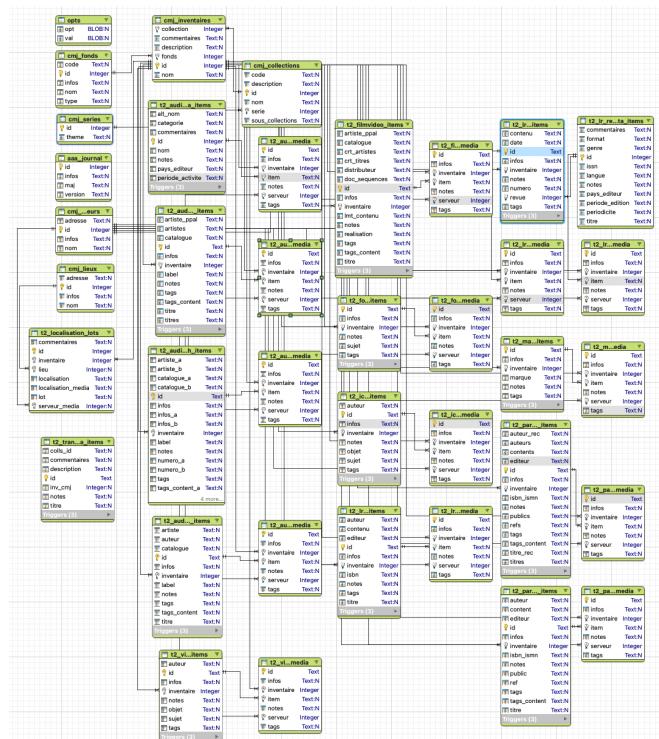


## Base de données CEMJazz2

## *Petit précis d'indexologie (appliquée au CEMJazz)*



## Base de données CEMJazz2 (vue partielle)



Cette nouvelle version (juillet 2025) du document initial d'octobre 2023 est cohérente avec la nouvelle structuration de la base de données CEMJazz2. Rappelons néanmoins que cette restructuration n'a pas affecté, sinon à la marge, la description des items indexés, rangés maintenant dans de nouvelles tables. L'indexation des items media (images, sons, ...), en revanche, a été unifiée et rationalisée.

## Sommaire

<b>1 Indexation des lots</b>	<b>5</b>
<b>2 Définir les champs</b>	<b>7</b>
2.1 Identification . . . . .	7
2.2 Tables Media . . . . .	8
2.3 Catalogues . . . . .	9
2.4 Attributs communs . . . . .	10
2.5 Description . . . . .	10
2.6 En pratique . . . . .	10
2.7 Tables dans CEMJazz2 . . . . .	11
<b>3 Les collections</b>	<b>11</b>
3.1 Série « Audio » . . . . .	12
3.1.1 Musique mécanique – Cylindres . . . . .	12
3.1.2 78t – Saphir – Edison diamant . . . . .	12
3.1.3 16t – 33t – 45t – Cds – Mds – Bmag – K7 . . . . .	13
3.1.4 Labels . . . . .	14
3.2 Série Films et Vidéos . . . . .	15
3.3 Série Iconographie . . . . .	16
3.4 Série Vidéographie . . . . .	17
3.5 Série Livres et revues . . . . .	17
3.5.1 Livres . . . . .	18
3.5.2 Revues . . . . .	18
3.6 Série Partitions . . . . .	20
3.6.1 Partitions, œuvres uniques . . . . .	20
3.6.2 Partitions, recueils . . . . .	21
3.7 Série Fonds documentaires . . . . .	21
3.7.1 Conférences, dossiers et papiers épars . . . . .	21
3.8 Série Matériel [de collection] . . . . .	22
3.9 Collections reconstituées . . . . .	22
<b>4 Media</b>	<b>23</b>
<b>5 Pour conclure</b>	<b>25</b>
<b>Table des documents</b>	<b>27</b>



CEMJazz 10 000 : [...] le 10 000<sup>e</sup> item indexé dans la base de données du CEMJazz est une bande magnétique pré-enregistrée Atlantic ALC 1468. Bobine 7", bande 1/4", 7 1/2 ips, 4-pistes stéréo (USA, 1966), disponible sous la référence mpf-bmag-bbn-pre-1 [...]



Doc. 1 – L'item mpf-bmag-bbn-pre-1

**L**E BUT premier de la base de données du CEMJazz est de collecter les résultats de tous les travaux, inventaires, indexations, ... effectués dans les différents fonds. Le but de ce petit document de travail est de présenter la façon dont ces données sont organisées dans la base de données. En effet, s'en inspirer facilite grandement par la suite le travail de collecte et permet de façonner des ensembles de données homogènes. Nous allons donc préciser *mais au fait... qu'indexe-t-on dans cette base?* Puis nous ferons ensuite un tour d'horizon des données relatives aux diverses collections déjà présentes dans la base.

## 1 Indexation des lots

Donc, la base CEMJazz2 sert à indexer des lots... Mais qu'est ce qu'un, lot ? Commençons par divers rappels, car tout cela a déjà été détaillé par le menu dans les documents de travail précédents.

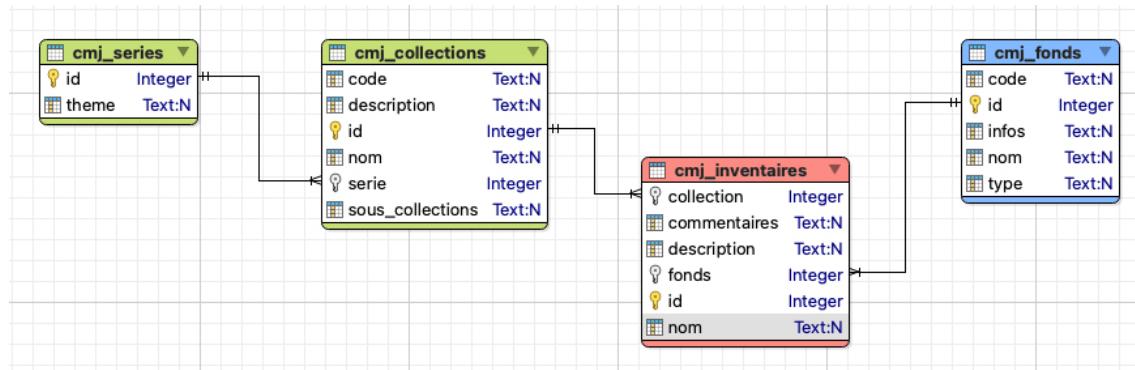
Rappelons le **plan de classement** utilisé pour l'indexation au CEMJazz. Nous distinguons, dans un premier temps, des **Séries** de documents, pour le moment (décembre 2025) huit sont clairement identifiées :

- Audio ;
- Films et Vidéos ;
- Iconographie ;
- Vidéographie ;
- Livres et Revues ;
- Partitions ;
- Fonds documentaires ;
- Matériel [de collection].

À l'intérieur de ces **Séries**, nous définissons des **Collections** permettant de regrouper des items, disons par affinités, ainsi : *les disques 33t, les DvDs, les documents photographiques, les livres, les partitions manuscrites*, et bien d'autres encore.

Mais, si chaque objet fait partie d'une **Collection**, il appartient aussi, de façon duale, à un **Fonds** provenant de l'un des donateurs du CEMJazz.

La démarche consiste alors à identifier une **Sous-collection** pertinente d'objets, disons homogènes, et, pour la reconstituer, à effectuer dans chaque **Fonds** un **Inventaire**. Ainsi : *les disques 33t de 25cm du fonds O. Brard (audio), les documents photographiques du fonds M. Pfau (iconographie), les partitions manuscrites du fonds P. Vernon, etc.*



Doc. 2 – La table des inventaires

En pratique, certaines sous-collections peuvent être affectées à un fonds fictif (notamment CEMJazz) et reconstituées dans un seul « meta-inventaire » à partir de plusieurs inventaires effectués dans les fonds réels. C'est typiquement le cas pour une sous-collection constituée de tous les exemplaires différents d'une revue donnée. Nous notons alors l'origine des différents items (la « mémoire » du CEMJazz) dans la définition du meta-inventaire avant élimination des doublons.

