



[Blog](#) [Talks](#) [Publications](#) ▾ [Tools](#) ▾ [Radar / Plan](#) ▾ [Team](#) [About](#)

# Master Data Management (MDM) : concepts, exemples, architectures et bonnes pratiques

Posted on 2025-08-20 by [Isabelle Boydens](#)

## *Nederlandstalige versie*

Dans cet article, nous illustrons la problématique à laquelle répond la gestion des Master Data (« Master Data Management (MDM) »). Nous définissons ensuite cette branche de l'informatique sur la base d'exemples et montrons ses liens avec la gouvernance et la gestion de données ("data governance" et "data management"). Nous présentons également une méthode de mise en œuvre ainsi que des architectures associées avec leurs avantages et inconvénients.

## La problématique à résoudre

La gestion des Master Data vise à résoudre deux types de difficultés.

D'une part, certaines données peuvent être dispersées, dupliquées et hétérogènes dans différentes applications et bases de données (DB) sémantiquement liées entre elles.

D'autre part, d'une DB à l'autre, les données peuvent diverger (quant au format ou au domaine de définition, par exemple) ou évoluer à des rythmes différents. En conséquence, des problèmes métier graves peuvent se poser. On observe ainsi un manque de traçabilité entre bases de données interdépendantes ou encore, l'absence de données de qualité « AI ready », ce qui constitue un obstacle à la mise en place efficace des projets d'IA croissants à l'heure actuelle.

Les buts de la gestion des Master Data sont donc l'élimination des incohérences et des dysfonctionnements opérationnels ainsi que

l'amélioration de la qualité des données et du service rendu.

## Définition et exemples

Le Master Data Management est davantage une discipline métier qu'un logiciel et repose sur la gouvernance des données et le data management.

La gouvernance des données ("data governance") est la spécification par la direction de l'institution de plusieurs éléments. Un cadre de responsabilité de haut niveau doit spécifier les modalités de gestion des Master Data (« policy settings »). Celles-ci sont fondamentales pour le business (par exemple : adresse de l'employeur ou de l'entreprise) et sont partagées entre plusieurs bases de données. A cette fin, des rôles doivent être identifiés à différents niveaux : gestion métier et technique du MDM, du système de méta-information, de la qualité des données, de l'architecture, ... Enfin, une organisation appropriée pour l'évaluation, la création, la consommation et le contrôle des données est mise en place.

