir un acteur de premier plan dans une société axée sur les données. La création d'un marché unique des données permettra de fluidifier la circulation de ces dernières au sein de l'UE et entre les secteurs, dans l'intérêt des entreprises, des créateurs, des administrations publiques et du public.

Les données sont au cœur de la transformation numérique. Elles façonnent notre façon de produire, de consommer et de vivre. L'accès à un volume sans cesse croissant de données et la capacité à les utiliser sont essentiels pour l'innovation et la croissance. L'innovation fondée sur les données peut apporter des avantages considérables concrets aux citoyens — par exemple sous la forme de l'amélioration de la mobilité — et à l'économie européenne, tels qu'une plus grande efficacité du processus d'élaboration des politiques et l'amélioration des services publics.

Plusieurs textes en préparation ou récemment adoptés visent ainsi à établir un écosystème ouvert, sécurisé, de confiance et transparent pour une meilleure circulation des données à travers l'Europe et le développement de nouveaux modèles d'affaires. L'ambition est de créer un marché unique des données conformes aux valeurs de l'Union Européenne.

- Data Governance Act
- Digital Market Act (DMA)

- Digital Services Act (DSA)
- Data Act (DA)

La régulation européenne, à travers les projets de règlements DSA et DMA, a un rôle considérable à jouer pour insérer les nouveaux acteurs dans l'écosystème européen et s'assurer d'une concurrence plus équitable avec les opérateurs classiques notamment les médias.

L'acte sur la gouvernance des données mettra en place des mécanismes solides pour faciliter la réutilisation de certaines catégories de données protégées du secteur public, renforcer la confiance dans les services d'intermédiation de données et favoriser l'altruisme en matière de données dans l'ensemble de l'UE.

Il s'agit d'une composante importante de la stratégie européenne pour les données, qui vise à renforcer l'économie fondée sur les données, à accroître la richesse et le bien-être et à conférer à l'Europe un avantage concurrentiel au profit de ses citoyens et de ses entreprises.

L'acte sur la gouvernance des données crée un cadre visant à promouvoir un nouveau modèle commercial - les services d'intermédiation de données - qui fournira un environnement sûr dans lequel les entreprises ou les particuliers pourront partager des données.

Pour les entreprises, ces services peuvent prendre la forme de plateformes numériques qui permettront le partage volontaire de données entre entreprises ou faciliteront le respect des obligations en matière de partage de données fixées par la loi. Grâce à ces services, les entreprises pourront partager leurs données sans craindre qu'elles soient utilisées de manière abusive ni risquer de perdre leur avantage concurrentiel.

4. Un soutien français

Dans la droite ligne de la stratégie numérique de l'Etat, le Ministère de la Culture a publié, le 15 septembre 2021, une feuille de route « Données et contenus culturels » qui vise, en particulier, à renforcer l'accessibilité et la découvrabilité des contenus culturels grâce à un enrichissement des métadonnées et à un partage facilité de ces données.

Par ailleurs, plusieurs initiatives de financements fléchés par les pouvoirs publics concernent les industries culturelles et créatives, en particulier de la part de la Banque des territoires (CDC) et de la BPI.

En outre, le plan de relance 2030 contient un volet substantiel portant sur les ICC.

S'agissant plus spécifiquement des données numériques, le Gouvernement Français soutient l'initiative Gaia-X depuis son démarrage et encourage le développement d'une infrastructure et d'un écosystème Cloud européen à même d'assurer la libre circulation des données et de créer de la valeur en Europe.

Dans tous les cas, l'innovation est un axe central pour que les ICC bénéficient des appuis publics liés au numérique.

LES DIFFERENTS TYPES DE DONNEES

1. Contenus

Les contenus numérisés, les jumeaux numériques ou les contenus nativement numériques constituent un premier type de données dont il faut assurer la circulation tout en en protégeant l'accès. Ces données possèdent une valeur intrinsèque en ce qu'elles constituent les œuvres elles-mêmes.

La question de la numérisation des œuvres devient également cruciale. Ces œuvres numérisées devront être stockées, identifiées et protégées. La standardisation des formats, la description des contenus, la réversibilité et l'entretien des données stockées sont des problématiques à traiter dans le cadre de ce groupe de travail.

2. Métadonnées

Les métadonnées sont des données associées à une autre donnée. Elles sont indispensables afin d'assurer la découvrabilité des données auxquelles elles sont associées ainsi que la rémunération des ayants-droits. Leur qualité est ainsi primordiale. Les métadonnées couvrent de nombreux champs [3] :

Catégorie	Description	Exemples
Identifiants uniques	Chaînes de caractères alphanumériques qui ont pour fonction d'identifier de manière stable un document, une ressource ou une entité, quelle que soit sa nature. En principe, un identifiant devrait être unique pour chaque ressource.	ISBN (livre) ISAN (film) ISRC (enregistrement musical) ISWC (œuvre musicale) IPN (*International Performer Number*) ISNI (*International Standard Name Identifier* pour auteurs, compositeurs, interprètes, producteurs, etc.), etc.
Descriptives	Décrivent précisément et de manière objective le contenu.	Nom de l'auteur Titre de l'œuvre Matériaux Date de création Date de naissance de l'auteur, etc.
Administratives	Fournissent des renseignements sur la manière dont les fichiers ont été créés, enregistrés, modifiés.	Date de création du fichier Nom de l'entreprise créatrice du fichier Format utilisé Identification du document source ayant été numérisé, etc.

[3] Inspiré de « *Etat des lieux sur les métadonnées relatives aux contenus culturels* » de l'Observatoire de la culture et des communications du Québec (2017).

Juridiques (ou de propriété intellectuelle)	Indiquent si l'œuvre est sous copyright, identifient les ayants-droits et la société de gestion, indiquent si l'œuvre est assujettie à une licence d'utilisation et déterminent les mentions de source devant apparaître avec l'œuvre.	Nom du titulaire ou du gestionnaire du copyright Identifiant du titulaire Libellé de la mention de source devant figurer avec le titre de l'œuvre
D'enrichissement	Donnent des renseignements complémentaires et parfois subjectifs sur un contenu. Le périmètre des métadonnées d'enrichissement est vaste. Elles peuvent être utilisées comme une valeur ajoutée par les fournisseurs de contenus culturels numériques pour se démarquer de la concurrence. Ces données d'enrichissement peuvent être produites par des algorithmes d'intelligence artificielle.	Biographie de l'artiste Photographies Notes Paroles de chansons Adresses des réseaux sociaux de l'artiste Image de la couverture / de l'affiche Texte d'une critique, etc.

Afin de faciliter le développement d'applications et l'échange, le partage ou la valorisation de ces données, il devient nécessaire, pour le secteur, d'utiliser des ontologies communes. A ce titre, l'Union Européenne de Radio-télévision (UER) a développé et maintient les standards EBUCore et EBU CCDM qui visent à faciliter la circulation des contenus entre producteurs et diffuseurs. Cette recommandation est utilisable sans contrepartie financière. La généralisation de l'usage de standards partagés permettrait une simplification des processus et un gain d'efficacité certain.

De même, dans une démarche initiale d'harmonisation de ses différentes collections, l'INA a développé et déployé, sur son périmètre de gestion (dépôt légal et dévolution réglementaire de l'audiovisuel public), un modèle de données librement inspiré de FRBR / CIDOC CRM. Ce modèle se caractérise par son évolutivité et sa flexibilité, par l'importance accordée aux vocabulaires contrôlés, articulée avec des entités très génériques. Il couvre ainsi tout type d'objet, audiovisuel, écrit, photographique ou web, diffusé ou non, dont il décrit à la fois la nature intellectuelle (contenu) et la matérialisation, physique ou numérique (contenant). Ce principe d'agnosticité et d'adaptabilité à tout contexte de gestion rend ce modèle déployable dans tout environnement de gestion de contenus, de toutes natures. C'est pourquoi l'INA propose d'ouvrir largement ce modèle, pour accompagner sa diffusion et faciliter son appropriation et son utilisation par tous les acteurs intéressés.

3. Les données de consommation

Les données de consommation regroupent à la fois les données d'usage générées de manière automatique lors de la consommation en ligne d'un contenu culturel particulier et les données (ou mesures) d'audiences ou de ventes collectées de manière plus classique.

Ces données sont indispensables à la compréhension des usages et des marchés et à l'adaptation des offres à la demande. Elles sont également nécessaires à la valorisation des espaces publicitaires qui constituent une source de revenus essentielle pour certains acteurs des ICC.

Le DMA prévoit, en particulier, d'imposer un accès, pour les fournisseurs de contenus, à certaines données de consommation collectées par les grandes plateformes dominantes (« contrôleurs d'accès »). Le Data Act (proposition législative publiée en février 2022) a pour objectif de stimuler le partage de données de consommation générées via l'usage de certains objets connectés, ainsi que d'encourager le développement de standards pour la mise en place d' « espaces de données ».

4. Les données personnelles

Les données personnelles sont des données se rapportant à une personne physique pouvant être identifiée. Il n'est pas interdit d'utiliser ce type de données mais des précautions particulières s'imposent. Le traitement des données personnelles est fortement régulé. L'Union Européen a élaboré un règlement spécifique traitant de la particularité de ces données : le RGPD (Règlement général de protection des données) du 27 avril 2016.

En ce qui concerne le secteur des ICC, ces données entrent en jeu notamment dans la gestion des abonnements à une plateforme ou dans le ciblage publicitaire et la gestion des consentements.

5. Les autres données

Chaque processus industriel génère son propre flux de données : des données logistiques, les logs des machines, la consommation énergétique, les rapports d'exécution, etc. Ces données peuvent également gagner à être mieux exploitées.

ASSURER LA DECOUVRABILITE GRACE AUX METADONNEES

Face à l'explosion du nombre de contenus culturels disponibles et accessibles, notamment via internet, la visibilité et la capacité à mettre en relation une œuvre avec un public ad hoc deviennent des enjeux majeurs.

Le World Wide Web est lui-même un espace d'exploration quasi-infini dans lequel le public évolue grâce à des outils tels que les moteurs de recherche qui utilisent des algorithmes aboutissant à des référencements naturels (SEO) ou payants (SEM) des contenus.

Rendre visible une œuvre revient donc à lui associer les métadonnées qui permettront aux algorithmes de détecter sa compatibilité avec les goûts des utilisateurs ou de la leur proposer comme découverte. Dans un certain nombre de secteurs des ICC, ces métadonnées restent souvent parcellaires et incompatibles avec les métadonnées des autres secteurs faute de standards communs. Il en résulte une sous-utilisation ou une utilisation réduite aux plateformes compatibles. Il existe, toutefois, des mappings utiles pour passer d'un système à un autre.

Le premier défi consiste, ainsi, à généraliser l'utilisation de l'un ou l'autre des systèmes existants et à amorcer le mouvement vers une forme de standardisation. Le second défi sera de s'assurer de la qualité des métadonnées associées aux œuvres. Une forme de gouvernance mériterait vraisemblablement d'être mise en place. Une telle gouvernance permettrait, notamment, d'instaurer des guides de bonnes pratiques en matière de métadonnées relatives aux contenus culturels.

AFFINER LES PROPOSITIONS GRACE AUX DONNEES D'USAGE

Quel que soit le modèle d'affaires suivi, aujourd'hui, l'accès aux données d'usage des consommateurs, aux données liées à leurs habitudes de navigation sur internet ou à leurs goûts, aux données

d'audience, est primordial. Ces données permettent notamment d'adapter l'offre de contenus, leurs modes de promotion et/ou d'exploitation, de développer de nouvelles offres ou encore d'optimiser la communication publicitaire et d'en mesurer les effets.