L'idéal est que l'**Inventaire** reste d'une taille raisonnable, typiquement au maximum quelques centaines d'items (?); dans le cas contraire, un pré-traitement voire un partitionnement s'impose !

L'inventaire est ensuite divisé en **Lots**. Les noms des lots peuvent être sans signification : *lots 1, 2, 3, ...*, ou désigner certaines catégories d'objets, par exemple pour des disques 33t où nous séparerons les albums (d'un seul artiste) des compilations : *lots al1, al2, al3, ... puis cp1, cp2, ...*

La granularité du **Lot** est liée au fait qu'il s'agit aussi d'une unité de rangement physique : *15 disques 78t, les numéros des années 1960 à 1965 d'une revue, ...* L'endroit où est rangé le lot après indexation est noté dans une table dédiée, ce qui permet de localiser ensuite les items. Le stockage



Doc. CEMJazz : obr-pub-mus-1-10

physique sera donc incohérent et inutilisable en pratique si les grands inventaires ne sont pas pré-triés.

L'**Inventaire** est bien le concept clef dans tout cela car il propose aussi une méthode de travail. Chaque **Inventaire** est déclaré et décrit dans la base de données, en explicitant éventuellement les **Lots** spécifiques envisagés : *Fonds x. X., disques microsillon 33 tours de 30cm (ou 12"). Lots : sp (spéciaux), al (album avec interprète principal), cp (compilation), did (didactiques), ....*

Une fois tout ce travail d'identification, de tri, de classement, de partitionnement (et autres) effectué, il reste à indexer les **Lots**. Ce petit précis parle donc tout simplement de l'indexation d'un **Lot** dans la base de données du CEMJazz.

## 2 Définir les champs

L'idéal, côté base de données, est de limiter le nombre des champs définissant les objets dans les tables. Ce n'est pas toujours simple. Par exemple, si nous nous intéressons à la table des livres, nous placerons spontanément une colonne `titre` et une colonne `auteur`, mais s'il y a deux auteurs il faudra peut-être envisager `auteur_1` et `auteur_2`? que faire alors s'ils sont trois et s'il y a un traducteur, *etc*. L'organisation des tables ne peut pas être (trop) évolutive, il faut donc définir une colonne `auteur` qui recevra toutes les informations concernant l'origine de l'œuvre. La même problématique se pose, par exemple, pour une partition, nous avons le compositeur, l'arrangeur, le parolier original, les traducteurs éventuels des paroles dans d'autres langues, *etc*. Là aussi, `auteur` est une colonne générique destinée à recevoir toutes ces informations.

Cette approche est valable pour tous les attributs définissant un objet dans la base de données comme cela sera détaillé par la suite, collection par collection. Elle est particulièrement adaptée au contexte SQL. Dans les tables, la recherche se fait en mode plein texte et les données rentrées par les requêtes SQL qui parcourent ces tables peuvent être restructurées en fonction des informations recherchées. Il est primordial de savoir *où chercher quoi*, peu importe la forme exacte pourvu qu'il n'y ait pas trop de *où!* ... Nous essayons aussi de réutiliser ou d'adapter *a minima* les schémas de données pour les collections et nous utilisons le plus possible les mêmes attributs avec la même sémantique.

La plus value, si l'on peut dire, d'une base de données est dans la structure, pas dans les données elles-mêmes.

### 2.1 Identification

Chaque item indexé au CEMJazz reçoit un numéro unique, forgé à partir du nom de l'**Inventaire** et de l'identifiant du **Lot** auquel il appartient. Dans toutes les tables, l'identifiant **id**, indiqué ici en rouge, est fabriqué par le système à partir de ces informations. Pour la personne qui réalise l'indexation, à travers un formulaire ou directement sur un fichier Excel ou autre, il s'agit juste de donner l'identifiant du **Lot** et le numéro d'ordre de l'item dans le **Lot**, l'identifiant peut être déduit automatiquement de ces informations et du nom de l'inventaire.

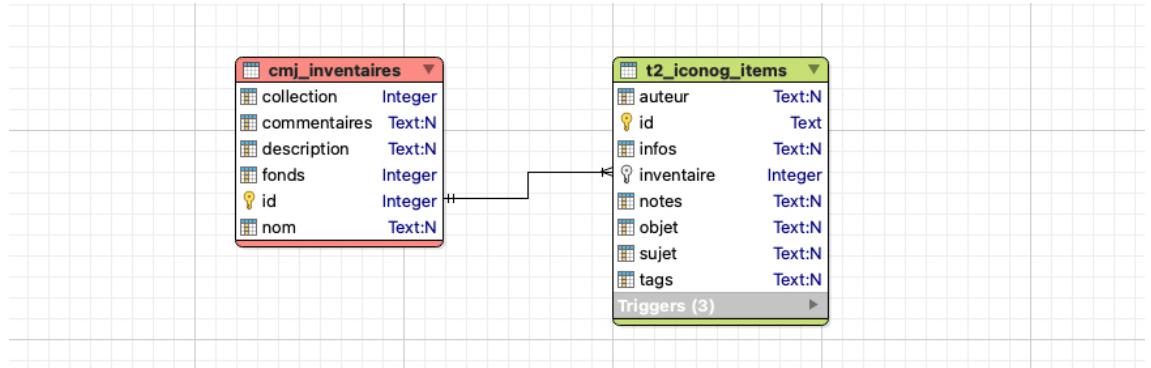
À cet effet, chaque item comprend, dans les tables un enregistrement appelé (judicieusement) **inventaire**. Il s'agit, ici, de l'**id** de l'**Inventaire** vu comme un objet de la table correspondante dans la base de données. Par exemple l'**Inventaire** des disques 33t de 17cm du fonds Jean Pierre Dauresse, *alias jpd-33t-17cm* a pour **id** 54. Cette information n'est nécessaire que pour importer les données dans la base. Elle est bien sûr identique pour tous les items de tous les lots d'un même **Inventaire**.

On comprend qu'il n'y a pas trop à s'occuper de toutes ces affaires et que l'item **13** du lot **a12** de l'inventaire **jpd-33t-17cm** recevra un identifiant ressemblant à **jpd-33t-17cm-a12-13**, après les vérifications d'usage. Le module de saisie en ligne effectue cela automatiquement, sinon une « moulinette » fait l'affaire.