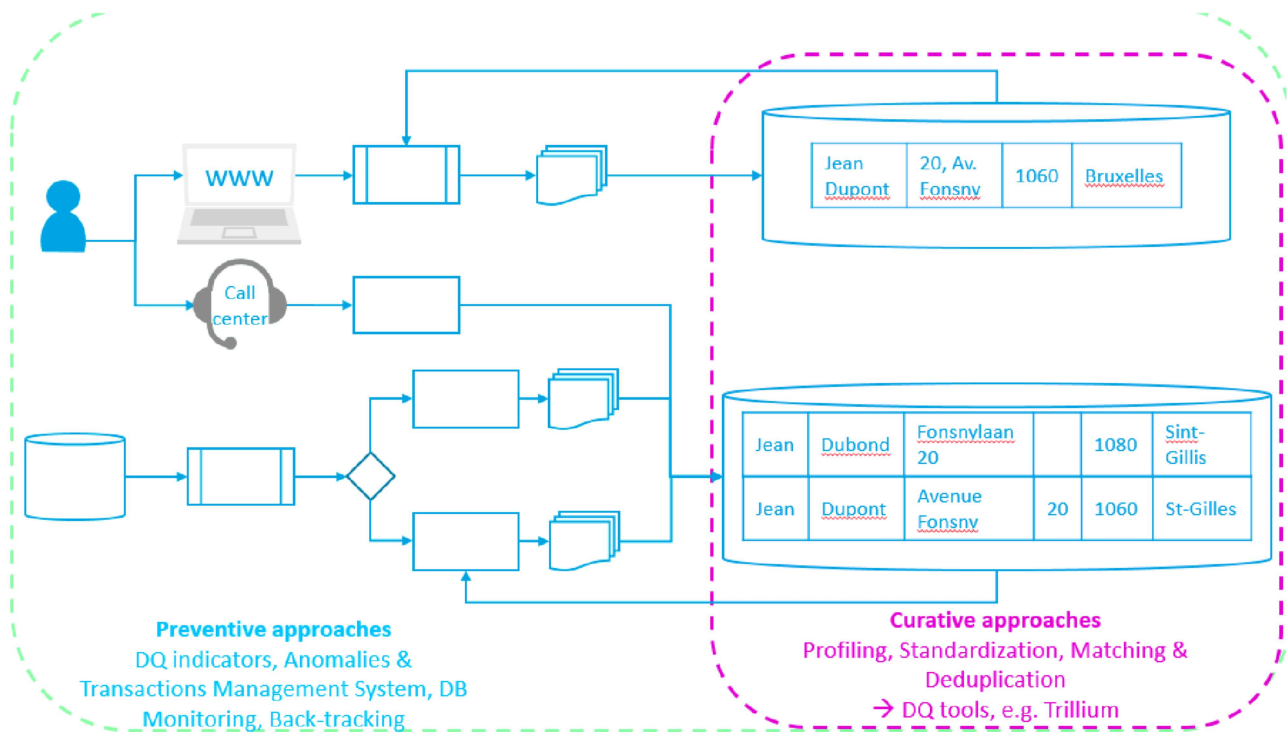
Le « data management » permet, sur la base des « policy settings », de manière itérative et incrémentale, l'identification, la définition et la modélisation des Master Data, via un premier "case study" validé par le métier concerné, en vue d'une approche plus large, ce qui comporte trois axes concomitants.

En premier lieu, un système de méta-information ou glossaire transversal de données à haute valeur ajoutée (« data catalog ») doit être mis en place, sujet à propos duquel nous avons publié un article de blog en français et néerlandais en mars 2025 (6). Le catalogue de données peut aussi gérer les données non partagées entre bases de données et dont la documentation est importante.

En second lieu, une approche focalisée sur la qualité des données (6, 7, 8, 9, 10, 11) doit être mise en place. Elle inclut deux types de méthodes. Les *méthodes curatives* (9), via des *data quality tools*, (Batch et Rest API sur le catalogue ReUse) permettent de traiter les problèmes (anomalies, présomptions de doublons, adresses à nettoyer,...) quand ils sont déjà présents dans les bases de données. Les *méthodes préventives* (7, 8, 11) permettent quant à elles de prévenir l'émergence d'anomalies en en détectant la cause (ou les causes) entre institutions et expéditeurs dans

les flux d'information (par exemple, problème d'interprétation de la loi, émergence d'un nouveau concept (mutations de virus, ...), définitions incohérentes, bugs, ...) et en les supprimant structurellement à la source, de sorte qu'elles ne se présentent plus/pas dans les bases de données (à venir, voir [Catalogue ReUse](#)).

La figure ci-dessous illustre les deux méthodes schématiquement, lesquelles peuvent interagir entre elles.



Enfin, en troisième lieu, un système de Master Data Management assure le choix d'une intégration entre les DB concernées et le catalogue de données (ou système de méta-information), il existe différentes architectures, que nous verrons plus loin.

Ces étapes importantes permettront de gérer transversalement des données synchronisées dans différents projets et applications. L'objectif de l'approche MDM est de mutualiser les efforts et d'assurer la synchronisation, la qualité, le partage et le contrôle des données à travers les différents silos. Et ce, quasiment en temps réel ou en mode différé (quand un workflow d'approbation s'impose pour la validation des modifications et des versions).

Par exemple (2, 5) : le MDM permettra de traiter les cas où deux termes différents sont utilisés dans le même sens et doivent être harmonisés (*montant à payer, montant dû*) ou encore, ceux où un même terme peut avoir plusieurs sens différents (ainsi, *salaires* peut signifier *salaires brut, salaires de base, salaires et traitements*, ou encore, *salaires net, ...*). Dans notre contexte belge, il faudra prendre également en considération l'harmonisation entre les termes des différentes langues usitées, ce qui est un facteur de complexité supplémentaire. En effet, il n'y a pas de relation biunivoque nécessaire, pour un concept donné, entre les différents termes des différentes langues nationales.