Lorsque les acteurs culturels obtiennent l'accès à des données sur l'exploitation de leurs contenus (soit de la part des plateformes et des distributeurs de services, soit à partir de leurs propres canaux de distribution comme leur site Web), ils n'ont pas toujours les moyens (techniques, humains ou financiers) de les harmoniser, de les analyser et d'en retirer toute l'information dont ils pourraient profiter, notamment dans l'objectif d'établir une stratégie optimale de découvrabilité de leur contenu. Ceci est particulièrement le cas pour les plus petites structures culturelles.

Par ailleurs, chaque acteur de la chaîne de valeur collecte aujourd'hui ces données pour son compte et les traite de son côté, ce qui ne lui permet que d'avoir une vue très parcellaire des goûts et des besoins des utilisateurs.

L'ensemble du secteur culturel bénéficierait d'une mise en commun, la plus exhaustive possible, de l'ensemble de ces données. Les analyses qui en découleraient seraient plus fines et permettraient un meilleur ciblage des offres culturelles. Le public aussi en bénéficierait, grâce à une meilleure adéquation des propositions avec ses goûts, ses centres d'intérêt ou ses perspectives de découverte.

GARANTIR LA REMUNERATION DES AYANTS DROIT

Au regard de la masse et de la diversité des modes d'exploitation des œuvres, de l'importance des flux, de la variété des ayants droit, des contrats et des modèles économiques, de la nature mondiale des exploitations devant respecter des législations différentes et de la vitesse de réaction nécessaire pour accompagner les évolutions techniques et des usages, la capacité de traitement des données associées aux œuvres est capitale pour assurer les gestion des droits et des rémunérations des ayants droit et de leurs partenaires économiques. Il est évidemment essentiel que l'ensemble des données relatives à une œuvre puisse être traité par les mêmes processus afin d'assurer la juste rémunération des ayants droit. En particulier, les données de consommation de contenus sur de grandes plateformes devraient être rendues accessibles dans des conditions raisonnables.

Le travail sur les formats de données et les moyens de les récolter et de les analyser est au cœur de la capacité des industries culturelles et créatives à s'inscrire pleinement dans les mutations en cours.

Il est également à noter que de nombreux acteurs récupèrent et traitent les mêmes données avec des objectifs légèrement différents mais que les opportunités de mutualisation et de simplification des flux de données sont très nombreuses.

RATIONALISER LES RESSOURCES

Pour que les différents domaines de la culture puissent tirer profit de leurs données, les entreprises et les organisations auraient avantage à se concerter et à canaliser leurs efforts, en tenant compte de l'ampleur de la tâche pour les plus petites organisations des différentes filières, qui n'ont pas les mêmes moyens que les grandes pour investir dans les infrastructures technologiques ou dans la normalisation de leurs métadonnées.

Ces défis ne pourront pas être relevés sans l'adoption, par les différents intervenants des domaines culturels, d'un langage commun (ou, à tout le moins, de langages harmonisés) et d'une vision globale, où les besoins de leur propre organisation ne sont plus les seuls déterminants de l'importance d'un champ de métadonnées. Pour ce faire, il importe que tous soient sensibilisés au potentiel des données

pour le plein déploiement du secteur de la culture et que chacun en comprenne l'impact et l'intérêt pour le développement de son activité.

Il est possible d'établir des liens entre les divers secteurs ou domaines culturels, des ponts qui permettraient une convergence minimale. Par exemple, l'adoption généralisée d'un même système de nommage dans les différents secteurs de la culture faciliterait le rapprochement de sources d'origines variées. Que le contenu culturel soit un spectacle, un film, une émission de télévision, un livre, un objet de musée, un enregistrement musical ou un élément patrimonial, un système unique de nommage permettrait d'identifier de manière unique les entités relatives à ce contenu (nom de l'auteur, du lieu, de la maison de production, etc.). Il permettrait de distinguer des noms qui, autrement, pourraient être confondus et de lier entre elles les données concernant un nom en particulier.

INTERET DE L'INITIATIVE GAIA-X

L'initiative Gaia-X entend développer un écosystème ouvert, sécurisé et transparent permettant à chaque acteur de l'économie d'accéder à des ressources de diverses natures dans des conditions transparentes et non discriminatoires. Ces ressources peuvent être des capacités de calcul, des services de traitement de données, des accès à des données, des services d'authentification, etc…

Grâce à une labellisation des services fondée sur des « Policy Rules » conformes aux valeurs européennes, Gaia-X répond à l'ambition de l'Europe d'établir un marché unique des données (et traitements associés) dans une infrastructure Cloud sécurisée, de confiance et interopérable.

Gaia-X entend également soutenir la création d'espaces de données sectoriels. Le concept d'espace de données fait référence à un écosystème d'échange, de partage et de mise à disposition, contre rémunération ou non, de données entre partenaires de confiance. Il ne s'agit pas de copier ou de rapatrier des données de manière centralisée mais d'assurer à chaque détenteur de données un contrôle total sur les conditions d'accès à ses données : qui, quand, dans quelles conditions. En donnant de la visibilité sur les données existantes, en facilitant l'accès aux différentes sources de données, les espaces de données sectoriels ont vocation à faire émerger de nouveaux services et à créer de la valeur, pour les Européens, en Europe et au-delà. L'association Gaia-X compte développer les briques architecturales et/ou logicielles permettant d'assurer des passerelles sécurisées entre espaces de données.

4 CAS D'USAGE

Les cas d'usage décrits dans cette partie ont vocation à donner lieu aux lancements de projets de développement.

La réalisation de ces projets constituera le socle de la création d'un espace de données sectoriel pour les Industries Culturelles et Créatives, regroupant la partie Média et la partie Patrimoine Culturel.

Les parties intéressées par l'une ou l'autre des initiatives décrites dans la section qui suit sont invitées à contacter le hub France.

CARACTERISTIQUES DES PROJETS GAIA-X

Les projets soutenus par Gaia-X pourront bénéficier des aides nationales (voire Européennes).

Afin d'obtenir un soutien de Gaia-X, ceux-ci doivent répondre aux objectifs de l'association :

- Pouvoir passer à une échelle européenne ;
- Participer à la création d'un écosystème ouvert, non propriétaire et sans approche monopolistique ;
- Adopter les règles et caractéristiques issus des policy rules, de l'architecture et des labels et pouvoir être associé, à termes, à un espace de données.

ORGANISATION DES CAS D'USAGE SELON LA CHAINE DE VALEUR

Comme indiqué dans les chapitres précédents, le secteur des Industries Culturelles et Créatives regroupe une diversité d'acteurs ayant tous un rôle dans la création de valeur du secteur. Leurs activités se situent dans :

- La création ;
- La production ;
- La fabrication ;
- L'édition ;
- La distribution ;
- La vente.

Les cas d'usage décrits ci-dessous impliquent un ou plusieurs niveaux de cette chaîne de valeur.

CAS D'USAGE POUR LA CREATION, LA PRODUCTION OU LA FABRICATION

1. Dépôt et référentiel des œuvres - Mise en valeur de la création

1.1 Création d'un référentiel de Jumeaux Numériques (Projet IWA - Interoperability Within the Ark)

A l'initiative de :

Monark, agence de création de jumeaux numériques du patrimoine architectural, de sites naturels sensibles et d'objets remarquables.

Participants potentiels :

Garou, plateforme multi user XR pour créateurs et entreprises, métavers

La problématique :

La numérisation du patrimoine immobilier, mobilier et d'objets remarquables à des fins de conservation, de transmission, de valorisation et de communication prend tout son sens dans notre société numérique contemporaine. La création de jumeaux numériques (JN) de notre patrimoine architectural ou pictural, de certains lieux typiques, détenus par les propriétaires des œuvres, pourrait constituer une nouvelle source de revenus pour les propriétaires de biens et d'inspiration pour la filière des Industries Culturelles et Créatives.

A ce jour, il n'existe aucune plateforme professionnelle pour permettre aux acteurs de la filière ICC d'exploiter pleinement les ressources liées aux jumeaux numériques. Les flous juridiques, le manque de partage, d'interopérabilité, de dialogue ou l'absence de visibilité entraînent souvent une nouvelle numérisation des œuvres.

Les grands hyperscalers investissent massivement dans les nouveaux supports de diffusion comme la réalité virtuelle (cf. Facebook avec Meta), la réalité augmentée ou le mapping vidéo (projection architecturale). Afin d'éviter la captation de valeur liée aux jumeaux numériques par ces acteurs, le projet IWA vise à l'émergence d'un acteur français ou européen qui organiserait la collecte, le stockage, la description, la mise en visibilité et la commercialisation des données de jumeaux numériques suivant un cahier des charges précis et transparent.

Le projet ambitionne également d'éveiller les consciences des acteurs ICC sur l'importance de la transposition de leurs biens matériels dans le monde virtuel et les bénéfices qu'ils pourraient en tirer.

Solution apportée :

IWA est une base de données de jumeaux numériques remarquables pouvant être exploités comme décors virtuels par les professionnels du cinéma, du gaming et de l'art contemporain…

Le projet IWA propose de :

- Monter une plateforme de mise à disposition de jumeaux numériques ;
- Créer des jumeaux numériques ;
- Sécuriser l'utilisation de ces jumeaux grâce au fingerprinting et aux NFT dans une blockchain ;
- Commercialiser (sous licence) l'utilisation de jumeaux numériques en rémunérant le propriétaire du bien matériel.

Par le biais d'une plateforme utilisant la technologie WebGL (GPU), tous les acteurs de la filière des industries culturelles et créatives pourront consulter les jumeaux numériques disponibles.

Bénéfices attendus :

Consolider le patrimoine immatériel des industries culturelles et créatives et le valoriser à l'international.

Monark permet aux institutionnels et aux privés d'investir dans un jumeau numérique et de le monétiser à travers des scenarii ouverts sur de nombreuses opportunités. IWA constitue une alternative aux GAFA pour garder la main sur le contenu et en conserver la valeur en Europe.

Difficultés/barrières identifiées :

Plusieurs éléments sont de nature à complexifier la réalisation du projet :

- L'absence de standardisation des formats de jumeaux numériques ;
- Les propriétaires et institutionnels sont peu familiers de ce nouveau marché ;
- L'absence de plateforme commune pour la filière ICC (Industries Culturelles et Créatives) ;

- L'hégémonie des GAFA (Google, Apple, Facebook, Amazon…).

Partenaires du projet :

- Éditeur de logiciel, spécialisé en captation et restitution audio 3D
- Fournisseur de solutions de Blockchain évolutives, interopérables, souveraines et intégrées
- Société de production d'animation 3D
- Instituts de recherche (éventuellement)

Principales technologies Gaia-X en jeu :

Plateforme data : partage, échange des données entre plusieurs acteurs de la filière ICC

Sécurité et confidentialité : Authentification, gestion de fédération d'identité, gestion, sécurisation des données gestion du consentement et contrôle de l'utilisation.

Conformité : portabilité des données, gestion des droits et certification.

1.2 Création/exposition d'une plateforme de description des œuvres

A l'initiative de :

Cap Digital – pôle de compétitivité Paris/Région (1.000 adhérents)

Participants :

Consortium en cours de constitution.