Les champs notés en rouges sont des **clefs primaires**, ainsi **id**, identifiant unique d'un item dans une table. Un champ orange, comme **inventaire**, est une clef **étrangère**, identifiant unique d'un item dans une autre table



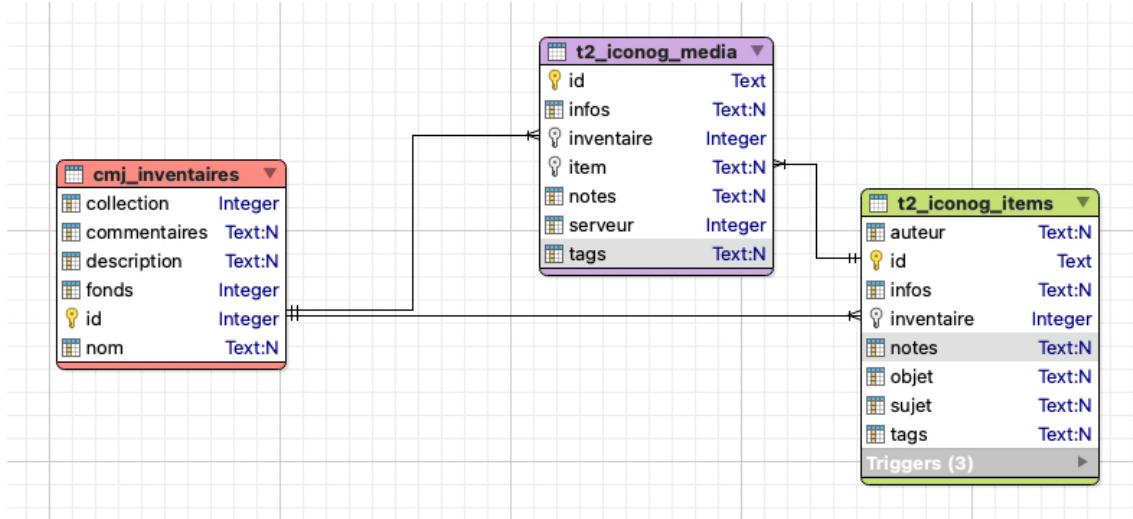
Doc. CEMJazz : obr-pub-mus-1-9



Doc. 3 – La table des items « iconographie » reliée à son inventaire

## 2.2 Tables Media

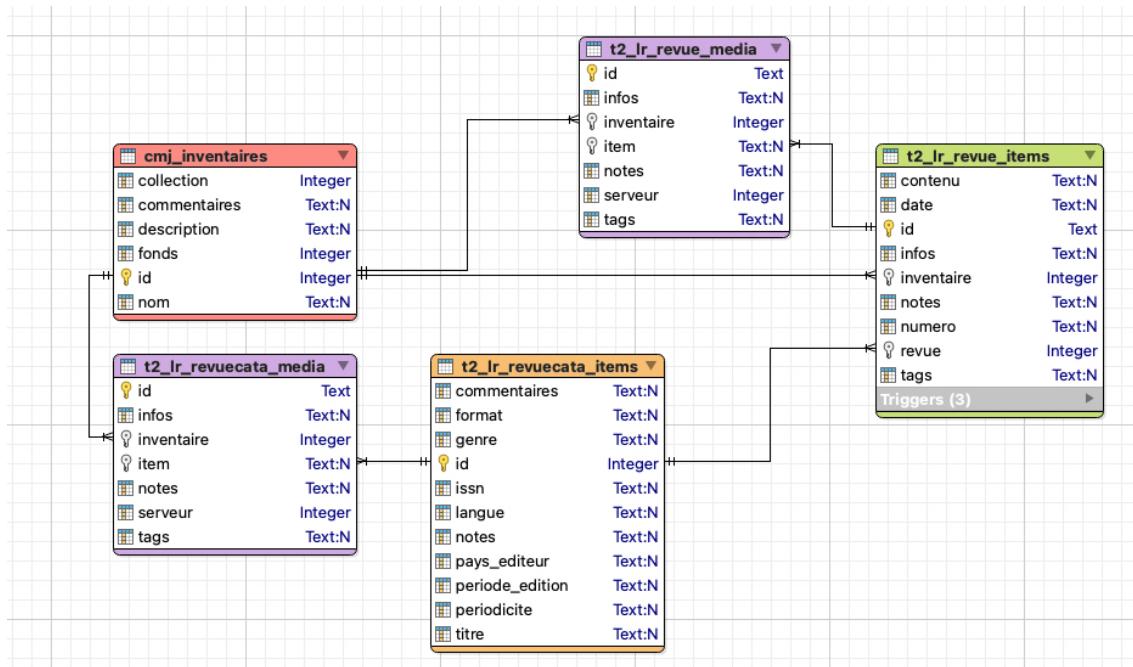
À chaque table « items » de la base est associée une table « media ». Ces tables ont toutes la même structure. Un média référence un unique item dans la table items correspondante (clef étrangère) et appartient au même inventaire (clef étrangère). Toutes les tables media fonctionnent donc de la même façon, nous en verrons quelques exemples par la suite. Un media n'existe que par l'item référencé et son identifiant n'est pas forgé selon les règles des items « comuns ».



Doc. 4 – La table des items « iconographie » et sa table media

### 2.3 Catalogues

certaines collections sont associées à des catalogues. C'est typiquement le cas des revues. Le catalogue n'est pas associé à un inventaire, c'est une donnée disons « primaire » de la base, comme la liste des séries ou des fonds. Chaque item revue sera associé à un titre dans le catalogue. Comme précédemment, des media peuvent être associés à un numéro précis d'une revue, mais aussi, de façon générique au titre (cf catalogue visuel des revues).



Doc. 5 – La table des items « revue », catalogue et tables media

## 2.4 Attributs communs

Tout item ou media dispose des attributs (codés en bleu) **infos**, **tags** et **notes**, qui seront complétés par des attributs plus précis selon la (sous) collection considérée.

- **infos** renseigne sur l'objet lui même, description physique, état...
- **tags** concerne la vie de l'objet avant le CEMJazz, par exemple *France, 1938, ..., photo dédicacée, ...*
- **notes** concerne la vie d'objet au Cemjazz, pour indiquer, par exemple, qu'il est rangé à un endroit hors norme (accroché au mur, associé à un autre objet, accessible sur un serveur exotique, ...).

## 2.5 Description



Doc. CEMJazz : obr-pub-mus-1-12

Voici maintenant les champs qui, eux, s'intéressent aux « contenu(s) » porté(s) par l'objet en tant que support : *qui ? quoi ? pour qui ? etc.* nous trouverons donc (ce sont des champs verts) des choses comme **titre**, **artiste**, **objet**, **sujet**, **auteur**, **public**, ... À tous ces contenus sont aussi associés des tags, souvent sous la forme **tags\_content** qui permettent de mieux qualifier les contenus ainsi, pour rester dans le registre musical, *jazz*, *dixieland*, ..., *enregistré en 1929*, ..., *musique du film "XXX"*, ...

## 2.6 En pratique

Lors de la saisie d'un lot, **id** et **inventaire** sont des affaires réglées une fois pour toute. Il reste, pour chaque item, à le décrire physiquement avec **infos** et à regarder un peu son histoire : les **tags**. Les **notes** restent généralement vides dans un premier temps et sont plutôt à visée interne.

L'on peut alors s'intéresser aux contenus, par exemple titre, artiste, et autres. Cela posé, nous pouvons maintenant (enfin !) commencer notre petit tour des tables associées aux collections.

Pour avoir une bonne idée de la façon dont tout cela fonctionne, le mieux est d'interroger la base de données du CEMJazz *via* le site. Attention néanmoins, les données sont retournées sous une forme compacte qui ne liste que des champs réputés pertinents pour les visiteurs du site.

## 2.7 Tables dans CEMJazz2

Les noms des tables indiquent la série et la ou les collection(s) concernée(s); « la ou les » car certaines bénéficient de descriptions identiques. À chaque table *items* est associé une table *media*. Nous voyons aussi, dans cette liste, les catalogues.