Plus rigoureusement, on distingue des données suivantes dans le cadre du MDM (pour plus de facilité, nous parlerons par la suite de Master Data) :

- *Master Data* : le plus petit nombre d'ensembles cohérents d'identifiants et d'attributs qui décrivent de manière unique les entités principales d'une institution ou d'une entreprise et sont utilisés dans plusieurs bases de données et processus métier liés conceptuellement et fonctionnellement entre eux.
- *Data "non master"* : donnée (provisoirement) non partagée par plus d'une base de données à un instant t (exemple fictif : nom de l'ouvrier affilié à l'ONVA, Office National des Vacances Annuelles – prélève et redistribue les pécules des ouvriers)
- *Reference Data* : taxonomies, listes de codes partagés (exemple : codes pays) – *bijlage – annexes dans les glossaires de la sécurité sociale*
- *Operational Data* : données sensibles potentiellement constituées des trois premières et d'un traitement confidentiel, demandant un échange sécurisé (exemple : le montant de prestations de maladie professionnelles dues au travailleur T, dont le n° de RN est ... pour l'instant t.).

## Master Data Management : méthode de mise en oeuvre

Sachant qu'un projet de Master Data Management est avant tout un projet métier, avant la mise en place d'un système informatique, il comporte les étapes suivantes souvent itératives (1, 3, 4) :

- Définir l'étendue du projet (commencer par un projet "modeste" essentiel, qui peut être incrémental et itératif, "nice to have")

- Définir un agenda haut niveau (planning) de mise en œuvre, projet continu (conception et maintenance) soutenu par la hiérarchie, ce qui inclut les points suivants :
- Mettre du relief (analyse), et identifier :
  - Les utilisateurs des données, leurs objectifs
  - Les sources authentiques
  - Les concepts principaux, et les Master Data (identifiant unique, catégories principales, ...). Attention : des compromis sont parfois nécessaires, le choix des Master Data n'est pas nécessairement déterministe
  - Les événements/processus pouvant affecter les Master Data (mise à jour, partage, et suppression, dus à l'évolution de la législation ou du réel appréhendé – par exemple : mutations de virus et évolution des concepts en médecine, ...)
  - l'organisation associée (p. ex. workflow de validation)
  - La gestion des versions des Master Data et des métadonnées (6)
  - La qualité des données : évaluation et amélioration (6, 7, 8, 9, 10, 11) – voir supra
  - La sécurité et la vie privée
- Définir des KPI ou métriques pour valider, mesurer et suivre les résultats de l'approche MDM, par exemple :
  - Baromètres de qualité DmfA: suivi des anomalies, indicateurs financiers (AR-KB 2017), ... (11)
  - Mesures de la traçabilité entre bases de données sémantiquement liées entre elles.
- Établir de façon incrémentale, comme évoqué plus haut, un référentiel ou glossaire des données ou système de méta-information (6) en retenant les fonctionnalités importantes pour une mise en production ultérieure :
  - Gestion des versions (planning) des Master Data et des métadonnées (législation évolutive, durée de prescription, force probante, émergence de nouveaux concepts, ...)
  - workflow de validation
  - Multilinguisme
  - Héritage
  - Standard et format d'échange (« Write Once Publish Many »)
  - Recherche multibase et multilingue