Problématique :

Ce projet de plateforme de description des œuvres consiste en une place de marché permettant l'identification et la description de l'ensemble des œuvres produites et distribuées en Europe, pour en faciliter l'exposition et la distribution. Cette place de marché pour la circulation des données descriptives des œuvres répondrait à trois besoins :

1. donner un accès aux données ouvertes décrivant les œuvres afin de faire circuler l'essentiel des informations descriptives propices à leur exploitation/dissémination. Ces données seraient principalement fournies par les producteurs et leurs partenaires, afin de constituer, dans le domaine de l'audiovisuel et du cinéma, une alternative à la plateforme américaine IMdB ;
2. permettre la circulation sécurisée de données plus stratégiques, impliquant cette fois l'interconnexion de serveurs privés de membres industriels (producteurs, post-producteurs, laboratoires, ayant-droits…) visant principalement à recenser, exposer et vendre des supports numériques (masters, PAD, sous-titres, doublages…) et/ou des droits. Ce besoin pourrait s'appuyer techniquement sur le projet national collaboratif TAMIS qui est déjà engagé (voir ci-après) et regroupe tous les grands acteurs français (Ficam, CST, laboratoires…) ;
3. proposer des connecteurs logiciels (API) associés à un modèle économique adapté à la création de valeur par des startups.

1.3 Création d'un catalogue d'actifs numériques

A l'initiative de :

CST

Participants potentiels :

Problématique :

De plus en plus de productions utilisent aujourd'hui des actifs numériques (ou assets numériques), monument, paysage, bâtiments, etc. Ces actifs sont généralement créés à la demande pour une production donnée et se retrouvent archivés à l'issue de la production. Certains actifs créés pourraient toutefois être réutilisés dans d'autres productions pour un coût bien moindre que la création d'un nouvel actif numérique. Il faudrait alors savoir que ces actifs existent et comment y accéder.

1.4 Création et exposition/valorisation d'un catalogue de décors

A l'initiative de :

CST

Participants potentiels :

Problématique :

Les productions audiovisuelles, pour la télévision ou le cinéma, utilisent de nombreux éléments de décors bien souvent fabriqués à la demande. La réutilisation de ces décors est encouragée dans le cadre de la démarche écoresponsable Ecoprod. Toutefois, la réutilisation reste faible car aucune base, à ce jour, ne répertorie l'ensemble des décors fabriqués et disponibles avec les éléments de description permettant un choix éclairé (matière, couleur, lieu de stockage, état, disponibilité, etc.).

Bénéfices attendus :

Un tel référencement, avec des métadonnées standardisées permettrait d'économiser des matières premières et de rentabiliser la construction des décors grâce à une valorisation renforcée. Une base transeuropéenne permettrait de valoriser les savoir-faire locaux et augmenterait la base de décors disponibles.

1.5 Dépôt des œuvres et certificat de propriété sécurisé dans la blockchain

A l'initiative de :

Participants potentiels :

France Télévisions

Problématique :

Toute œuvre numérisée devient facilement reproductible. Les technologies blockchain et NFT (Non Fungible Token) doivent permettre de garantir la propriété intellectuelle de ces œuvres aux ayants droit et de s'assurer de leur authenticité. Elles facilitent ainsi la lutte contre le piratage, en particulier grâce à la traçabilité de la blockchain.

1.6 Numérisation et pérennité des contenus patrimoniaux

A l'initiative de :

Participants :

Problématique :

La généralisation des accès aux réseaux de télécommunication facilite l'accès aux œuvres numériques. En revanche, le patrimoine audiovisuel (principalement le cinéma) reste en grande partie stocké sur des supports physiques. L'accès à ces œuvres reste difficile.

Un grand projet de numérisation de ces œuvres permettrait sa mise en valeur et sa diffusion à un large public. Cette numérisation s'accompagne évidemment de la création de toutes les métadonnées nécessaires au référencement et à la découvrabilité des œuvres.

Il s'agirait de réaliser :

- Numérisation des œuvres
- Catalogage
- Enrichissement des métadonnées
- Entretien des formats
- Espace de stockage fiable
- Données associées assurant la lisibilité des contenus
- Accessibilité des archives

2. Faciliter et documenter la fabrication

2.1 Portail de partage de données avec les producteurs de contenus

A l'initiative de :

France Télévisions avec l'appui de Cap Digital

Participants :

Producteurs des œuvres éditées et diffusées

La problématique :

L'exploitation des œuvres numériques nécessite un volume de métadonnées important afin de référencer l'œuvre le plus justement possible. Ce référencement permet ensuite une meilleure découvrabilité, à la fois, directement par une exploration manuelle de la plateforme et surtout par le biais de la recommandation directe ou indirecte ainsi que par la personnalisation.

Aujourd'hui, ces métadonnées sont complexes à rassembler et obtenir pour les diffuseurs. Les producteurs et les ayants droit n'ont pas de processus systématiques pour créer, stocker et distribuer ces métadonnées.

Le portail producteur se veut être un lieu où ces métadonnées peuvent être livrées en une fois ou de façon incrémentale afin d'accompagner le processus de production. Il permet également aux équipes du diffuseur d'évaluer la qualité et la pertinence des métadonnées fournies afin de faire corriger ou compléter ce qui doit l'être. Une fois validées, ces données sont reversées au data product Programme de France Télévisions pour utilisation sur chacun des vecteurs de diffusion.

2.2 Echange des métadonnées de fabrication (Projet TAMIS)

A l'initiative de :

CST

Participants :

Les participants opérationnels actuels du projet :

- La CST, qui porte le projet et ses prestataires,
- MediaADN
- Lum::Invent (ex Media-IO)
- Startin'blox

Avec l'appui de :

- L'agence ISAN-IA
- La Procirep
- Cap Digital – pôle de compétitivité Paris - Région
- Titra

La problématique :

Dans un contexte de multiplication des livraisons dématérialisées des œuvres audiovisuelles, le nombre et la variété des fichiers médias associés se multiplient demandant des efforts croissants de référencement et de suivi de la part des différents acteurs du secteur.

L'immatriculation ISAN est demandée par le CNC en France depuis 2017, mais son utilisation réelle reste timide dans les livrables audiovisuels. C'est le cas également pour d'autres identifiants, comme l'EIDR demandés par les plateformes OTT.

Les validations des versions se font de manière artisanale (mail, SMS, etc.), générant incompréhensions et confusions dans la commande et la livraison. L'immatriculation de variantes et manifestations est centralisée dans les mains de l'ayant droit, complexifiant la déclaration de nouveaux livrables par les autres acteurs de la chaîne.

Solution apportée :

Le projet TAMIS vise à favoriser l'échange de métadonnées de fabrication, au-delà des identifiants, afin d'optimiser les échanges entre acteurs de la chaîne de fabrication.

La solution envisagée repose sur l'architecture Solid, qui est basée sur une pile de technologies Web existantes. Cette technologie est en particulier promue par Tim Berners Lee pour revenir à un web des données réellement décentralisé. La pile technologique Solid repose sur des technologies web établies, une modélisation sémantique, telle que le RDF, l'utilisation d'API en restful, une authentification par WebId basée sur OpenID, pour n'en citer que quelques-unes.

Pour la modélisation sémantique, le projet envisage l'utilisation de EBU-CCDM, une généralisation de EBUCore proposée par l'UER (EBU).

Le projet TAMIS-1 prévoit une évaluation des options envisagées et une modélisation fine des cas d'usages.

L'objectif est de poursuivre sur un projet TAMIS-2 qui serait une preuve de concept (PoC) regroupant des représentants typiques de la chaîne de valeur du secteur, avec un laboratoire, une entité d'immatriculation, un diffuseur et un ayant droit.

Les livrables seraient, outre une documentation détaillée, un SDK pour faciliter le développement de connecteurs, et une brique "prête à l'emploi" facile à mettre en œuvre avec une interface graphique permettant aux acteurs faiblement informatisés de faire de la saisie manuelle pour entrer dans la boucle TAMIS.

Bénéfices attendus :

La bonne circulation des identifiants est un changement systémique. L'adoption massive par la filière audiovisuelle de cette solution distribuée permettra d'optimiser les échanges de ces identifiants ainsi que des métadonnées qui doivent les accompagner au sein de la chaîne de fabrication des œuvres audiovisuelles. Tamis permettra ainsi de :

- fluidifier les processus de production d'une œuvre :
  - faciliter la collaboration entre acteurs de la chaîne en réduisant l'effort d'adaptation des systèmes d'information ;
  - engendrer des gains de productivité par une automatisation fiable des différents processus et en réduisant les erreurs de saisies ;
  - optimiser l'organisation des plannings des différentes étapes de fabrication d'une œuvre ;
  - réduire le "Time to Market".
- contribuer au suivi de la commercialisation et de l'exploitation des œuvres audiovisuelles produites ;
- favoriser la circulation des informations liées à la rémunération des ayants droit ;
- mettre en valeur les compétences du marché en améliorant la visibilité des acteurs ;
- contribuer à la sobriété numérique.

Difficultés/barrières identifiées :

- Intéresser les producteurs de cinéma à un projet technique ;
- Mobiliser des investissements des entreprises du secteur pour financer un connecteur ;
- Agencer les différents vocabulaires contrôlés employés dans la profession (EBUCore ou EBU-CCDM, ou l'ontologie MDDF proposée par MovieLabs, le laboratoire des studios hollywoodiens).

Principales technologies Gaia-X en jeu :

L'architecture de la fédération de services Gaia-X propose des services d'identité et de confiance, qui sont à priori résolus par l'architecture Solid retenue, mais une étude plus approfondie pourra faire apparaître des convergences technologiques et architecturales.

En revanche, le service de catalogage doit permettre d'indexer et d'exposer des points d'entrées au service.

A ce stade, cette approche est plutôt envisagée techniquement comme un connecteur Solid vers les services fédérés de Gaia-X.

2.3 Service d'aide au calcul de l'empreinte Carbone des ICC

A l'initiative de :

Cap Digital – pôle de compétitivité Paris/Région (1.000 adhérents)

Participants :

Consortium en cours de constitution

Problématique :

La mesure précise de l'empreinte Carbone des chaînes d'édition, production et distribution des œuvres va progressivement devenir nécessaire à l'obtention d'un financement. Le projet vise à réunir les acteurs de la chaîne de valeur qui ont déjà réalisé un calcul de l'empreinte carbone liée à l'ensemble de leurs activités. Plusieurs objectifs sont définis :

- créer un référentiel des mesures par type d'activité ;
- standardiser les méthodologies de mesure ;
- mettre en œuvre un service européen d'assistance au calcul de ces empreintes, sous la forme d'une plateforme logicielle.

2.4 Lutte contre les fake news

A l'initiative de :

France Télévisions

Participants :

France Télévisions, l'INA

Problématique :

De plus en plus d'images d'amateurs, postées notamment sur les réseaux sociaux, montrent une partie d'une actualité.

Pour le public, mais aussi pour les sociétés qui produisent de l'information, l'enjeu est de distinguer les images réelles des images fausses ou hors-contextes afin de démêler le vrai du faux.

Solution apportée :

Créer une plateforme qui permettrait à tous les factcheckeurs volontaires de marquer les images (fixes ou vidéos) vérifiées et de stocker ses empreintes.

Cette plateforme serait un outil pour les citoyens et/ou les journalistes qui leur permettrait d'interroger la banque d'empreinte afin de savoir si l'image a déjà été utilisée ou non, si elle est falsifiée ou hors-contexte.

Un POC a déjà été réalisé chez France Télévisions

Bénéfices attendus :

Permettre aux citoyens d'avoir un outil facile à utiliser qui leur éviterait de propager de fausses informations.

Difficultés/barrières identifiées :

Embarquer le plus de factchekeurs possibles pour que la base de données soit suffisamment efficace. Coût des empreintes. Maintenance de la plateforme qui serait publique et gratuite.

Principales technologies Gaia-X en jeu :

Plateforme data : partage, échange des données entre plusieurs acteurs de la filière ICC.

Sécurité et confidentialité : Authentification, gestion de fédération d'identité, gestion, sécurisation des données gestion du consentement et contrôle de l'utilisation.

Conformité : portabilité des données, gestion des droits et certification.