Table	t2_audio_labcata_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	205
Table	t2_audio_labcata_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	464
Table	t2_audio_mu_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	1879
Table	t2_audio_mu_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	121
Table	t2_audio_rl_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	10
Table	t2_audio_rl_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	0
Table	t2_audio_sh_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	3356
Table	t2_audio_sh_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	135
Table	t2_filmvideo_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	130
Table	t2_filmvideo_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	0
Table	t2_fondoc_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	118
Table	t2_fondoc_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	162
Table	t2_icong_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	702
Table	t2_icong_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	1153
Table	t2_localisation_lots	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	594
Table	t2_lr_livre_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	93
Table	t2_lr_livre_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	17
Table	t2_lr_revue_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	1213
Table	t2_lr_revue_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	14
Table	t2_lr_revuecata_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	180
Table	t2_lr_revuecata_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	246
Table	t2_materiel_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	9
Table	t2_materiel_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	0
Table	t2_partition_rec_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	23
Table	t2_partition_rec_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	0
Table	t2_partition_uni_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	122
Table	t2_partition_uni_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	3
Table	t2_trans_cata_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	10
Table	t2_videog_items	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	7
Table	t2_videog_media	Naviguer	Structure	SQL	Rechercher	Insérer	Exporter	Importer	Renommer	Vider	Supprimer	17

Doc. 6 – Les collections (items, media, catalogue) dans la base CEMJazz2

## 3 Les collections

Classées ici par **Séries**. Sauf cas particulier, tout item présent dans la base de données est vu sous deux angles, le support matériel et le message visuel, sonore, ... qu'il porte, ce qui correspond, respectivement, aux champs bleus et verts de la description comme expliqué ci-dessus..

Aucun item n'échappe à son identifiant unique (rouge) ni à son code inventaire (orange); toutefois, les entrées des catalogues, qui ne sont pas des items matériels, n'ont pas de référence inventaire.

Enfin, rappelons que les notes (en bleu) décrivent « physiquement », la vie de l'item avant le CEMJazz, puis au CEMJazz.

Le fonctionnement des tables media est présenté, à titre d'exemple, dans le cadre des labels (cf 3.1.4) puis en détail à la fin de ce document (cf 4).

### 3.1 Série « Audio »

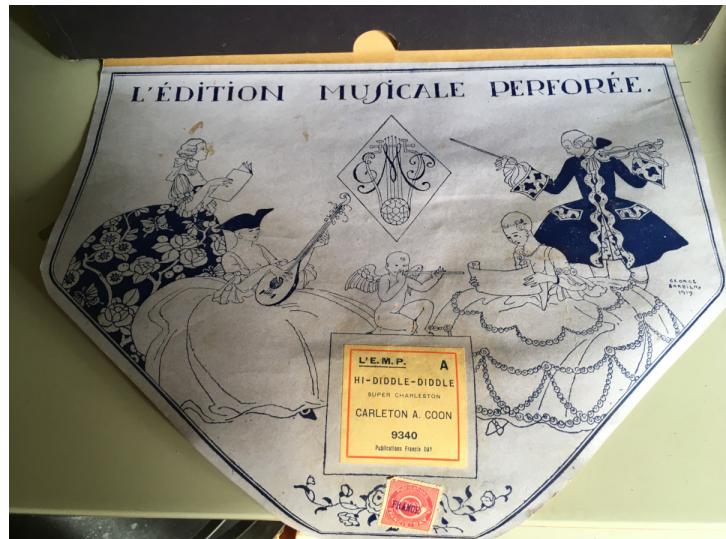
L'analyse des différentes collections de la série Audio, fait apparaître trois descriptions possibles des items et de leurs contenus. certaines sont ainsi regroupées dans une même table.

#### 3.1.1 Musique mécanique – Cylindres

Ces items ne comportent qu'un message sonore. Pour la musique mécanique, le champ `artiste` n'est généralement pas pertinent mais l'objet à pu être conçu d'après une interprétation spécifique de l'œuvre. À noter, dans notre description, si le `label` est associé à l'objet physique le champ `catalogue` est, lui, associé à l'objet sonore.

En plus des classiques `id`, `inventaire` puis `infos`, `tags`, `notes`, les données à saisir sont `label` pour l'objet matériel et `catalogue`, `titre`, `auteur`, `artiste`, `tags_content` pour l'objet sonore. Ce dernier champ permet de préciser le genre musical, l'année d'enregistrement ou de composition, ... : la vie de l'objet musical.

<code>id</code>	
<code>label</code>	
<code>catalogue</code>	
<code>infos</code>	
<code>tags</code>	
<code>titre</code>	
<code>auteur</code>	
<code>artiste</code>	
<code>tags_content</code>	
<code>inventaire</code>	
<code>notes</code>	



Doc. 7 – Structure de données Audio « Rouleaux » (11 champs)

#### 3.1.2 78t – Saphir – Edison diamant

Les trois collections disques 78t, disques saphir et disques Edison Diamant, partagent la même définition. Sur les disques anciens, les deux faces n'ont souvent aucun rapport entre-elles, parfois chacune porte même un numéro de catalogue différent (particularité héritée des cylindres...).

Nous avons bien affaire, exceptionnellement, à trois objets en un : le disque lui même et deux objets sonores à décrire indépendamment.

Les champs `catalogue_a` et `catalogue_b` font référence au label. Généralement, un autre identifiant (`numero_a`, `numero_b`) est associé au morceau par rapport à un autre catalogue de titres (du label ou d'un autre éditeur). Enfin d'autres informations peuvent être gravés sur les faces du disques (numéro de matrice, ...) qui sont généralement des

détails propres aux documents sonores et qui seront alors relevés dans les champs `infos_a` et `infos_b` et non dans `infos`.

Les `tags_content` propres à chaque faces, servent là aussi à préciser le genre musical, l'année d'enregistrement et même les noms des musiciens si ceux-ci sont précisés sur l'étiquette, ce qui est parfois le cas sur des disques récents ou produits par des labels spécialisés (bien sûr dans le Jazz...).

<code>id</code>	
<code>label</code>	
<code>infos</code>	
<code>tags</code>	
<code>titre_a</code>	
<code>artiste_a</code>	
<code>catalogue_a</code>	
<code>numero_a</code>	
<code>infos_a</code>	
<code>tags_content_a</code>	
<code>titre_b</code>	
<code>artiste_b</code>	
<code>catalogue_b</code>	
<code>numero_b</code>	
<code>infos_b</code>	
<code>tags_content_b</code>	
<code>inventaire</code>	
<code>notes</code>	



Doc. 8 – Structure de données Audio « Shellac » (18 champs)

À noter : cette table est la première à avoir été définie en commun au CEMJazz à la fin de l'année 2021, lors des premiers tests avec une base de données.

### 3.1.3 16t – 33t – 45t – Cds – Mds – Bmag – K7

Les microsillons (16t, 33t, 45t), les Discs (Cds et Mds) ainsi que les bandes magnétiques (bobines ou cassettes) partagent la même définition. Nous n'avons plus que deux objets (cas normal...), le support et l'objet sonore.

Outre le catalogue, la description de l'objet comprend `titre` et `artiste principal` (`artiste_ppal`), en général ce qui est lu sur la pochette, puis `titres` et `artistes` qui permettent de noter les pistes et les musiciens. Si le disque est une compilation (et non un pot-pourri d'un même musicien), ce mot apparaîtra à la place de l'artiste principal.

Nous avons choisi de ne pas noter systématiquement les informations gravées sur les disques (cela peut être fait, ponctuellement dans la rubrique [infos](#)) en effet, tous ces items sont beaucoup plus normalisés que les disques 78t, Saphir, Edison diamant et ces informations sont de moindre intérêt. Pour une recherche précise, il est toujours possible de scruter à la loupe les objets.

L'état actuel du travail sur les bandes magnétiques n'a pas permis à ce jour (juillet 2025) de décider s'il fallait créer une nouvelle série où ranger des audiogrammes (sur bandes magnétiques) qui, à l'instar de la série vidéographie, seraient réalisés dans un cadre amateur, non édités, non montés et de nature documentaire.