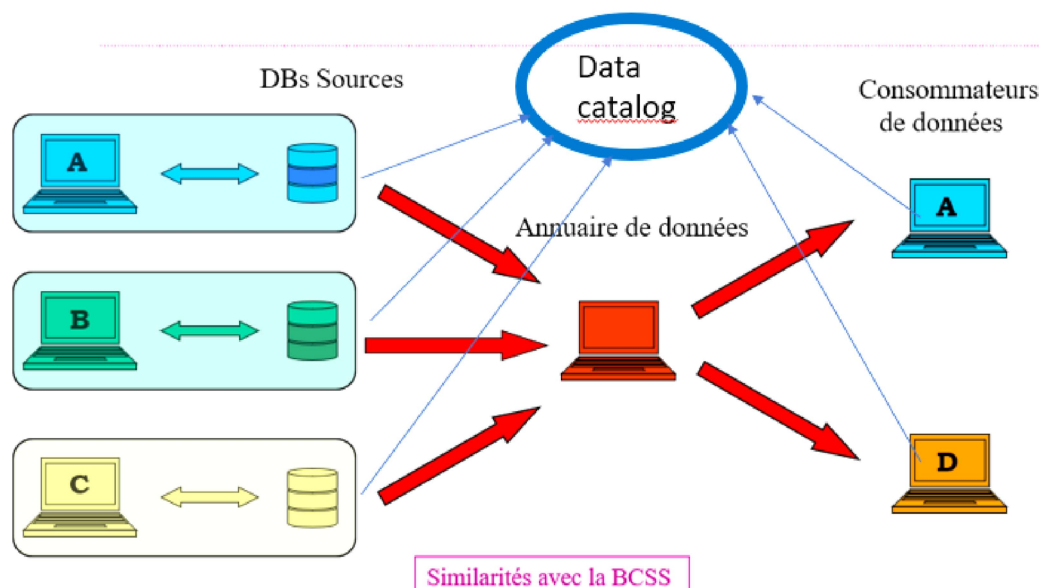
- Établir une stratégie d'évolution, de gestion du changement pour le passage de la situation antérieure ou actuelle (« AS IS ») à la situation à venir (« TO BE »)
- Établir des rôles et des équipes business et IT (MDM, Data Quality, Architecture, ...)
- Définir les standards (recommandé pour l'analyse : (12))
- Établir une méthode d'intégration des bases de données concernées ; via le référentiel ou glossaire transversal des données, ce que nous allons voir au point suivant : les architectures d'intégrations, avec leurs avantages et inconvénients sur le plan :
  - de la qualité des données (10)
  - de la sécurité et de la vie privée
  - de la performance
  - du caractère plus ou moins intrusif dans les systèmes d'information concernés

Architectures d'intégration entre bases de données et data catalogs éventuellement liés, avec leurs avantages et inconvénients

Les schémas qui suivent sont adaptés de (5) et modifiés selon les solutions actuelles (1, 2, 3, 4)

