



---

# Qu'est-ce qu'un data catalog ?

C'est un fait, les données sont un levier de croissance et de compétitivité. Elles sont perçues comme **des informations stratégiques favorisant l'innovation**. De ce constat, les entreprises se réorganisent pour adopter une démarche dite "data-driven". Fini l'intuition ou les expériences personnelles, les projets sont guidés par les données.

**Devenir data-centric** est une culture qui se pense et s'organise avec des process mais aussi des outils. Pour atteindre cet objectif, **le monde de la data voit apparaître de nouveaux outils** centralisant ces assets stratégiques. On les appelle **les data catalogs**.

## Pourquoi un data catalog ?

Les sujets data sont encore aujourd'hui la chasse gardée de profils trop techniques. Or, l'innovation par la donnée n'est possible que si celle-ci est **partagée au plus grand nombre**. Le métier doit avoir l'autonomie d'accéder aux données pour mesurer, démarrer ou encore optimiser un produit ou service. Innover demande une certaine **flexibilité et agilité** qui est, à ce jour, trop peu présente dans les organisations.

**Démocratiser l'accès aux données !** C'est la raison même des data catalogs : permettre à l'ensemble de ses

collaborateurs de trouver les données dont ils ont besoin via **une plateforme simple d'utilisation** au-dessus des systèmes data. Les data catalogs ne nécessitent pas de compétences techniques pour enfin **découvrir ce qui est nouveau et saisir des opportunités**. Business analysts, data scientists ou encore équipes marketing deviennent autonomes dans l'exploration des données. Quant aux data stewards, ils sont enfin outillés pour **construire une gouvernance data** et évangéliser une culture data-centric au sein des organisations.

## Les objectifs d'un data catalog

Un data catalog permet d'**acquérir une vision métier des données** entreposées dans les systèmes data. Il **centralise et unifie les informations collectées** afin qu'elles puissent être partagées auprès des équipes IT et fonctions métier puis connectées aux outils de l'entreprise. Cette vue unifiée des données permet de :

### Construire une gouvernance data

Le data catalog permet de cartographier et visualiser le cycle de vie de ses données. Les professionnels de la data savent enfin où se trouvent leurs données, qui les utilise, dans quel but et comment celles-ci sont utilisées.

### Documenter les données

Le data catalog permet de créer un répertoire de métadonnées techniques et business. Cette documentation connectée stocke ces informations dans le but de faciliter la recherche et la découverte de données toujours up-to-date.

### Fédérer les collaborateurs autour des données de l'entreprise

Le data catalog devient l'outil data de référence pour l'ensemble des employés. Son interface web ne nécessite pas de compétences techniques pour découvrir et comprendre les données. Il permet également de collaborer avec ses pairs.

### Rendre la donnée intelligente

Grâce à la création de modèles prédictifs sur les données cataloguées, la productivité est accrue et l'innovation par la donnée devient de plus en plus accessible.

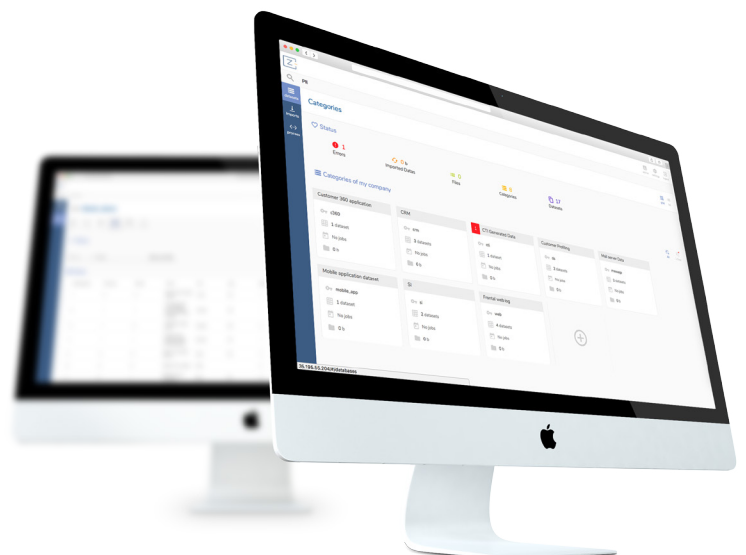
## Les fonctionnalités clés d'un data catalog

### Registre de métadonnées

Ce référentiel de métadonnées dynamique intervient à tous les niveaux : du jeu de données à la donnée en elle-même. Pour chaque élément, ce registre de métadonnées peut faire figurer **une description business et technique**, les propriétaires, des indicateurs de qualité ou encore créer une taxonomie (properties, tags, etc.).