2.5 Plateforme de services fondés sur l'intelligence artificielle pour enrichir les métadonnées de programme

A l'initiative de :

France Télévisions, INA

Participants potentiels :

lum::invent (ex Media-IO)

Audiovisuel Public

Problématiques :

Sur la base d'un projet open-source initié en 2018 : https://media-cloud.ai, l'objectif consiste à proposer des services de traitement des données multimédia (images, images animées, son, sous-titres, données interactives…) exploitables sur une instance de plateforme open-source existante mise à disposition sous la forme d'un service Cloud, à l'attention des utilisateurs dépourvu des compétences d'ingénierie logicielle nécessaires à sa mise en œuvre.

Au-delà des traitements classiques des fichiers et flux audiovisuels, cette plateforme permet l'orchestration de traitements basés sur l'Intelligence Artificielle à partir de modèles explicables.

Cette plateforme pourrait être développée en commun entre différents acteurs du secteur et notamment les acteurs publics en France et en Europe.

3. Intelligence artificielle et autorégulation des contenus publicitaires – Projet Invenio

A l'initiative de :

ARPP

L'Autorité de Régulation Professionnelle de la Publicité -ARPP- est l'organisme d'autodiscipline publicitaire en France. Créée en 1935, l'organisation regroupe aujourd'hui près de 700 membres cotisants, des marques, des agences, des médias, des plateformes, des fournisseurs de solutions technologiques publicitaire (AdTech)… qui ensemble adoptent des règles déontologiques, s'engagent à les respecter et confient à l'ARPP le soin de veiller à la bonne application de ces règles.

Ecosystème du projet :

- Régies publicitaires digitales ;
- Éditeurs de sites web ;
- Plateformes de monétisation des inventaires ;

- Plateformes/Réseaux Sociaux.

Problématique :

Un des principaux enjeux auquel est aujourd'hui confrontée l'ARPP réside dans la démultiplication des canaux et des créations publicitaires, particulièrement depuis l'avènement de la publicité programmatique, de leur format (display, vidéo, audio, texte…), rendant complexe le contrôle des contenus publicitaires.

Face aux attentes de la société civile et des pouvoirs publics, les professionnels doivent apporter la preuve de leur capacité à mettre en œuvre un dispositif d'autorégulation efficient, moderne dans les technologies qu'il utilise, agile dans sa faculté d'adaptation à l'évolution du marché, représentatif, dans son inclusion de tous les acteurs de la chaîne de valeur, capable de maintenir des standards éthiques élevés dans un marché en pleine mutation digitale et devenu complexe, et de continuer à œuvrer en faveur d'une publicité légale, éthique, loyale et respectueuse de tous les publics, en concertation renforcée avec la société civile.

Description de la solution :

La publicité étant une combinaison d'images, de textes et/ou de sons, les technologies de traitement automatique du langage naturel, de Computer vision, de réseaux de neurones à convolution etc. basées sur l'apprentissage automatique peuvent aider à la détection d'éléments basiques dans les contenus publicitaires comme par exemple l'alcool (cf. la Loi « Evin »), la taille des caractères (lisibilité des mentions légales), d'hommes et de femmes (représentation contemporaine des genres, des âges, des diversités en publicité), de références textuelles ou sonores spécifiques (allégations santé, allégations environnementales…)…

Solution de Compliance As a Service : Mettre à disposition des acteurs du marché de la publicité digitale un catalogue de détectables aidant à la conformité légale et déontologique des contenus publicitaires. En effet, l'avènement de la publicité programmatique et l'automatisation des campagnes, particulièrement en RTB (système d'enchères en temps réel) ont rendu difficile l'identification et le filtrage manuel des pratiques publicitaires contraires aux règles légales ou déontologiques.

Bénéfices attendus :

La mise à disposition de modèles de détection automatique de suspicions de non-conformité, contribuera :

- à sécuriser la responsabilité des acteurs de la chaîne de valeur et leur permettra de se prévaloir de leur démarche éthique notamment auprès de toutes parties prenantes (consommateurs, pouvoirs publics, marques…).
- A mettre en œuvre un environnement numérique sain, de confiance, permettant au marché de se développer sur le long terme.

Dimension européenne

L'ARPP et ses homologues européens fédérés par l'EASA (European Advertising Standards Alliance) échangent dans le cadre d'un groupe de travail « Data driven self-regulation Committee ». Les partenaires tech des organisations européennes y sont invités, les codes sources y sont partagés. L'objectif est une mise à l'échelle européenne de ces outils pour des manquements communs.

Difficultés/barrières identifiées :

Un défi identifié réside dans la constitution de dataset correctement annotés, en volumétrie suffisante, et correspondant à des cas d'usage de «non-conformité» (ex : allégations environnementales trompeuses, véhicule sur un espace naturel, consommation alimentaire excessive…) afin de disposer de modèles dont la précision est jugée suffisante.

CAS D'USAGE POUR L'EDITION, LA DISTRIBUTION OU LA DIFFUSION

1. Gestion unifiée du consentement des utilisateurs

A l'initiative de :

DAG Conseil

Participants potentiels :

France Télévisions

Problématique :

Comme mentionné précédemment dans ce document, les revenus issus de l'exposition du public à de la publicité constituent une des trois sources de revenus liées aux ICC.

Ces revenus publicitaires s'appuient sur :

- La taille des audiences ;
- La qualification des audiences ;
- La capacité de l'éditeur du contenu à s'adresser à une audience pour l'exposer à un message.

Le paradigme actuel consiste, pour chaque éditeur (ou chaque régie associée à un pool d'éditeurs), à construire sa propre audience afin de la monétiser. Depuis plusieurs années, chaque régie s'attache donc à accroître la taille de son audience. Toutefois, devant la demande croissante de ciblage exprimée par les annonceurs ces régies se voient ensuite dans l'obligation de segmenter leurs audiences avec souvent le risque de se retrouver individuellement (régie par régie) avec des bassins d'audience beaucoup trop faibles.

Par ailleurs, l'acquisition du consentement publicitaire est actuellement donnée par l'usager pour chaque responsable de traitement (souvent une régie ou un éditeur). Ce fait conduit à la difficulté majeure pour le consommateur d'opérer la traçabilité/portabilité de son consentement tel que la réglementation RGPD l'y autorise. Cette difficulté conduisant à une défiance et à une attrition du partage des données par les consommateurs.

Le projet consiste à partager au sein d'un écosystème regroupant l'ensemble des éditeurs/régies, le consentement, les usages de contenu et les centres d'intérêt recueillis auprès de l'usager du contenu. Cet écosystème permettrait à l'usager de gérer son consentement pour plusieurs éditeurs/régies au sein d'une même interface.

Cette gestion des consentements au travers de cet écosystème HuB (qui ne se substitue pas aux entrepôts de données détenues par les différentes régies) permettrait aux acteurs de la filière de développer et proposer aux annonceurs de nouvelles offres de services :

- Statistiques cross régies ;
- Audiences segmentées cross régies ;
- Informations anonymisées ou pseudonymisées sur l'ensemble de la filière.

Description de la solution

Le projet consiste à mettre en place une couche de gouvernance, de stockage et d'analyse d'une liste précise de données sur laquelle s'entend l'ensemble des acteurs de la filière tout en garantissant la souveraineté de l'utilisateur sur la gestion de ses données et la conformité au RGPD.

Ces données partagées par les acteurs doivent non seulement être listées, mais doivent faire l'objet d'une formulation normalisée afin de favoriser l'interopérabilité entre les principaux acteurs que sont les éditeurs, les annonceurs, et les agences média.

Cette liste de données et les modalités de gouvernance doivent être définies dans la phase de cadrage du projet et pourrait entrer dans le périmètre du « Data Space » que le GT appelle de ces vœux dans ce présent document.

Bénéfices attendus :

- Bénéfices côté régies et éditeurs :
  - Meilleur respect de l'évolution du consentement des usagers dans le temps ;
  - Minimisation des risques juridiques liés au non-respect des textes RGPD ;
  - meilleure connaissance des comportements d'usage et centres d'intérêt des usagers (granularité à définir lors du cadrage projet) ;
  - Possibilité de monétiser la data et/ou la connaissance « as a service » au travers de l'écosystème partagé.
- Bénéfices côté usagers :
  - Meilleure visibilité du statut de ses consentements ;
  - Meilleure expérience dans la mise à jour de ses consentements ;
  - Meilleure visibilité sur ces centres d'intérêt prédictifs et possibilité de mise à jour.
- Bénéfices côté annonceurs :
  - Possibilité de disposer d'une visibilité des volumes d'audience cross régies ;
  - meilleure connaissance des affinités des usagers.
- Bénéfice côté régulateur :
  - Meilleure applicabilité des textes.

Difficultés/barrières identifiées :

Comme pour beaucoup de projets consistant à partager certaines données, il est probable que le principal frein concerne la volonté des éditeurs et/ou régies à partager certaines données.

Partenaires/écosystème du projet :

- Les principaux éditeurs et les principales régies digitales et TV ;
- Les agences média ;
- Les opérateurs fixe et mobile ;
- Les plateformes souveraines d'AD exchange et de digital programmatic.

Identification des principales technologies Gaia-X en jeu :

Les fonctionnalités proposées par un socle technique Cloud + gouvernance des données au travers des technologies WEB3 semblent permettre de construire un écosystème décentralisé tel que celui nécessaire au projet proposé.

Le périmètre du besoin technique devra être défini à l'issue de la phase de cadrage du projet.

2. Accélérer la rémunération des ayants droit

A l'initiative de :

SACEM

Participants potentiels :

France Télévisions

Projet de Hub Médias.

La Sacem protège, représente et défend les intérêts de plus de 182 520 membres en France et à l'international, auteurs, compositeurs et éditeurs. Au service de la création, elle dispose du plus vaste répertoire mondial et s'engage pour la diversité musicale en soutenant de nombreux projets. Société privée à but non lucratif, la Sacem est gérée par ses membres qu'elle accompagne à travers 3 missions essentielles :

- Collecter et répartir les droits d'auteur
- Promouvoir et soutenir les créateurs
- Défendre et protéger ses membres

La Sacem assure la collecte et la répartition des droits d'auteur pour la diffusion publique (médias audiovisuels, salles de concert, festivals, services internet, cinémas, magasins, etc.) et pour la reproduction sur support (disques, vidéos, fichiers numériques légaux, DVD, CD-Rom, jeux vidéo…) des œuvres qu'elle représente. Les droits d'auteur sont répartis au plus près de l'utilisation réelle des œuvres : 80% œuvre par œuvre, 9% par sondage et 11% en tenant compte des habitudes de consommation.

La problématique et les parties prenantes

Afin d'assurer la collecte et la répartition des droits d'auteur pour la diffusion publique dans les médias audiovisuels, la SACEM utilise actuellement des déclarations fournies par chaque média. Chaque média investit des ressources dédiées au monitoring de ses chaînes et à l'établissement de ces relevés. Malgré tout, il manque parfois des informations clés relatives aux programmes et aux œuvres musicales associées pour permettre la juste répartition des droits associés aux auteurs-compositeurs et éditeurs de ces œuvres. Cette situation entraîne à son tour une charge de travail importante chez les sociétés d'auteur pour analyser les relevés fournis par les diffuseurs et réaliser la meilleure répartition possible.