## 1. Les répertoires virtuels

### Les solutions: Répertoire Virtuel

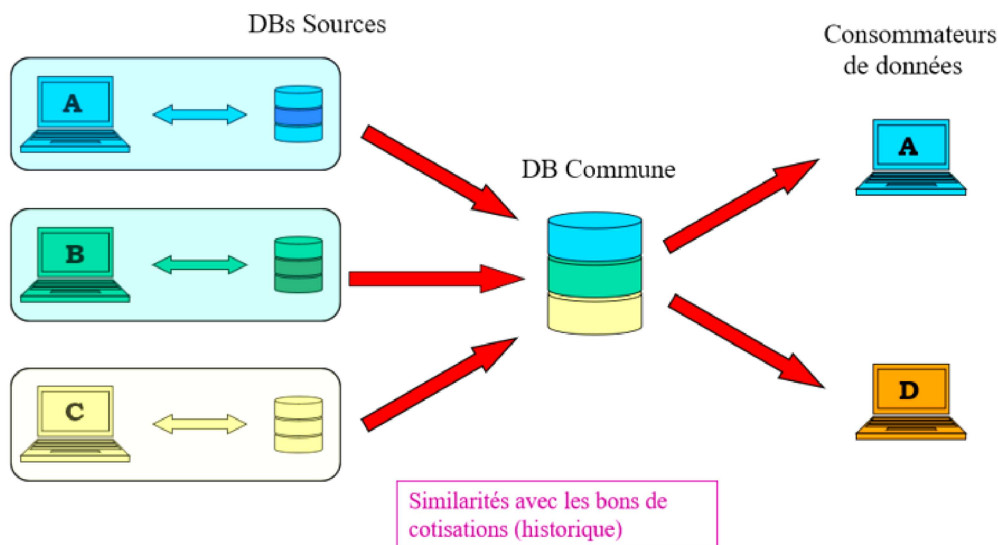


Avec un Répertoire virtuel, l'annuaire des données (au centre du schéma, en rouge) permet, en fonction d'une base de connaissance sur les autorisations de consultation, le transfert des données des bases de données sources vers les consommateurs. Il y a des similarité avec la BCSS qui est un réseau stellaire permettant l'échange de données entre institutions. S'agissant de la sécurité sociale, les consommateurs de données peuvent consulter un glossaire des données (ou "data catalog" dans le schéma) au fil des versions, en ligne sur le portail de la sécurité sociale (6). *Il s'agit d'une bonne solution dans le contexte évoqué.*

- *Avantages* : respect de la vie privée, sécurité, les bases de données sources ne doivent pas être modifiées
- *Désavantages* : questions de performance éventuelles, les producteurs de données doivent partager leur data catalog, l'unifier, en assurer la mise à jour au même rythme et le rendre accessible aux consommateurs.

## 2. la consolidation

### Les solutions: Consolidation

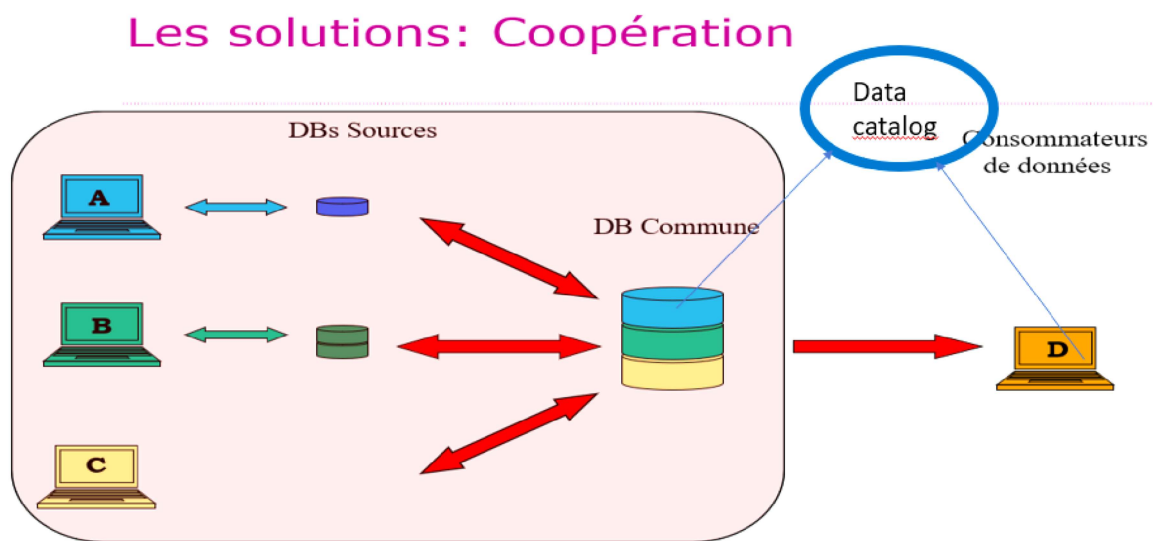


Dans la consolidation, les bases de données sources sont copiées dans une base de données commune en une seule fois ("one shot"). Aucun "data catalog" ne se trouve sur le schéma car la base de données commune est ensuite désynchronisée des bases de "données sources" qu'elle est censée représenter. *Cette solution n'est pas conseillée.* Elle fut appliquée

par le passé à la gestion des bons de cotisations en soins de santé, puis abandonnée.