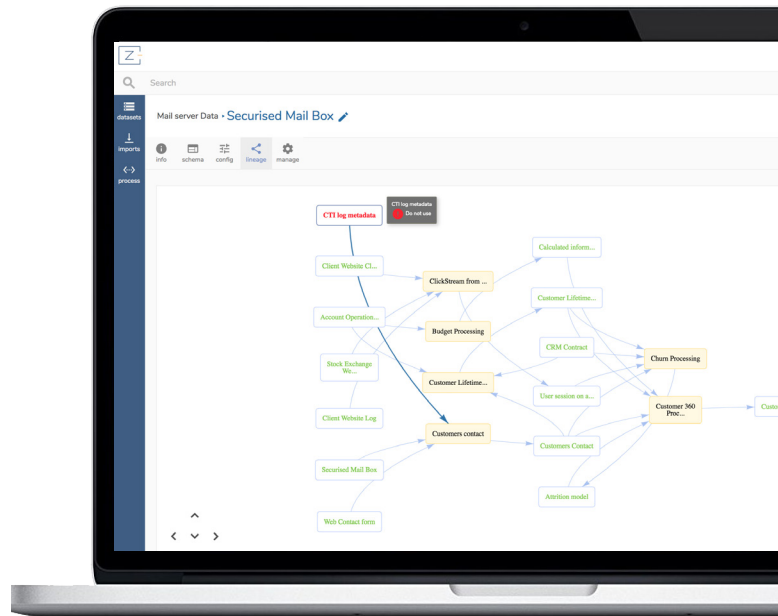
### Moteur de recherche

Toutes les métadonnées renseignées dans le registre sont **requêtables depuis le moteur de recherche** du data catalog. Les recherches peuvent être triées, filtrées à tous les niveaux.



## Data lineage et registre de traitements

Grâce au data lineage, il est possible de **visualiser dans son ensemble l'origine et les transformations** d'une donnée spécifique au fil du temps. Cela vous permet de comprendre d'où proviennent les données, quand et où elles se séparent et fusionnent avec d'autres données. Ces transformations et traitements effectués sur la donnée sont ainsi répertoriés dans ce qu'on appelle **un registre de traitements**, indispensable pour répondre aux attentes de la réglementation européenne (GDPR).



## Fonctionnalités collaboratives

Dans une **approche user centric**, un data catalog est l'outil data de référence d'une entreprise. Il permet de visualiser la donnée comme un asset et de travailler de manière transparente sur celle-ci. Partager, assigner, commenter, qualifier à l'intérieur même de l'outil pour augmenter la productivité et la connaissance auprès de tous les collaborateurs.

## Détection de données sensibles

Les algorithmes de machine learning et d'intelligence artificielle peuvent **reconnaître les données sensibles** au sein du data catalog et lors de nouveaux imports dans l'outil. Le data catalog est à même de surveiller l'activité de ces données et d'avertir les personnes référentes en cas de problèmes.

## Les cas d'usage d'un data catalog ?

### Data steward

**Le data catalog, un outil de gouvernance data.**

- **Référencer dans un espace dédié** les connaissances techniques et fonctionnelles récupérées sur les données.
- Organiser les données.
- Donner un accès aux informations utiles et nécessaires aux utilisateurs de la donnée.
- Identifier les différents utilisateurs de la donnée et gérer le niveau d'habilitation de chacun.
- **Qualifier** la valeur des données.
- **Démarrer une mise en conformité** avec la réglementation européenne (GDPR).

### BI analyst / Data scientist

**Le data catalog, un accélérateur de projet.**

- **Trouver très facilement vos données**, peu importe où elles sont stockées.
- Consulter l'historique des jeux de données : date de création et les actions effectuées dessus.
- **Comprendre le contexte métier des données.**
- Identifier les sachants par jeu de données.
- **Collaborer** facilement avec ses pairs.
- **Créer une documentation automatisée** de part mes actions au sein du data catalog.
- Recommandation de données pertinentes par rapport aux autres jeux de données consultés.

## Les bénéfices d'un data catalog ?

À mesure que les fonctionnalités du data catalog deviennent de plus en plus automatisées et sophistiquées, les entreprises bénéficient de trois avantages principaux :

### Maximiser la valeur des données

En rassemblant l'ensemble des données d'une entreprise sur un outil data de référence, il devient possible de croiser ces assets et en tirer de la valeur plus facilement. La collaboration des équipes techniques et métier à l'intérieur même du data catalog permet des innovations répondant à des besoins marché avérés.

### Produire mieux et plus vite

Vos équipes le confirmeront : plus de 70% du temps consacré à l'analyse des données est engagé dans des activités de "querelles de données". Le catalogage simplifie la recherche des données, l'identification des sachants et donc, la prise de décisions intelligentes.

### Assurer le bon contrôle des données

Mal interprétées ou erronées, les entreprises s'exposent à fonder leurs décisions sur de mauvaises informations. Les data catalogs connectés permettent d'accéder à des données toujours à jour. Les utilisateurs des données peuvent s'assurer que les données et leurs informations sont corrects et utilisables.

En 2019,  
**80%**  
des data lakes\*  
implémentés dans  
les entreprises  
seront inefficaces  
sans une bonne  
gestion des  
métadonnées.

*\*Gartner survey :  
Data catalog is the new black*

Zeenea intègre sans effort les données des systèmes Big Data d'une entreprise dans son **data catalog Plug&Play** collaboratif et simple d'utilisation.

Véritable self-service de la donnée, Zeenea permet de **documenter et explorer rapidement vos données** selon les besoins et les autorisations de chacun.