L'amélioration de ce processus concerne ainsi l'ensemble des parties prenantes de l'écosystème de la création audiovisuelle :

- les Producteurs de programmes. Ils disposent des données originales concernant les œuvres diffusées, mais les communiquent peu ou mal aux autres acteurs ;
- les Médias. Ils remplissent leur obligation légale de manière individuelle, en faisant potentiellement appel à des prestataires pour produire et enrichir les données de diffusion, ce qui se traduit pour eux par des charges importantes sur le plan humain, technique et financier ;
- les Organismes de Gestion Collective (OGC). Ils reçoivent des données souvent incomplètes ou incohérentes, ce qui nécessite un énorme travail de vérification et d'enrichissement manuel qui se traduit par un temps de traitement long, alors que l'exactitude et la rapidité de versement des droits est un enjeu clé pour l'activité des créateurs ;

- les Ayants droit, en bout de chaîne. Ils dépendent, pour leur rémunération, aussi bien en termes de justesse que de délai, de la qualité des données et de la mise en œuvre de processus efficaces entre les parties prenantes citées ci-dessus.

Si cette problématique n'est pas nouvelle, elle devient de plus en plus pressante pour les acteurs traditionnels à mesure que le secteur audiovisuel se numérise. En effet le numérique a généré un développement massif de la consommation de contenus audiovisuels en ligne donc une augmentation colossale des quantités de données à traiter par les OGC. En même temps, le numérique constitue une formidable opportunité de faciliter la mutualisation et l'analyse des données.

Les producteurs, les médias et les organismes de gestion collective font ainsi face à un triple défi :

- Technologique : accélérer les temps de traitement des données, les fiabiliser et simplifier leur réconciliation et leur enrichissement ; limiter la production et le traitement manuel des données ;
- Économique : réduire les coûts consacrés à ces activités en mutualisant une partie des traitements et des investissements nécessaires à la construction des outils informatiques. Eventuellement, dans un second temps, créer de nouveaux revenus à partir des nouvelles données ainsi générées ;
- Politique : développer une stratégie partenariale entre les acteurs du secteur pour défendre une vision commune de la création culturelle et de la souveraineté technologique.

<u>Objectifs du projet</u>

Pour répondre à ces problématiques, le "Hub Médias" se positionne de manière transverse au sein de l'écosystème. Il aura pour mission de mettre à disposition des producteurs, des médias et des OGC une plateforme technologique commune, permettant l'accélération et la fiabilisation du processus de déclaration des données concernant les œuvres et la diffusion de celles-ci, en :

- Facilitant la production de relevés de diffusion par les médias ;
- Assurant la qualité et la complétude des données, notamment les données relatives aux œuvres, intégrées dans ces relevés ;
- Partageant les données nécessaires entre les parties prenantes au processus de répartition des droits ;
- Mettant en commun et en industrialisant des moyens techniques entre les partenaires sous la forme d'une plateforme
- Assurant la sécurité et la pérennité de la plateforme en permettant aux partenaires de choisir le niveau de confidentialité des données partagées dans le cadre de ce Hub ;
- Permettant la reproduction des processus à même de proposer, à terme, des services commerciaux dérivés

<u>Bénéfices attendus</u>

Le déploiement du « hub média » apportera plusieurs bénéfices :

- Rémunération plus juste et plus rapide des ayants droit et donc soutien global à la création dans un modèle responsable et soutenable
- Réduction globale des coûts et gain de productivité pour l'ensemble des parties prenantes
- Simplification des processus de recueil et de traitement des données nécessaires à la répartition des droits pour les OGC et les diffuseurs
- Prise en compte des nouveaux usages et innovation dans le domaine de la gestion collective, afin de pérenniser le leadership français dans ce secteur

<u>Principales technologies Gaia-X en jeu :</u>

Plateforme data : partage, échange et potentiellement intelligence des données entre plusieurs acteurs de la filière ICC, avec potentielle implication des pouvoirs publics, pour défendre un modèle à la française de défense du droit d'auteur et de la création, tout en prenant en compte l'évolution des usages face au risque de désintermédiation.

Sécurité et confidentialité : Authentification, gestion de fédération d'identité, gestion, sécurisation des données gestion du consentement et contrôle de l'utilisation.

Conformité : portabilité des données, gestion des droits, onboarding et certification.

3. Plateforme collaborative de création et de diffusion de contenus éditoriaux – projet CREA

A l'initiative de :

Agencia EFE, BSC SME, Cap Digital, Ceipes, Panodyssey, Pro Progressione, Venetian Cluster, Voxeurop, Worldcrunch

Panodyssey est une plateforme collaborative de création et de diffusion de contenus éditoriaux de fiction et de non-fiction au service des créateurs professionnels et amateurs. Elle rassemble une communauté de lecteurs et d'auteurs autour de valeurs fortes en **réconciliant éthique** et **numérique**.

La plateforme est pensée pour permettre aux auteurs de développer leur créativité en explorant de nouveaux formats narratifs. Elle intègre nativement les fonctionnalités qui permettent aux auteurs de développer leur communauté et de gérer leur rémunération en fonction de leurs objectifs - dans le **respect du droit d'auteur** et en toute transparence grâce à la technologie blockchain.

Participants actuels et futurs :

Le consortium autour du projet CREA réunit 9 partenaires internationaux et 5 pays (Bulgarie, Espagne, France, Hongrie, Italie) dans les secteurs des médias, de la culture, de l'éducation et de la technologie. Il est créé et soutenu par la Commission européenne dans le cadre du programme Créative Europe Lab Innovation.

Membres du consortium :

- **AGENCIA EFE** : Agence de presse espagnole, leader européen dans le secteur audiovisuel (1939)
- **BSC SME** : Organisation non gouvernementale bulgare dans le secteur culturel (1996)
- **CAP DIGITAL** : Pôle européen de la transformation numérique et écologique (2006)
- **CEIPES** : Établissement d'enseignement italien pour l'éducation et le développement (2007)
- **PANODYSSEY** : Start-up française dans le secteur des contenus numériques (2018)
- **PRO PROGRESSIONE** : Organisation hongroise multidisciplinaire pour les arts et la culture (2008)
- **VENETIAN CLUSTER** : Cluster italien pour la culture et l'innovation basé à Venise (2014)
- **VOXEUROP** : Média en ligne multilingues au rayonnement européen (2014)
- **WORLDCRUNCH** : Média en ligne anglophone international (2011)

Ce consortium est ouvert à de nouveaux partenaires associés.

Projet CREA :

L'objectif du projet est de créer un nouvel espace numérique européen grâce à une approche innovante autour de la création et la diffusion de contenus numériques en abordant des sujets sociétaux tels que la désinformation et l'éducation aux nouveaux médias qui sont des priorités de l'Union Européenne. Le projet ambitionne de réunir une communauté de 10 millions de créateurs et 100 millions d'internautes à horizon 2030 et ainsi créer une alternative solide aux GAFAM.

En constante évolution, le consortium tisse de nouveaux partenariats économiques européens pour devenir un écosystème de référence dans les secteurs audiovisuels, culturels et créatifs.

Financement du projet :

Ce projet est entièrement financé pour toute la durée de ses activités en 2022 et 2023. La Commission européenne contribue à hauteur de 60% du budget et les partenaires fondateurs du projet CREA financent les 40% restant.

Problématique et ambition :

L'objectif du projet CREA, soutenu par toute la filière, est de créer une alternative européenne pour ne plus dépendre exclusivement des plateformes américaines. Il est indispensable de faire émerger des opérateurs locaux pour favoriser et valoriser économiquement les contenus des créateurs en Europe.

Les nouvelles formes de **créations hybrides** et les **nouveaux formats créatifs** sont un enjeu majeur pour l'innovation des secteurs culturels. Ils ont besoin d'un nouvel outil transsectoriel pour accompagner le développement de ces nouveaux formats.

La vision du collectif et ses valeurs est de proposer des solutions professionnelles, en adéquation avec les **normes européennes**, et faciliter la **coopération transfrontalière** dans les secteurs des industries culturelles et créatives en Europe.

Pour lutter contre la désinformation, Panodyssey a mis en place un système sécurisé permettant de garantir la certification de l'identité (ID Check) des créateurs de contenu. Dès lors, les faux profils, les faux contenus et les robots n'ont pas accès à la plateforme. L'absence de publicité est un moyen complémentaire et efficace pour lutter contre les mécanismes économiques qui financent la désinformation. L'action du CREA va continuer ce travail d'éducation, notamment avec son activité "Laboratoire d'éducation média", qui consiste à aborder des thématiques sociétales, comme la désinformation, l'inclusion et l'accessibilité digitale, au travers d'initiatives diverses, impliquant tous les secteurs culturels, audiovisuels et créatifs.

Ces secteurs sont particulièrement sensibles à la protection et à la valorisation de la propriété intellectuelle. Pour lutter contre l'utilisation de contenu, sans l'autorisation des ayants-droits, la solution est dotée de processus de sensibilisation et de gestion de la propriété intellectuelle. En effet, la certification d'identité permet de garantir la paternité des œuvres et de remplir les obligations légales en matière de transactions financières en toute **transparence**. Pour industrialiser la valorisation et la protection de la propriété intellectuelle, la plateforme de contenu est dotée de **technologie blockchain,** à une expertise IP et un **Smart Contract IP**. Ce **Smart Contract IP** est l'actif technologique clé pour industrialiser, sans limite, la gestion des collaborations entre créateurs de contenu (tous types de contenus confondus) et l'automatisation de la gestion des droits et reversements. L'utilisation de la **blockchain** est parfaitement appropriée à cet objectif parce qu'elle permet d'assurer une transparence totale en décentralisant les registres. La blockchain renforce quant à elle **la confiance** par son caractère infalsifiable.

Pour encourager la **diversité** et renforcer la **visibilité des contenus européens**, le CREA s'inscrit pleinement dans une **démarche transsectorielle,** étant précisé que chaque partenaire du CREA est une tête de réseau dans des secteurs complémentaires : média et audiovisuel, culture, secteurs créatifs et éducatifs.

Bénéfices attendus :

- **Bénéfices attendus pour les créateurs de contenus** : Valoriser la création de contenus européens par la protection de la propriété intellectuelle et la mise en valeur des œuvres de l'esprit. Favoriser les nouvelles expériences en ligne autour de l'écriture créative et collaborative.
- **Bénéfices pour les utilisateurs** : Permettre aux citoyens d'avoir un outil facile à utiliser et garant des informations diffusées. Favoriser l'accessibilité à des contenus de qualité autour de la littérature et de la connaissance. Déployer une nouvelle vision autour d'un algorithme choisi par l'utilisateur.
- **Bénéfices côté régulateur** : Anticipation de la mise en application des nouvelles normes européennes telles que le *Digital Services Act*

Points forts :

- Un projet ambitieux, regroupant des acteurs européens de qualité et spécialistes dans leur secteur ;
- Une équipe avec une expertise technologique et/ou sectorielle ;
- Un projet soutenu et cofinancé par la Commission européenne ;
- Une capacité à fédérer des parties prenantes dans les secteurs culturels et créatifs à travers l'Europe.

CAS D'USAGE POUR LA VENTE DE BIENS CULTURELS

1. Agrégation des données de billetterie

A l'initiative de :

Ministère de la Culture

Participants :

Arenametrix

La problématique :

La billetterie est une brique stratégique dans le champ culturel, essentielle pour les modèles économiques des opérateurs mais aussi pour la connaissance des publics, la communication et la fidélisation.

La dématérialisation des systèmes de billetterie soulève de multiples enjeux et le secteur est aujourd'hui globalement dépourvu d'acteur en capacité de rassembler et de mettre en visibilité les offres portées par des structures culturelles aux statuts multiples, intervenant dans des champs d'activité très divers et générant un nombre très élevé de transactions (plus de 50 millions de transactions pour les principaux musées français en 2019 uniquement, auxquels s'ajouteraient les théâtres, monuments et autres lieux culturels avec une billetterie dédiée).