<a href="#">id</a>	
<a href="#">label</a>	
<a href="#">catalogue</a>	
<a href="#">infos</a>	
<a href="#">tags</a>	
<a href="#">titre</a>	
<a href="#">artiste_ppal</a>	
<a href="#">titres</a>	
<a href="#">artistes</a>	
<a href="#">tags_content</a>	
<a href="#">inventaire</a>	
<a href="#">notes</a>	



Doc. 9 – Structure de données Audio « Musique » (12 champs)

### 3.1.4 Labels

Les tables « label », regroupent le catalogue et les images des étiquettes des labels rencontrés lors de l'indexation des disques catalogués « Jazz <sup>1</sup> ». Il s'agit d'une collection transverse pour le moment limitée aux disques dits « shellac », c'est à dire les 78t (au sens large), les disques à saphir et les Edison diamant.

L'indexation des labels repose sur l'utilisation d'un catalogue. L'indexation des media se fait en reconstituant des inventaires similaires à ceux des disques (« 78t 25cm », « 33t 30cm », ...) affectés à un fonds fictif **media**. Ces deux tables servent donc essentiellement à référencer des collections d'images.

Le **Catalogue des labels** comporte des champs nécessaires à une description sommaire : [catégorie](#) (shellac, microsillons, CDs, *etc.*), [nom](#) du label, [noms alternatifs](#), [pays](#) (il peut y en avoir plusieurs), [période d'activité](#) et [commentaires](#) variés. Le champs [notes](#) recueille ici (par exemple) des liens internet permettant d'approfondir le sujet.

À noter : il n'y a pas d'[inventaire](#) (c'est un catalogue) ; il n'y a pas non plus de champs [infos](#) et [tags](#) un item du catalogue n'est pas un objet matériel.

1. Avec quelques exceptions pour les disques plus exotiques, décorés, flexibles et autres...

Les **media** associés sont les avatars en image des entrées du catalogue, c'est à dire les différentes variantes des étiquettes gravées ou collées sur les disques. Par convention, l'**id** d'un media est le nom du fichier correspondant (nom forgé selon des règles bien précises).

La description d'un media renvoie, bien sûr à un élément du catalogue (**id** de l'item représenté qui est une clef externe), à un **inventaire** et à un **serveur** où l'image est stockée (clef externe), les champs **infos** et **tags** se réfèrent uniquement à l'avatar (objet matériel) photographié, le champ **notes** peut servir à commenter le media lui même (par exemple Photo CEMJazz...).

<b>id</b>	
<b>nom</b>	
<b>categorie</b>	
<b>alt_nom</b>	
<b>pays_editeur</b>	
<b>periode_activite</b>	
<b>commentaires</b>	
<b>notes</b>	

**Inventaires "Labels" dans la série Audio**

Labels 78t, Saphir et Edison Diamant (classés par noms de label)



“Mini base de données” de travail permettant de retrouver des informations sur les labels des disques à saphir (80t), des disques 78t (aiguille et quelques vinyles) et Edison diamant. Cette base présente des images de tous les labels rencontrés au CEMJazz lors de l’indexation de ces disques. Elle contient un minimum d’information (typiquement: pays, période d’activité, ...) sur chaque label mais renvoie vers des sites où tout cela est étudié ou listé de façon plus détaillée.

La rubrique **ressource:labels** présente toutes les photographies des labels inventoriés selon les caractéristiques des disques (saphir, aiguille, format, ...)

Choisir un Label dans le menu ci-dessous

Label \* AEROPHONE OK

Doc. 10 – Structure de données Catalogue des labels (8 champs)

Toutes les tables média de la base CEMJazz2 ont exactement la même structure, nous y reviendrons (cf 4).

<b>id</b>	
<b>item</b>	
<b>infos</b>	
<b>tags</b>	
<b>inventaire</b>	(spécifique media)
<b>serveur</b>	
<b>notes</b>	



Doc. 11 – Structure de données Catalogue des labels, media (7 champs)

### 3.2 Série Films et Vidéos

Dans cette série, nous trouvons des DVDs, des VHS et des Films 16mm. Le terme **distributeur**, que nous utilisons est plus adapté dans ce contexte que celui de **label**, mais il s'agit bien (plus ou moins) de la même chose.

Nous verrons qu'il existe aussi une série vidéographie (pour les vidéogrammes). Si toutes les vidéos et les films sortis dans un circuit professionnel sont concernés par la série Films et Vidéos, la frontière est parfois ténue entre le document amateur et un produit fini qui n'a pas été édité. Beaucoup de nos films 16mm — produits finis, qui seront notées sans **distributeur** et sans (numéro de) **catalogue** — rentrent bien dans cette série Films et Vidéos.

Côté contenu le **réalisateur** fait son apparition ; l'indexation prévoit des champs, à remplir au choix, pour trois types de documents rencontrés au CEMJazz : des concerts filmés (**crt\_titres**, **crt\_artistes**), des long métrages (**lmt\_contenu**) et des documentaires (**doc\_sequences**). Cela semble aller à l'encontre du concept de champ générique, mais à l'usage ce dernier s'est, pour une fois révélé, peu efficace.

<b>id</b>	
<b>distribution</b>	(cf label)
<b>catalogue</b>	
<b>infos</b>	
<b>tags</b>	
<b>realisation</b>	
<b>titre</b>	
<b>artiste_ppal</b>	
<b>crt_titres</b>	(concert)
<b>crt_artistes</b>	
<b>lmt_contenu</b>	(long métrage)
<b>doc_sequences</b>	(documentaire)
<b>tags_content</b>	
<b>inventaire</b>	
<b>notes</b>	



Doc. 12 – Structure de données Films et Vidéos (15 champs)

### 3.3 Série Iconographie

Pour cette série, les données à saisir pour la partie objet visuel varient selon le type de document. Trois types d'informations peuvent être saisies pour un document : son **auteur** (pas toujours indiqué selon le type de document), son **objet** (pas toujours pertinent hors affiches ou publicité, ...), son **sujet** (le message visuel).

Une table commune est utilisée pour l'indexation des documents iconographique, toutefois, dans la pratique nous privilégiions :

- pour les collections Documents Photographiques et Œuvres Graphiques, le couple « **auteur** / **sujet** » ;
- pour les collections Programmes et Publicités, le couple « **objet** / **sujet** » ;
- et pour les collections Affiches les trois attributs « **auteur** / **objet** / **sujet** ».

id	
objet	
auteur	
sujet	
infos	
tags	
inventaire	
notes	



Doc. 13 – Structure de données Iconographie (8 champs)

### 3.4 Série Vidéographie

Cette série regroupe des documents disons « bruts » sous forme de cassettes vidéos Hi-8 ou VHS et de DVD voire des items n'existant que sous forme numérique ; nous avons aussi, au CEMJazz, quelques films en 8mm et en 9.5mm dont le contenu reste encore à explorer.

La description des items est somme toute identique à ceux de la série iconographie mais une table séparée est jugée préférable car il s'agit bien de deux séries différentes. Une collection Audiogramme, évoquée ci-dessus, aurait sans doute la même structure.

id	
objet	
auteur	
sujet	
infos	
tags	
inventaire	
notes	



Doc. 14 – Structure de données Vidéographie (8 champs)

### 3.5 Série Livres et revues

Dans la série Livres et Revues, les « `tags_contenu` » habituels sont remplacés par un attribut appelé tout simplement « `contenu` ».