- *Avantages* : simple pour les producteurs de données : les bases de données sources ne doivent pas être modifiées
- *Désavantages* : les bases de données sources et leur data catalog peuvent évoluer à leur propre rythme et la base de donnée commune à un autre rythme ("ghost factory", source de redondance) : des problèmes se posent au niveau de la qualité des données "consolidée", partager un "data catalog commun" n'a plus de sens

### 3. La coopération



Avec la coopération, les bases de données sources partagent les "Master Data" communes dans une nouvelle base de données et continuent à gérer leurs données qui ne sont pas communes. Pour les données communes qui sont gérées en un seul exemplaire (donc, sans redondance), un "data catalog" est mis à disposition de tous les consommateurs de données et mis à jour de manière collégiale par les producteurs de données. *Il s'agit d'une bonne solution si les bases de données sources demandent une restructuration parce que, par exemple, elles sont devenues obsolètes techniquement et conceptuellement.*

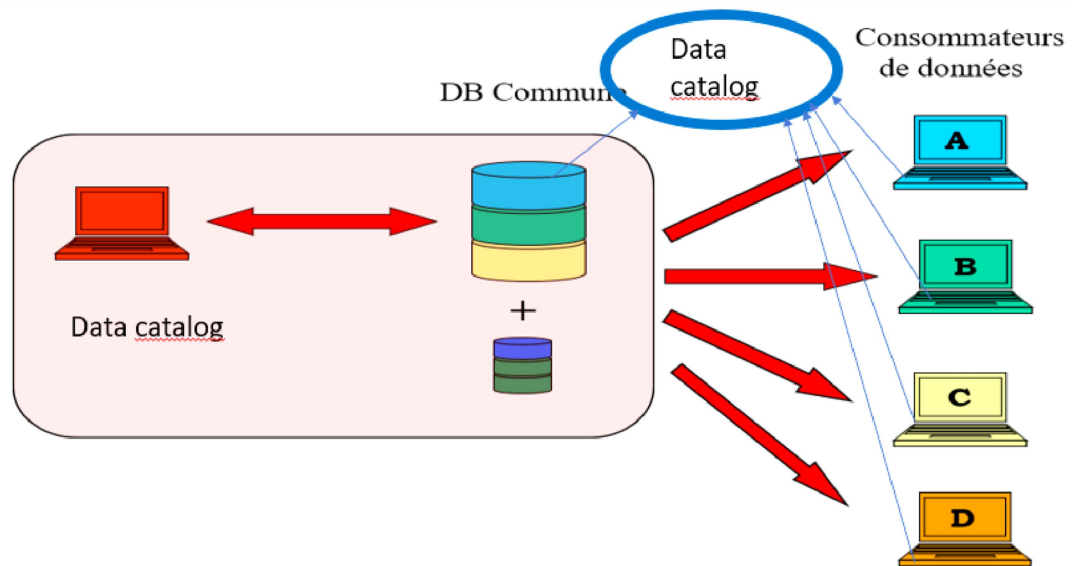
- *Avantages* : les Master Data sont accessibles en un endroit unique (ce qui est bon pour la qualité des données) et leur "data catalog" est commun, la partie accessible pour les consommateurs de données est diffusée. Le respect de la vie privée et la sécurité son

assurés. Chaque producteur de données continue à gérer les données qui lui sont propres et ne sont pas partagées.

- *Désavantages* : les bases de données sources doivent être restructurées, questions de performance éventuelles

#### 4. La centralisation

### Les solutions: Centralisation



Avec la centralisation, les bases de données sources partagent leurs Master Data dans une nouvelle base de données unique sur laquelle les producteurs de données n'ont plus seuls le pouvoir. Cette nouvelle base de données est gérée selon une organisation collégiale coordonnée et imposée aux producteurs de données. Un seul "data catalog" en définit le contenu harmonisé (on le voit représenté en rouge dans le schéma et à l'extérieur, accessible à toutes les parties, producteurs et consommateurs). Ce « data catalog » est également géré de manière collégiale et coordonnée, imposée aux producteurs de données.