En outre, l'émergence de nouveaux opérateurs spécialisés dans la vente et la distribution électronique de billets et le développement de plateformes d'achats et de service en ligne ont significativement modifié le paysage concurrentiel (verticalisation de certains groupes, concentration et émergence d'acteurs mondiaux, nouveaux intermédiaires issus de la communication ou du paiement en ligne, etc.). Enfin les secteurs et les acteurs présentent une grande variété de maturité en termes de dématérialisation ; les plus petits acteurs étant aujourd'hui les moins bien équipés. L'évolution des usages à la faveur des grandes plateformes de distribution de billets ou d'hébergement et de tourisme fait porter un risque important sur la visibilité de la diversité culturelle des territoires.

Pour une grande majorité de producteurs (ICC, tourisme, loisir, etc.), les plateformes de distribution de billets, d'hébergement et de tourisme sont devenues incontournables pour l'acquisition de clients et leur développement économique. Ce modèle conduit naturellement à un éclatement de la valeur des offres et a des conséquences préjudiciables pour les producteurs et les usagers :

- La bataille pour l'accès direct aux des consommateurs et aux données les concernant
- D'importantes marges intermédiaires sur le prix des billets
- Le manque de visibilité pour les petits acteurs dépourvus de solutions

Solution recherchée :

Afin de ne pas laisser s'installer une dissymétrie dommageable en particulier aux plus petits acteurs, la mise en commun et l'agrégation des données issue des différents systèmes de billetterie permettrait à l'ensemble du secteur de mieux comprendre les comportements et les aspirations du public et, ainsi, d'être en capacité d'adapter la programmation des événements culturels et de favoriser la rencontre des artistes et des publics. Ces données pourraient également être partagées avec d'autres secteurs (tourisme, transport, etc.) pour l'adaptation de leurs propres offres.

Les données à rapprocher sont :

- Les données de ventes ;
- Les lieux de vente ;
- Les informations disponibles sur les clients.

Bénéfices attendus :

- Meilleures connaissances des publics ;
- Adaptation des programmations culturelles ;
- Visibilité sur les offres des petites structures ;
- Possibilité de créer des offres combinées avec des offres touristiques.

Parties prenantes :

- Les producteurs ;
- Les diffuseurs ;
- Les intermédiaires de billetterie.

2. Sécurisation de la billetterie

Assurer la traçabilité de la billetterie et lutter contre la fraude et le marché noir.

CAS D'USAGE TRANSVERSES

1. Alliance Culture Data : projet de plateforme d'échange de données de la filière des Industries Culturelles et Créatives

A l'initiative de :

BnF-Partenariats

BnF-Partenariats, filiale de droit privé de la BnF (Bibliothèque Nationale de France), a été créé en 2012 afin d'accroître la numérisation et la visibilité des collections de la BnF grâce à la construction de partenariats avec de nombreux acteurs du numérique et des ICC (Edition, Presse , Musique en ligne...).

Forte de son positionnement et tournée vers l'innovation, BnF-Partenariats a entrepris de fédérer les ICC autour d'un projet de plateforme d'échange de données sectorielle à l'instar d'autres industries.

Initialement amorcé et formalisé par BnF-Partenariats dans Alliance Culture Data, ce projet a besoin désormais d'un ou plusieurs acteurs à même d'en assurer le portage et le développement.

Ecosystème du projet :

Le secteur des ICC.

La problématique :

La donnée : un pilier de l'économie de demain

Les données sont au cœur de la transition digitale en marche au sein de nombreuses industries. Selon la Commission européenne, la quantité de données générées par les organismes publics, les entreprises et les citoyens devrait être multipliée par cinq entre 2018 et 2025. Bien orchestrées et partagées, ces données sont à la fois source d'optimisation des organisations, d'innovation et de monétisation. De nombreuses initiatives visant à faciliter la circulation des données à l'échelle d'une filière ont d'ores et déjà vu le jour notamment dans l'agriculture, l'agro-alimentaire, la santé, le tourisme… et ont ainsi contribué à faire de la donnée un moteur clé de leur développement. Parallèlement, l'émergence d'un marché unique de la donnée conformément à la stratégie européenne - incarnée par le Règlement sur la gouvernance des données - devrait offrir un environnement juridique et économique plus favorable à la circulation des données.

Le partage et la valorisation de la donnée dans les Industries Culturelles et Créatives constituent un gisement à exploiter.

Le partage des données au sein des Industries Culturelles et Créatives (ICC) portent plusieurs enjeux clés pour le développement économique et la structuration de la filière :

- Générer de la valeur et renforcer la compétitivité des entreprises des ICC ;
- Fédérer les secteurs face aux géants de l'économie de la donnée ;
- Innover, stimuler l'écosystème des startups et développer l'IA.

Or, les nombreuses données produites depuis des années par les ICC au gré de la digitalisation de leur économie sont aujourd'hui sous-exploitées et sous-valorisées.

Des freins identifiés au développement des échanges de données au sein des Industries Culturelles et Créatives

Un certain nombre de contraintes financières, techniques, juridiques et commerciales freinent l'échange de données au sein des ICC. L'incapacité de chaque acteur à lever seul ces obstacles favorise

aujourd'hui le silotage des données et ne permet pas à la filière de tirer pleinement partie de leur valeur.

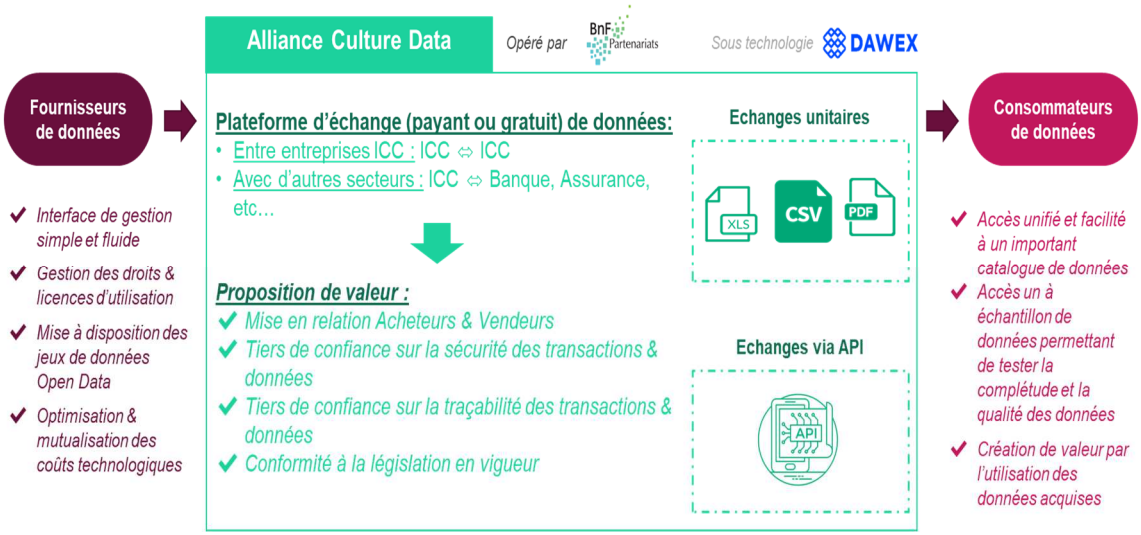
En dépit de quelques initiatives isolées, les ICC ne disposent pas aujourd'hui d'infrastructures techniques fiables et partagées permettant d'échanger de façon sécurisée tout en garantissant la traçabilité et le bon usage qui en est fait de leurs données.

Le projet Alliance Culture Data :

Avec Alliance Culture Data (ACD), BnF-Partenariats a initié un projet d'envergure proposant de répondre à ces enjeux en offrant un lieu d'échange et de partage des données sécurisé (1) et en accompagnant les différents acteurs via des briques servicielles (2).

1) La plateforme

ACD se définit comme une plateforme permettant à ses acteurs de fournir ou d'acquérir, gratuitement ou contre rémunération, tout type de données. La plateforme se constitue comme « intermédiaire de données » au sens du Règlement sur la gouvernance des données. La solution technique de la plateforme serait fournie par la société française Dawex dont la technologie est déjà utilisée par plusieurs industries.



2) Les services spécifiques autour de la data

- En amont (services indépendants des échanges de données) :
  - Structuration, mise en qualité et nettoyage de la donnée
  - Formation à la gouvernance de la donnée
  - Estimation de la valeur de la donnée
  - Packaging de la donnée répondant aux cas d'usage concrets
- En aval (services intervenant à la suite de l'échange de données auprès des acquéreurs et proposés par des entreprises spécialisées via la plateforme) :
  - Analyse des données quantitatives et qualitatives
  - Mise à disposition de compétences data, services de data marketing et data strategy
  - Packaging de la donnée répondant à des besoins spécifiques

La proposition de valeur ACD apportera ainsi une réponse concrète aux contraintes freinant aujourd'hui l'accès et l'exploitation des données culturelles et vise à développer un usage de la donnée massif et créateur de valeur, à l'image des autres secteurs industriels.

Des externalités positives fortes pour la filière

Les externalités positives liées à l'échange et au partage de données rendues possibles par la plateforme sont nombreuses, à la fois pour la filière et les acteurs la composant, mais également pour des secteurs extra-ICC, permettant ainsi de démultiplier les liens et échanges entre les ICC et les autres industries. Plus de trente cas d'usage et d'applications possibles facilités par ACD ont été identifiés dans le cadre d'entretiens menés par BnF-Partenariats avec les ICC.

ACD apparaît également comme un outil de lobbying à même de fédérer les ICC et d'influencer la stratégie et la réglementation sur les échanges de données face aux géants du web.

Un modèle économique bâti pour permettre à Alliance Culture Data de s'autofinancer au-delà de son financement initial

Par son intérêt stratégique pour la filière et par les externalités positives qu'elle génère, Alliance Culture Data est un projet permettant de préparer les conditions de création de valeur pour l'avenir. C'est également un projet d'envergure à la fois par son périmètre et par le degré d'innovation qu'il suppose.

Le modèle économique du projet Alliance Culture Data repose sur deux sources de revenus :

- Les ventes d'abonnements à la plateforme augmentées des services d'onboarding
- Les ventes de services annexes proposés, en amont et aval de l'échange de données sur la plateforme

Le projet nécessite un besoin de financement en phase d'amorçage estimé à 6M€.

Le projet a vocation à être autonome financièrement vis-à-vis du soutien public à l'issue de la phase d'amorçage. La trajectoire de déploiement prévue à date permet à ACD de s'autofinancer au bout de 5 ans suivant son démarrage et de dégager un retour sur investissement positif au bout de 8 ans.

3 BESOINS TRANSVERSES

DEFINITION D'UNE ONTOLOGIE DES METADONNEES LIEES AUX CONTENUS CULTURELS ET CREATIFS

A l'initiative de :

Participants potentiels :

Problématique :

Ce projet vise à cartographier les métadonnées permettant d'étiqueter les données culturelles et créatives, d'en définir les standards (existants ou à créer) et de définir l'écosystème technique permettant d'héberger ces métadonnées au cours du cycle de vie des contenus propriétaires ou partagées.

Ces métadonnées de description, de gouvernance, d'usage, d'intégrité et de valorisation favoriseront le partage des données entre les différents acteurs ainsi que leur gouvernance. Elles serviront de base à la mise en place de smart contrats entre les multiples acteurs de la filière.

Livrable principal : Dictionnaire des Métadonnées rattachées aux données culturelles et créatives (incluant Taxonomie/Ontologie) + Guide des standards d'étiquetage des données.

GOUVERNANCE D'UN DATA SPACE ICC

Création d'une instance pour :

- Dissémination
- Formations
- Modèles
- Transmission des standards
- API
- Définition des processus et critères d'acceptance ou rejet de participants ;
- Choix des briques technologiques ;
- Veille sur le cadre légal ;
- Recherche de financements ;
- Etc.