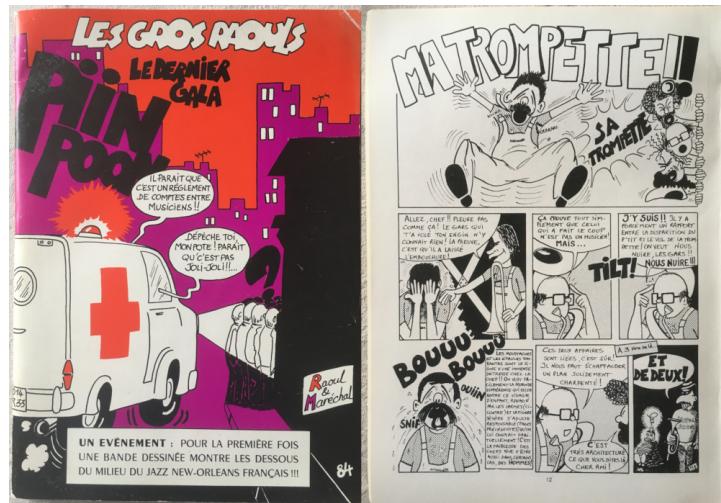
Pour un livre, cet attribut possède au moins la fonction de son prédécesseur, mais rien n'empêche d'en dire un peu plus et de décrire, fut-ce brièvement de quoi il s'agit.

Pour les revues, les « `tags_contenu` » au sens habituels sont considérés comme communs à tous les numéros d'une même revue. Ils sont donc explicités dans plusieurs champs du catalogue [des revues] pour le titre correspondant. L'attribut `contenu` sert bien, lui, à décrire le contenu du numéro indexé.

### 3.5.1 Livres

La description des livres comprend aussi les champs **titre**, **auteur**, **éditeur** et **lisbn** (pour les livres récents...). Tout le reste fonctionne comme d'habitude.

<b>id</b>	
<b>titre</b>	
<b>auteur</b>	
<b>éditeur</b>	
<b>isbn</b>	
<b>contenu</b>	
<b>infos</b>	
<b>tags</b>	
<b>inventaire</b>	
<b>notes</b>	



Doc. 15 – Structure de données Livres (10 champs)

### 3.5.2 Revues

Rappelons que les revues fonctionnent avec un catalogue. L'indexation se fait après la reconstitution de la sous-collection complète de tous les numéros disponibles d'une revue donnée. Cela se fait *a priori* en trois étapes (mais peut aussi être effectué au fur et à mesure de l'inventaire des fonds).

1. Ajout du titre au catalogue (informations génériques pour tous les exemplaires de la revue).
2. Recueil des différents items, en notant la liste des fonds « contributeurs ».
3. Élimination des doublons, puis création d'un inventaire *ad hoc* pour la revue, affecté au fonds fictif « CEMJazz ».
4. Et enfin... indexation des numéros.

Il y a donc deux tables définies dans la base de données du CEMJazz pour la collection des revues, et deux tables media associées, voir le schéma présenté plus haut Doc. 5.

### Catalogue des titres

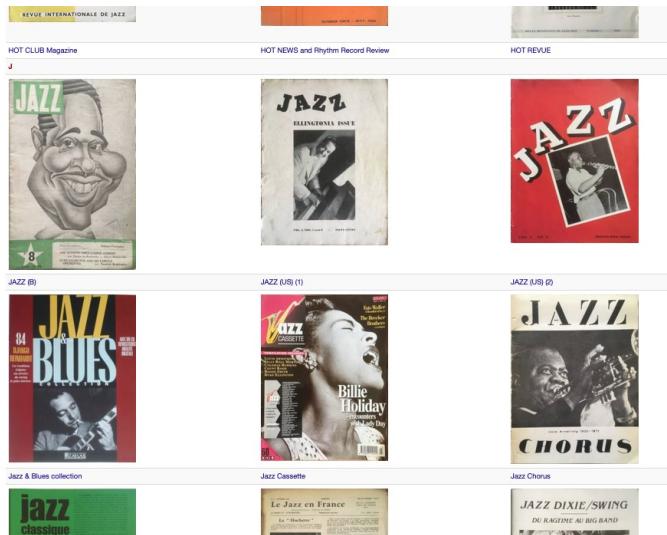
Nous essayons de définir en quelques rubriques le genre et la vie de la revue ; nous n'oublions pas l'**isbn** (souvent attribué rétroactivement). Ici les **commentaires** servent à noter ce que l'on peut éventuellement dire de plus sur la revue (prend la succession d'un autre titre ; la publication cesse puis reprend, ...). Comme vu précédemment, le catalogue ne possède pas de champ **inventaire**, ni les attributs **infos** et **tags**.

## Revues (Items)

Chaque item revue possède une clef étrangère **revue** (l'**id** du titre dans le catalogue). Bien sûr, **infos**, **tags** et **notes** se réfèrent uniquement au numéro indexé et ces champs peuvent souvent être vides.

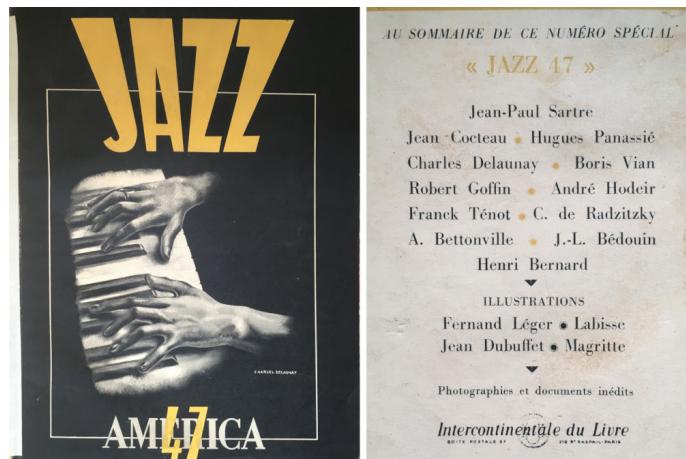
Les lots de revues indexés peuvent être constitués par années ou par volumes. Toutefois, si la revue est éditée par volumes, les numéros repartent à 1 à chaque volume, l'idéal est alors bien de créer un lot par volume.

<b>id</b>	
titre	
genre	
pays_editeur	
issn	
langue	
format	
periodicite	
periode_edition	
commentaires	
notes	



Doc. 16 – Structure de données Catalogue des revues (11 champs)

<b>id</b>	
<b>revue</b>	
<b>numero</b>	
<b>date</b>	
<b>contenu</b>	
<b>infos</b>	
<b>tags</b>	
<b>inventaire</b>	
<b>notes</b>	



Doc. 17 – Structure de données Revues (9 champs)

## 3.6 Série Partitions

Dans le plan de classement du CEMJazz, les partitions se répartissent en deux collections : **manuscrites** et **imprimées**. Les termes « imprimée » ou « manuscrite » doivent plutôt être lus dans le sens « éditée » ou « non éditée », un item manuscrit indexé peut tout à fait être un tapuscrit. Toutefois, pour ne pas multiplier les structures nous avons privilégié, dans la mise en œuvre, une approche **œuvres uniques** versus **recueils**.

C'est cette dernière analyse qui a conduit à la création des tables. Les inventaires sont donc forgés comme **pve-puni-man**, *Fonds Paul Vernon, partitions « œuvre unique »*, *manuscrites*, ... ou encore **obr-prec-imp**, *Fonds Olivier Brard, partitions « recueil »*, *imprimées*... avec en premier l'identifiant des tables.

Nous utilisons un champ **ref** dans le sens « interprète de référence », cette précision peut être utile pour retrouver des œuvres rendues célèbres (par tel artiste ou tel orchestre connu) mais dont les compositeurs ou paroliers sont restés un peu dans l'ombre. Un autre champ **public** permet d'indiquer simplement à qui s'adresse cette partition : chanteur, instrumentiste (pianiste, ...) ou orchestre (sextet, octet, big band, ...). Les différentes parties présentes étant bien sûr détaillées dans le contenu.

Pour les partitions manuscrites, le champ **éditeur** sera judicieusement remplacé par la mention « Partition manuscrite ». Les partitions éditées reçoivent normalement un numéro **isbn** mais les recueils sont souvent gratifié d'un **isbn**.