*Pour les raisons évoquées (dans les inconvénients ci-dessous), au sein du domaine d'application de l'administration publique, cette solution n'est pas conseillée. Elle est parfois pratiquée dans le secteur privé au sein d'une multinationale, par exemple (elle fut appliquée dans les années 1990 chez T&T Laboratories aux USA).*

- *Avantages* : les Master Data et leur catalogue sont accessibles en un endroit unique (ce qui est bon pour la qualité des données) et leur data catalog est commun
- *Inconvénients* : les bases de données sources doivent être restructurées et perdent leur pouvoir sur les données dont elles avaient la gestion seules, il y a des questions potentiellement graves de vie privée, de sécurité et de performance (beaucoup de requêtes)

## Conclusions provisoires

En ce qui concerne les architectures proposées (et toutes les étapes préalables illustrées dans cet article étant prises en compte), le répertoire virtuel avec "data catalog" semble la meilleure solution. Si un (*reengineering*) de bases de données source est requis, la coopération peut-être envisagée.

Dans un prochain article de blog sur le Master Data Management, nous envisagerons les "lessons learned" en Belgique et à l'étranger sur le recours aux Master Data Management selon les outils et les méthodes de mise en oeuvre, ainsi qu'une typologie des MDM Tools avec un regard critique et constructif.

## Références

(1) DUBOIS P. et al. (University of Paris), *Harnessing Data Integrity: A Study of Master Data Management Best Practices*. MZ Computing Journal, vol 5, issue 1, 2023.

(2) *Hype Cycle for Data and Analytics Governance 2025*, Gartner, 19 June 2025 – ID G00827117.

(3) LEPENIOTIS P, *Master data management: its importance and reasons for failed implementations*. Doctoral Thesis, Sheffield Hallam University Press, 2020 (3 vol.).

(4) SINGH A. et al., *Best Practices for Creating and Maintaining Material Master Data in Industrial Systems* In International journal of research and analytical reviews, vol 10, issue 1, Janvier 2023.

) TRIGAUX J.-C., *Master Data Management – Mise en place d'un référentiel de données*. Bruxelles, Smals Research,

## Deliverable 2009/TRIM4/01.

(6) BOYDENS I., [Au coeur de la "data governance": les "data catalogs" ou systèmes de méta-information](#), Bruxelles, Smals Research, article de blog, 19/03/2025 (*inclut un lien vers la version néerlandaise*).

(7) BOYDENS I., HAMITI G. et VAN EECKHOUT R., *A service at the heart of database quality. Presentation of an ATMS prototype*. In Le Courrier des statistiques, Paris, INSEE, 2023, n°6, 11 p. (publié le 2/10/2023). [Lien vers l'article](#).

(8) BOYDENS I., HAMITI G. et VAN EECKHOUT R., [Data Quality : "Anomalies & Transactions Management System" \(ATMS\), prototype & "work in progress"](#), Bruxelles, Smals Research, article de blog, 02/12/2020, last update 04/07/2025. (*inclut un lien vers la version néerlandaise*).

(9) BOYDENS I., CORBESIER I. et HAMITI G., [Data Quality Tools : retours d'expérience et nouveautés](#), Bruxelles, Smals Research, article de blog, 07/12/2021.

(10) BOYDENS I., [Dix bonnes pratiques pour améliorer et maintenir la qualité des données](#), Bruxelles, Smals Research, article de blog, 16/06/2014, last update : décembre 2021.

(11) BOYDENS I., [Data Quality & Back Tracking : depuis les premières expérimentations à la parution d'un Arrêté Royal](#), Bruxelles, Smals Research, article de blog, 14/05/2018.