5 POURSUITE DES TRAVAUX

Ce document de positionnement est loin de décrire de manière exhaustive l'ensemble des cas d'usage pertinents dans le cadre de GAIA-X pour les ICC. Les membres du secteur sont invités à l'amender, à le compléter et à signaler aux porteurs du groupe de travail tout intérêt pour l'une ou l'autre des initiatives présentées, que ce soit en tant que contributeur ou en tant que porteur de projet.

6 ANNEXE

La numérisation des secteurs

Pour la musique, ce fut un bouleversement complet du secteur de la production et de la distribution, ayant abouti à une chute vertigineuse de chiffre d'affaires pendant plusieurs années de mutation, puis à une hausse aboutissant désormais à envisager de dépasser le chiffre d'affaires qui existait avant le début de la transformation.

Les nouveaux canaux de diffusion des œuvres, tels que les plateformes de streaming et les services de partage de vidéos, dont beaucoup sont musicales, ont bâti de nouveaux modèles économiques aboutissant à relancer la croissance du secteur grâce aux transformations techniques et à l'évolution des usages. En parallèle, la construction de référentiels de données pour faire découvrir les œuvres et les enregistrements sur ces nouveaux canaux et sur les canaux traditionnels accompagne cette transformation.

Dans le domaine du jeu vidéo, cette dématérialisation est certes native mais le développement des jeux en ligne est devenu un élément déterminant de cette industrie, qui a également profondément diversifié ses modes d'accès et ses modèles économiques, ainsi que la richesse de ses propositions créatives grâce à l'accroissement des puissances de calcul disponibles pour le public.

L'audiovisuel connaît un véritable bouleversement avec l'arrivée des plateformes de streaming mondiales qui dépensent des sommes gigantesques dans la création exclusive sur leurs canaux. Elles sont amenées, d'emblée mais également de façon encadrée par la régulation, à investir localement pour assurer une proximité avec le public des différents marchés dont elles ont besoin pour croître mais également pour renouveler leurs créations en permanence, dans un contexte de concurrence extrême. Les marchés financiers sont essentiels à leur développement et la régulation européenne s'adapte à cette puissance nouvelle qui a pris pied sur le continent européen.

En parallèle, les plateformes de partage de vidéos telles que YouTube, jouent un rôle de promotion des œuvres, de prescription voire de modèle économique nouveau pour certains types d'œuvres audiovisuelles comme le documentaire.

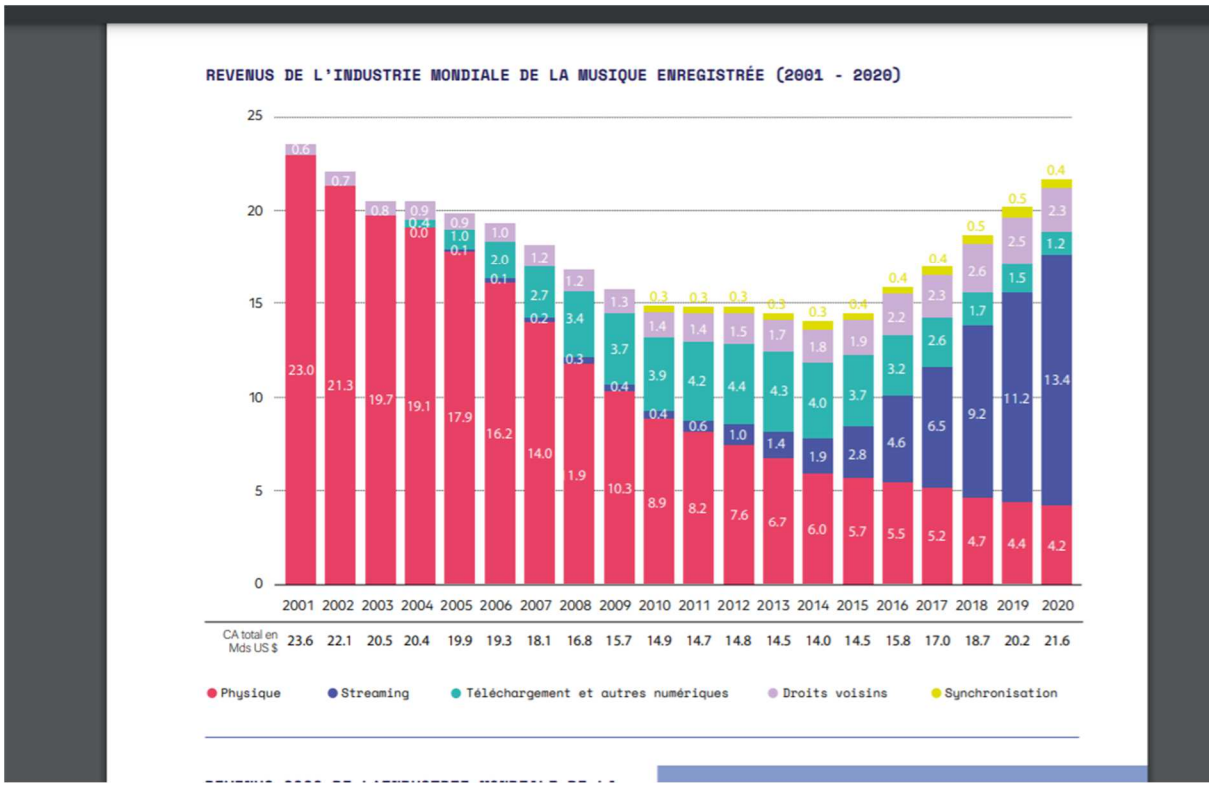
Les médias audiovisuels classiques continuent leur travail d'exposition des œuvres et de financement pour ce qui concerne les chaînes de télévision mais en réinventant leur modèle via l'introduction d'outils interactifs, à la fois pour délinéariser leurs programmes (podcasts et télévision de rattrapage) et pour coller aux goûts des publics notamment dans l'optique d'une meilleure proposition publicitaire (publicité ciblée).

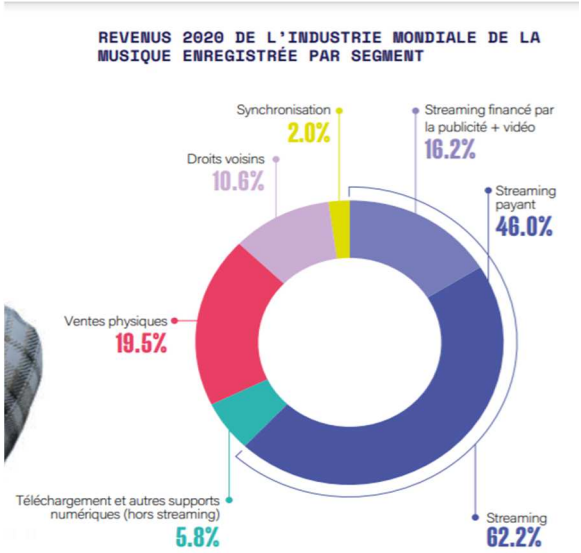
Néanmoins, dans tous les cas, les données portant sur les œuvres et les usages sont essentielles pour leur découvrabilité, qu'il s'agisse de référencement ou de compréhension des goûts du public et des modes de diffusion et de promotion les plus pertinents selon les œuvres.

Exemples :

i) Musique

(Source : SNEP – IPFI)





Quelques chiffres complémentaires :

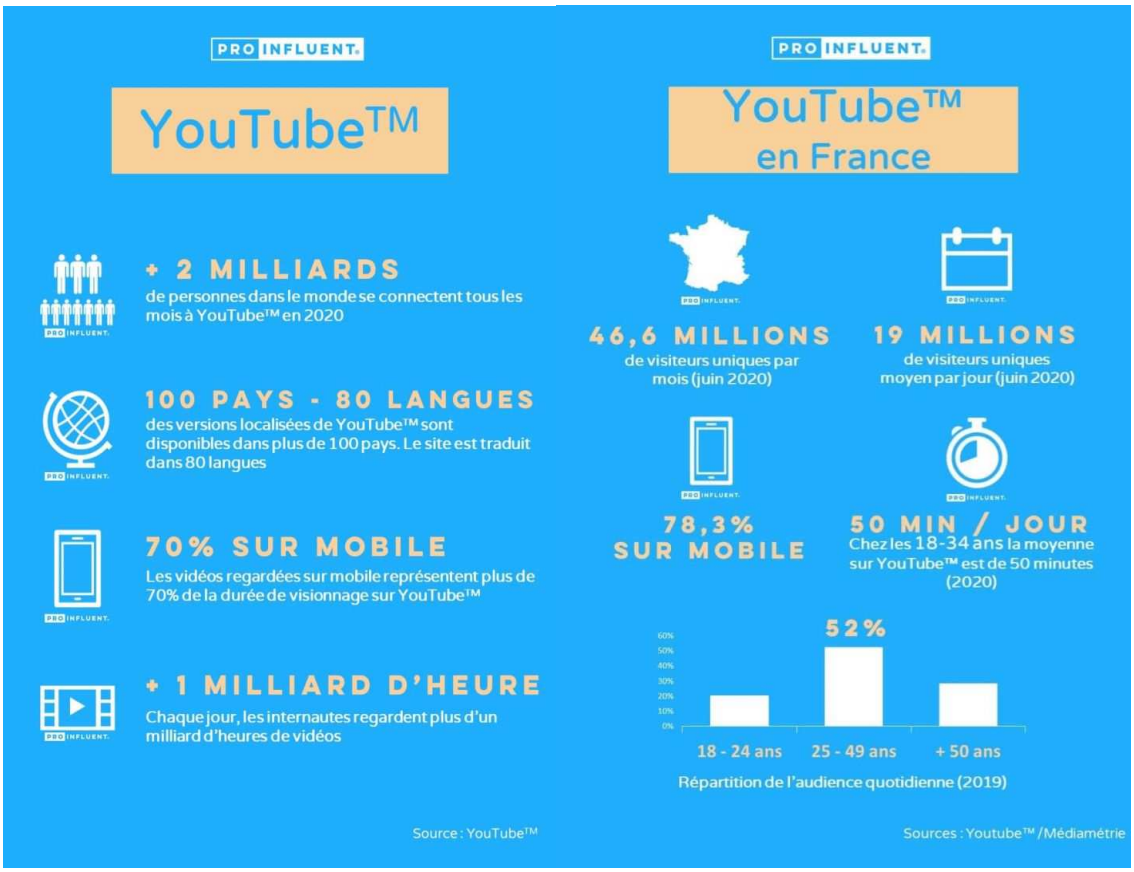
- Les revenus :
  - Depuis 2016, le numérique représente plus de 50% des revenus du marché de la musique enregistrée (Unesco 2020).
  - La croissance des revenus issus du numérique tire la croissance des revenus de la musique enregistrée : respectivement +19,9% et +7,4% en 2020 (IFPI).