### 3.6.1 Partitions, œuvres uniques

Comporte donc les champs **titre**, **auteur**, **ref**, **éditeur** et **isbn\_isnm**, puis **public**, **contenu** et **tags\_content**.

<b>id</b>	
<b>titre</b>	
<b>auteur</b>	
<b>ref</b>	
<b>éditeur</b>	
<b>isbn_isnm</b>	
<b>infos</b>	
<b>tags</b>	
<b>public</b>	
<b>contenu</b>	
<b>tags_content</b>	
<b>inventaire</b>	
<b>notes</b>	



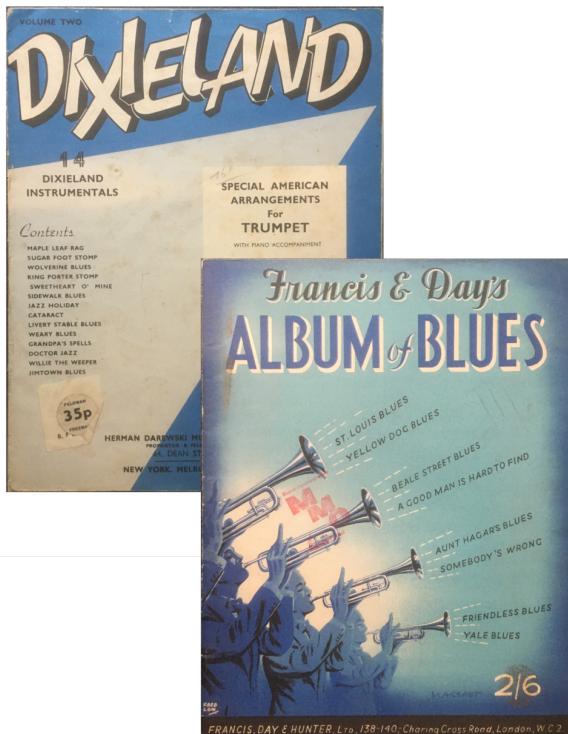
Doc. 18 – Structure de données Partition « œuvre unique » (13 champs)

### 3.6.2 Partitions, recueils

Nous avons là un autre exemple d'objet polymorphe. Pour le contenu nous décrivons la structure englobantes (le recueil) avec les champs `titre_rec`, `auteur_rec`, `editeur` et `isbn_isnm`, puis autant d'objets musicaux que de partitions présentes.

Pour cela, nous retrouvons au pluriel, tous les attributs d'un item œuvre unique qui contiennent cette fois des listes à la manière des titres d'un disque et de leurs interprètes, `titres`, `auteurs`, `refs`, `publics`, `contenus` et `tags_content`.

<code>id</code>	
<code>titre_rec</code>	
<code>auteur_rec</code>	
<code>editeur</code>	
<code>isbn_isnm</code>	
<code>infos</code>	
<code>tags</code>	
<code>titres</code>	
<code>auteurs</code>	
<code>refs</code>	
<code>publics</code>	
<code>contenus</code>	
<code>tags_content</code>	
<code>inventaire</code>	
<code>notes</code>	



Doc. 19 – Structure de données Partition « recueil » (15 champs)

## 3.7 Série Fonds documentaires

Dans cette série trois collections sont identifiées pour le moment (juillet 2025) « dosiers », « conférences » et « papiers épars ». Les champs importants ici sont `auteur`, `sujet` et `contenu`.

Une seule table est utilisée, toutefois, pour les collections Papiers épars nous ne retiendrons généralement que le champ `sujet`.

### 3.7.1 Conférences, dossiers et papiers épars

Les conférences et dossiers peuvent être « éclatés » en plusieurs items, notamment lorsqu'il ya des documents photographiques ou autres. Le champ `auteur` est toujours renseigné avec l'auteur du dossier ou de la conférence mais d'autres noms peuvent être bien sûr être cités dans le `contenu`.

id	
auteur	
sujet	
contenu	
infos	
tags	
inventaire	
notes	



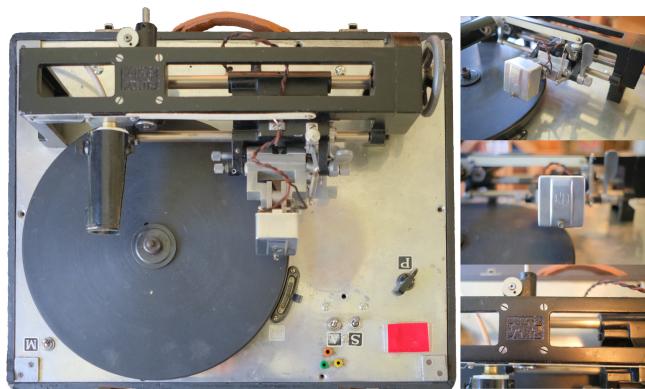
Doc. 20 – Structure de données Fonds documentaires (8 champs)

### 3.8 Série Matériel [de collection]

Dans cette série trois collections sont identifiées pour le moment (juillet 2025) « matériel audio », « matériel vidéo » et « divers ».

Un seul champ identifie le contenu, **marque**, le reste est naturellement décrit dans le champ **infos**.

id	
marque	
infos	
tags	
inventaire	
notes	



Doc. 21 – Structure de données Matériel de collection (6 champs)

### 3.9 Collections reconstituées

Dernier point, des sous-collections reconstituées transversales visent à mettre en évidence des items appartenant éventuellement à plusieurs collections mais logiquement reliés par une histoire ou une origine commune.

Comme exemples « intra-collection », citons les disques Swing (78t 25cm ou 30cm) ou les V-Disc (78t 30cm) qui se retrouvent tous dans des inventaires différents. Une façon simple de les relier entre eux consiste à former des lots spéciaux, il est alors facile de retrouver (dans la base de données) tous les disques Swing car ils possèdent tous, dans leur identifiant, un nom de lot commençant par **swing**, ainsi : **obr-78t-30cm-swing7-91, mpf-78t-25cm-swing6-78**, ... Dans la base... mais pas seulement car le lot est aussi

une unité de rangement. Une fois tous les disques Swing dénichés dans les divers fonds, il sera possible de faire mieux, par exemple (au moins) éliminer les doublons.

Mais l'affaire peut aussi concerner plusieurs collections, par exemple les disques du label AFAJNO qui se présentent en 33t et en 45t. Là aussi, des noms de lots spécifiques permettent de les repérer facilement : **obr-45t-17cm-afajno-2**, **jpd-33t-17cm-afajno-4**, dans chaque sous-collection.

Pour faire mieux, il suffit de travailler dans chaque collection comme avec les revues. Bien noter, l'origine des différents items puis supprimer les doublons. L'ensemble des items peut alors être relié à un inventaire fictif attribué au CEMJazz. À l'instar des revues, ce sont ces inventaires qui gardent la trace des fonds qui ont servis à reconstituer la sous-collection.

Ce travail a été effectué (par exemple) avec les disques TOM (tous des 33t 30cm) et Fat Cat's Jazz qui sont donc référencés maintenant par rapport à des inventaire **cmj-33t-tom**, et **cmj-33t-fcat**.

### Catalogue (collections transverses)

Le catalogue des collections reconstituées permet, pour chaque collection, de retrouver les collections originelles concernées ; c'est le champs **colls\_id** qui indique les **ids** des collections concernées (dans la table des collections). Le champ **inv\_cmj** (valeur 1 ou 0 *i.e.* vrai ou faux) permet de savoir si des inventaires spécifiques CEMJazz sont utilisés.

Certes, en toute rigueur, le contenu du champ **colls\_id** devrait être une liste de clefs externes, mais il n'est pas possible référencer ainsi plusieurs collections [c'est un souci]. Sinon, il s'agit bien d'un catalogue, pas d'**inventaire** ni d'**infos** ni de **tags**.