(12) XLS et le fichier FedVoc publié sur le GitHub de BelgIF [GitHub – belgif/fedvoc: Federal Vocabularies](#)

---

*Ce post est une contribution d'Isabelle Boydens, Data Quality Expert chez Smals Research. Cet article est écrit en son nom propre et n'impacte en rien le point de vue de Smals*

Annexe (3)

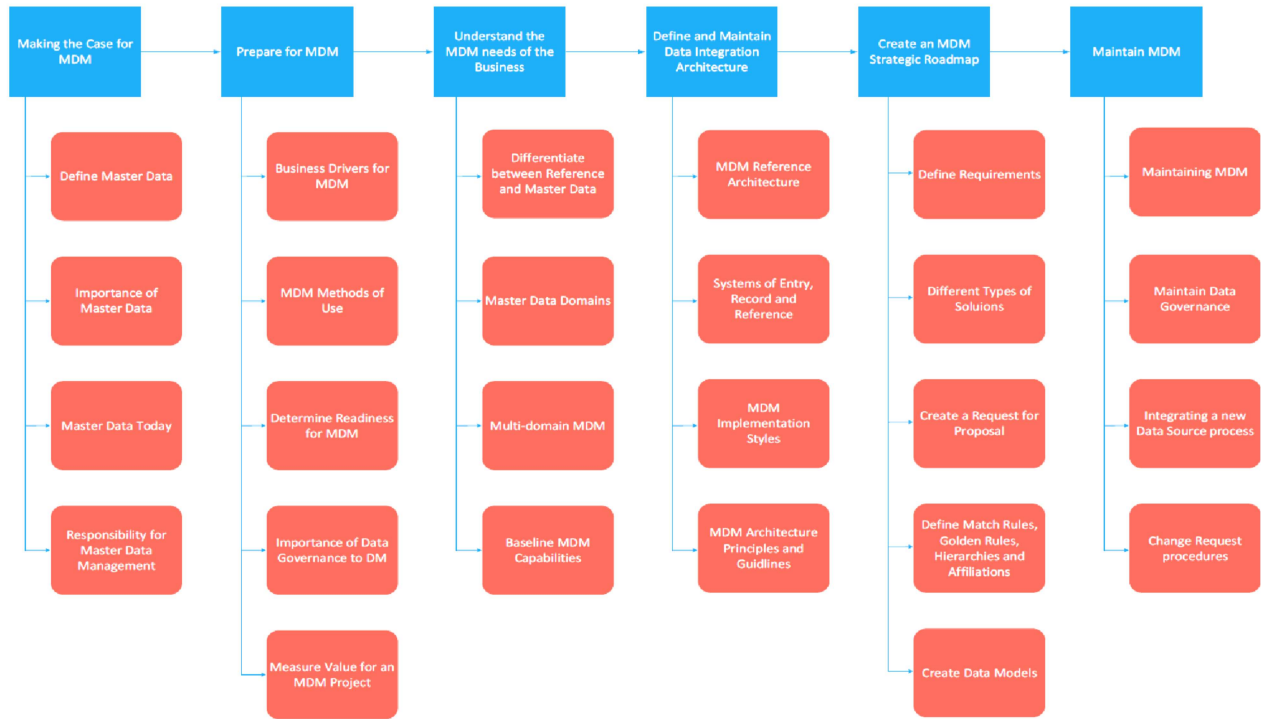


Figure 1 MDM Implementation Roadmap: Infotech Research Group 2014

- data quality
- information management
- master data management
- metainformation system

MORE POSTS

### Je data beschermen tegen beheerders: 'on-premise' Confidential Computing

2026-03

### Protéger ses données des administrateurs : l'informatique confidentielle « on-premise »

2026-03

### De performance van LLM's: Een vergelijkende analyse tussen Frans en Nederlands

2026-03

# Made by Smals Research – Privacyvriendelijk Kruisen van Persoonsgegevens

2026-02

Search

Newsletter &amp; webinars:

 Dutch  French

Keywords:

[analytics](#) [artificial intelligence](#) [big data](#) [blockchain](#) [bpm](#) [chatbot](#)[cloud computing](#) [cost cutting](#) [cryptography](#) [data center](#)[data quality](#) [development](#) [eda](#) [egov](#) [event](#) [gis](#)[information management](#) [machine learning](#) [managing it costs](#)[methodology](#) [mobile](#) [natural language processing](#) [open source](#) [privacy](#)[productivity](#) [security](#) [social](#) [software design](#)[software engineering](#) [standards](#)

## Smals Research

© Smals Research – License/Disclaimer: [FR](#) / [NL](#)