- Les utilisateurs :
    - On compte 443 millions d'utilisateurs de comptes payants de streaming musical en 2020, soit 30% de plus en un an (IFPI).
    - On compte 345 millions d'utilisateurs actifs par mois sur Spotify en 2021 (les Echos).

ii) Audiovisuel

Quelques chiffres complémentaires :

- Netflix compte plus de 200 millions d'abonnés dans le monde et pourrait atteindre 286 millions d'abonnés en 2026 selon Digital TV Research,
- Amazon Prime compte plus de 200 millions d'abonnés dans le monde dont au moins 175 millions ont regardé le service Amazon Prime Video en 2020,
- Disney + compte près de 120 millions d'abonnés dans le monde et, selon Digital TV Research, pourrait dépasser Netflix en nombre d'abonnés en 2026 en atteignant le chiffre de 294 millions,
- Chiffres YouTube 2021 (pour 2020) :



Compléments sur les définitions des données

Métadonnées

Dans son sens large, une métadonnée « est une donnée servant à définir ou décrire une autre donnée, quel que soit son support (papier ou électronique) » (wikipédia).
Le grand dictionnaire terminologique de l'Office québécois de la langue française s'avère pertinent pour la définition d'une métadonnée numérique :

> « Ensemble structuré de données accompagnant un ouvrage et servant notamment à en décrire le contenu et le format, à assurer son indexation dans les moteurs de recherche et les bases de données, et à faciliter la gestion des droits d'auteur qui y sont liés. […] Dans la perspective des entrepôts de données, les métadonnées sont un élément primordial et sont destinées à diverses catégories d'utilisateurs. Elles permettent notamment de connaître l'origine et la nature des données stockées dans l'entrepôt, de comprendre comment elles sont structurées, de savoir comment y avoir accès et comment les interpréter, de connaître les différents modèles de données en présence et les règles de gestion de ces données. »

Dans le cadre d'un rapport sur les métadonnées culturelles réalisé pour le ministère de la Culture et des Communications français, Pierre Lescure apporte pour sa part certaines précisions relatives aux métadonnées culturelles, qu'il décrit comme :

> « l'ensemble des informations décrivant une ressource quelconque, numérique ou non. Leur fonction première est de décrire le contenu de la ressource, tout en permettant de l'identifier, de le qualifier et de l'enrichir ; elles ne sont pas nécessairement contenues dans le document lui-même. […] [Les métadonnées] sont l'ADN des contenus culturels ».

Données d'usage

Selon la définition fournie par des rapports canadiens et français sur le sujet[4], les données d'usage sont des données sur la consommation d'un contenu culturel en particulier, constituées de manière automatique par l'activité des internautes fréquentant cette ressource en ligne.

Il existe plusieurs types de données d'usage pouvant être qualifiées de « données d'usage étendues », qui recouvrent un spectre assez large d'informations utiles au processus de création et d'exploitation des œuvres culturelles, notamment :

- les données brutes de consommation : nombre de vues, d'écoutes, heures et jours d'accès, etc.;
- les données relatives au contexte d'exposition des contenus : nombre de fois où le contenu a été proposé par un algorithme de recommandation, termes de recherche qui ont mené à l'accès au contenu, listes de lecture éditoriales dans lesquelles le contenu a été inclus, etc ;
- les caractéristiques des publics : données sociodémographiques, préférences ou goûts, nombre d'amis sur les réseaux sociaux, etc.;
- les données permettant de comparer la performance de ces contenus aux autres contenus du même type proposés sur la plateforme : nombre de vues moyen par catégories de contenus, etc.

Données personnelles

(Source : CNIL)

Une « *donnée personnelle* » est « *toute information se rapportant à une personne physique identifiée ou identifiable* ».

Une personne peut être identifiée :

- **directement** (exemple : nom, prénom)
- **ou indirectement** (exemple : par un identifiant (n° client), un numéro (de téléphone), une donnée biométrique, plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale, mais aussi la voix ou l'image).

L'identification d'une personne physique peut être réalisée :

- **à partir d'une seule donnée** (exemple : numéro de sécurité sociale, ADN)
- **à partir du croisement d'un ensemble de données** (exemple : une femme vivant à telle adresse, née tel jour, abonnée à tel magazine et militant dans telle association)

Exemple : une base marketing contenant de nombreuses informations précises sur la localisation, l'âge, les goûts et les comportements d'achats de consommateurs, y-compris si leur nom n'est pas

[4] Rapport de la mission franco-québecoise sur la découvrabilité des contenus culturels francophones en ligne (2020) - Laticce, Etude ACEI 2018-2020 – « *Etat des lieux sur les métadonnées relatives aux contenus culturels* », Observatoire de la culture et des communication du Québec (2017)

stocké, est considérée comme un traitement de données personnelles, dès lors qu'il est possible de remonter à une personne physique déterminée en se basant sur ces informations.

Les mégadonnées (« big data ») [5]

Dans le contexte du numérique, les données d'usage s'accumulent de manière automatique et à grande vitesse ; ce sont des mégadonnées (*big data*). Le terme « mégadonnées » est utilisé pour décrire un très grand nombre de données numériques structurées, mi-structurées et non structurées, ayant le potentiel de contenir une mine de renseignements si elles sont habilement exploitées.

Mais les mégadonnées sont si volumineuses qu'elles sont difficiles à entreposer, à gérer, à compiler et à analyser avec des outils classiques de traitement et d'analyse de bases de données.

Le concept de « mégadonnées » fait également référence à la capacité de mettre en relation plusieurs bases de données et, grâce à l'utilisation d'analyses avancées, d'identifier des *patterns* d'information qui demeureraient autrement invisibles.

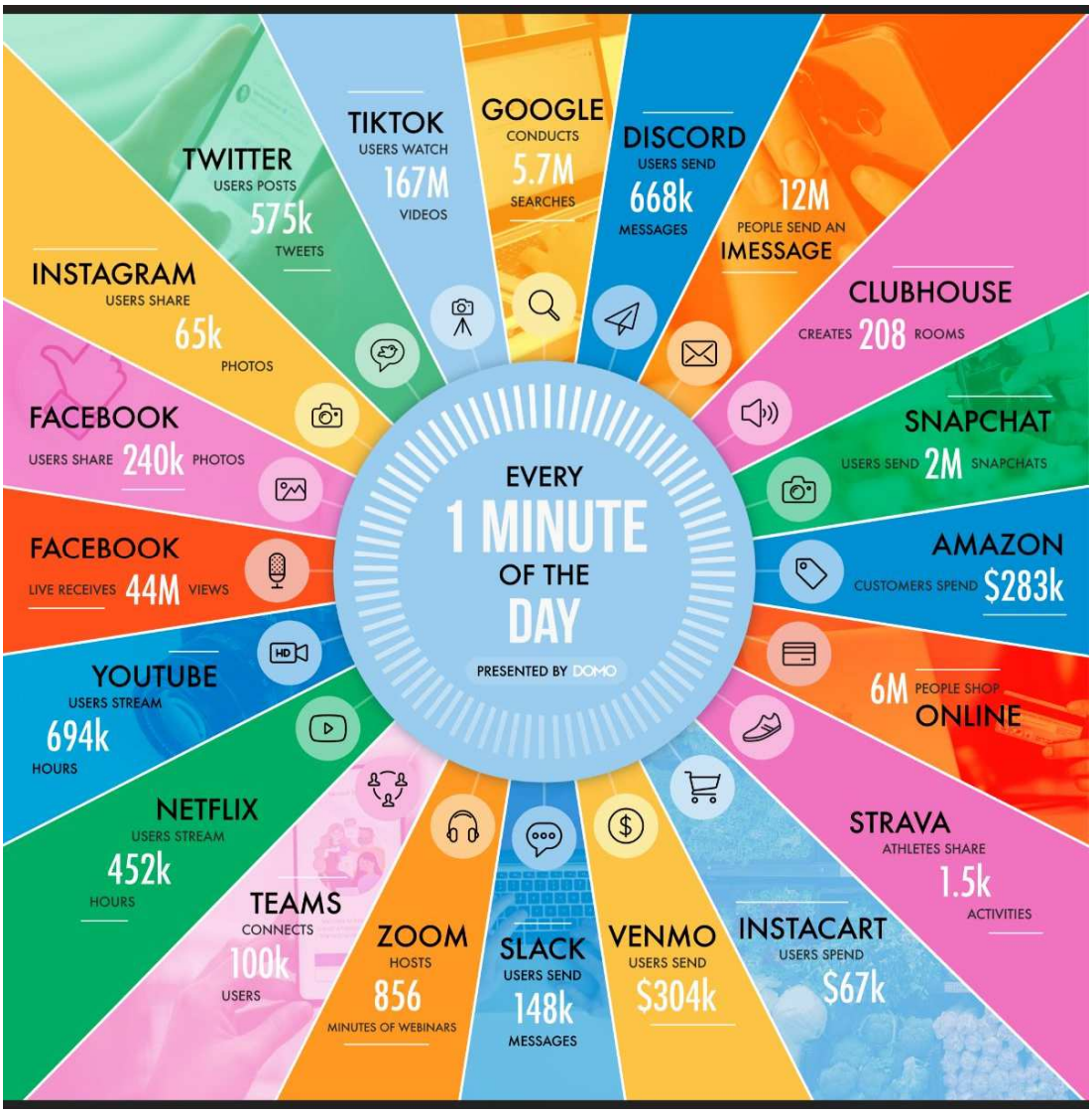
Les mégadonnées présentent trois caractéristiques clés, communément appelées 3Vs :

- Volume : Les lots de mégadonnées sont gigantesques, d'où le préfixe « méga ». Les termes souvent utilisés en référence au volume des mégadonnées sont « pétabytes » (ou pétaoctet) et « exabytes » (ou exaoctet) de données.
- Vélocité : Les mégadonnées sont caractérisées par la grande vitesse à laquelle elles se constituent (souvent de manière automatique, au fil des « clics » ou des transactions Web) et par le fait qu'elles sont en constant changement, reflétant une situation en temps réel.
- Variété : Il s'agit d'ensembles composés de données structurées (généralement du texte organisé dans des bases de données relationnelles traditionnelles) et de données non structurées (photos, vidéos, données textuelles).

Le recours aux mégadonnées pour mesurer la consommation des produits ou des contenus culturels numériques nécessite des infrastructures techniques, des capacités informatiques et une expertise en science des données différentes de celles utilisées pour la production de statistiques traditionnelles. En conséquence, elles représentent un défi pour les agences gouvernementales et les instances nationales ou internationales qui doivent produire des statistiques publiques sur la consommation de la culture, utiles à une diversité d'utilisateurs et s'appuyant sur une lecture objective et neutre de l'activité sur le Web. Un autre défi pour la production de statistiques publiques sur la culture numérique consommée vient du fait que les données d'usage sont souvent la propriété d'entreprises privées peu ouvertes à les partager sans frais.

Quantité de données générées sur Internet chaque minute (Source : www.domo.com)

[5] Rapport de la mission franco-québecoise sur la découvrabilité des contenus culturels francophones en ligne (2020) - Laticce, Etude ACEI 2018-2020 – « *Etat des lieux sur les métadonnées relatives aux contenus culturels* », Observatoire de la culture et des communications du Québec (2017)


EVERY
1 MINUTE
OF THE
DAY
PRESENTED BY DOMO
TWITTER
USERS POSTS
575k
TWEETS
TIKTOK
USERS WATCH
167M
VIDEOS
GOOGLE
CONDUCTS
5.7M
SEARCHES
DISCORD
USERS SEND
668k
MESSAGES
12M
PEOPLE SEND AN
IMESSAGE
CLUBHOUSE
CREATES 208 ROOMS
SNAPCHAT
USERS SEND 2M SNAPCHATS
AMAZON
CUSTOMERS SPEND $283k
6M PEOPLE SHOP
ONLINE
STRAVA
ATHLETES SHARE
1.5k
ACTIVITIES
INSTACART
USERS SPEND
$67k
VENMO
USERS SEND
$304k
SLACK
USERS SEND
148k
MESSAGES
ZOOM
HOSTS
856
MINUTES OF WEBINARS
TEAMS
CONNECTS
100k
USERS
NETFLIX
USERS STREAM
452k
HOURS
YOUTUBE
USERS STREAM
694k
HOURS
FACEBOOK
LIVE RECEIVES 44M VIEWS
FACEBOOK
USERS SHARE 240k PHOTOS
INSTAGRAM
USERS SHARE
65k
PHOTOS