<b>id</b>	
<b>titre</b>	
<b>description</b>	
<b>colls_id</b>	
<b>inv_cmj</b>	
<b>commentaires</b>	
<b>notes</b>	



Doc. 22 – Structure de données Catalogue collections transverses (7 champs)

## 4 Media

Revenons plus précisément sur nos affaires de media. Des fichiers media peuvent être associés à un item présent dans la base de données. Les tables correspondantes ne s'intéressent qu'à ce fichier media, l'item concerné étant déjà décrit par ailleurs (dans la

table des items). Chaque table items est donc associée à une table media. Ainsi, pour la série Iconographie, les tables **t2\_iconog\_items** et **t2\_iconog\_media** cf Doc. 4.

La description du media comporte une clef étrangère qui est l'identifiant de l'objet associé. Pour des media associés à la série Iconographie, cette clef sera recherchée dans la table des items Iconographie correspondante.

Comme tout item qui se respecte, une entrée dans une table média est rattachée à un inventaire. Par convention cet inventaire est celui de l'objet référencé. Ainsi une photographie ou une numérisation d'un disque dont l'identifiant est **jpd-33t-30cm-al3-7** sera rattachée au même inventaire **jpd-33t-30cm** où l'on s'intéresse aux disques 33t 30cm du fonds Jean Pierre Daubresse. Dans le cas des catalogues, où il n'y a pas d'inventaire, un ou plusieurs inventaires doivent être créés pour regrouper les media éventuels. Ces inventaires sont attribués à un fonds fictif « media ». Ainsi **med-rv-cat**, pour les media associés au catalogue des revues.

L'identifiant unique associé à un item media est le nom du fichier correspondant. Quelques règles définies par ailleurs (cf les bibliothèques media) permettent de construire ce nom d'une façon cohérente, à partir de l'identifiant unique de l'objet associé où encore à partir d'informations de contenu.

Ainsi, pour un item de la série Iconographie **mpf-docp-phot-1-14** nous trouvons deux entrées dans la table des media

- **mpf-docp-phot-1-14-r.jpg** (recto);
- **mpf-docp-phot-1-14-v.jpg** (verso).

Pour les audio numérisés utilisés dans le Juke-Box, nous préférons forger le nom à partir du titre et de la référence du disque :

- **tar-paper-stomp\_swaggle-records-jcs33780.mp3**, morceau numérisé du disque **jpd-33t-17cm-al1-10**.

La structure de données d'une table media, déjà présentée brièvement ci-dessus (voir au 3.1.4), comporte les champs habituels **id**, **infos**, **tags**, **notes** et **inventaire**.

Deux nouvelles clefs étrangères font leur apparition : **item** qui est l'identifiant dans la table *ad hoc* de l'objet photographié ou numérisé et **serveur**, la clef vers la table des serveurs, ce qui permet de localiser le fichier.

<b>id</b>	(règles de nommage)
<b>item</b>	(id de l'objet numérisé)
<b>infos</b>	
<b>tags</b>	
<b>inventaire</b>	(de l'item ou spécifique)
<b>serveur</b>	(cf table des serveurs)
<b>notes</b>	(uniquement le media)



Doc. 23 – Structure de données d'une table media (7 champs)

## 5 Pour conclure

Nous avons présenté, en détail, les structures des données configurées dans la base de données pour les différents items indexés dans la base CEMJazz2.

Pour indexer des données il est possible d'utiliser un système de formulaires associés à un inventaire et mis en place (très simplement) à la demande sur le site où d'utiliser tout procédé à sa convenance en s'inspirant le plus possible (...) des descriptions des tables données ici afin de faciliter l'importation des informations dans la base du CEMJazz.

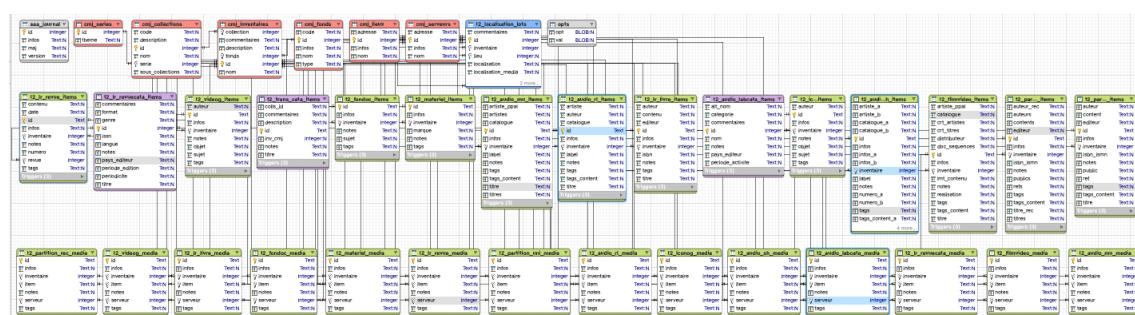
Ces structures de données sont au départ des propositions issues d'une analyse des objets de chaque collection : *Quelles sont les informations à retenir ? comment les présenter ?*, ... Ces structures ne peuvent être validées comme « opérationsnelles » que lorsqu'un « grand nombre » d'items ont été indexés, c'est le passage à l'échelle (qui passe ou pas !).

Pour les — déjà — relativement grandes collections, de nombreuses petites modifications ont du être effectuées avant de converger vers une représentation stable. Heureusement il est (relativement...) facile avec SQL de remodeler les choses sans perdre d'information.

L'idée ensuite est de présenter, de rendre facilement accessibles dans les locaux du CEMJazz tous les items qui ont été classés et indexés pour créer, disons, un parallèle entre le CEMJazz virtuel (le site internet) et CEMjazz réel (les locaux à Chevilly-Larue) où l'on retrouvera des inventaires de taille « raisonnable » et des lots physiquement localisés. C'est une autre affaire (à suivre)...



Doc. CEMJazz : obr-pub-mus-1-14





## Table des documents

1	L'item mpf-bmag-bbn-pre-1 . . . . .	5
2	La table des inventaires . . . . .	6
3	La table des items « iconographie » reliée à son inventaire . . . . .	8
4	La table des items « iconographie » et sa table media . . . . .	9
5	La table des items « revue », catalogue et tables media . . . . .	9
6	Les collections (items, media, catalogue) dans la base CEMJazz2 . . . . .	11
7	Structure de données Audio « Rouleaux » (11 champs) . . . . .	12
8	Structure de données Audio « Shellac » (18 champs) . . . . .	13
9	Structure de données Audio « Musique » (12 champs) . . . . .	14
10	Structure de données Catalogue des labels (8 champs) . . . . .	15
11	Structure de données Catalogue des labels, media (7 champs) . . . . .	15
12	Structure de données Films et Vidéos (15 champs) . . . . .	16
13	Structure de données Iconographie (8 champs) . . . . .	17
14	Structure de données Vidéographie (8 champs) . . . . .	17
15	Structure de données Livres (10 champs) . . . . .	18
16	Structure de données Catalogue des revues (11 champs) . . . . .	19
17	Structure de données Revues (9 champs) . . . . .	19
18	Structure de données Partition « œuvre unique » (13 champs) . . . . .	20
19	Structure de données Partition « recueil » (15 champs) . . . . .	21
20	Structure de données Fonds documentaires (8 champs) . . . . .	22
21	Structure de données Matériel de collection (6 champs) . . . . .	22
22	Structure de données Catalogue collections transverses (7 champs) . . . . .	23
23	Structure de données d'une table media (7 champs) . . . . .	24

Orsay sur Yvette, le 29 Frimaire an CCXXXIV (*vulg.* 19 décembre 2025) V. 1.6 provisoirement définitive  
Philippe d'Anfray [Philippe@d-Anfray.fr](mailto:Philippe@d-Anfray.fr)