

Table des matières

Avant-propos	XI
Introduction	XV

Première partie – Comprendre : les concepts

Chapitre 1 – La gestion des données de référence.....	3
1.1 Principes et notions élémentaires.....	3
1.1.1 Caractéristiques des données de référence	5
1.1.2 Typologie des données de référence	6
1.1.3 La notion de famille de données.....	7
1.1.4 Positionnement par rapport à d'autres notions	8
1.1.5 Gestion des données de référence (GDR)	11
1.1.6 Enjeux et besoins	13
1.1.7 Résumé des enjeux et besoins, importance de la GDR	17
1.2 Exemples d'entreprises	19
1.2.1 Un grand distributeur	19
1.2.2 Un producteur	20
1.2.3 Un fournisseur	21
1.3 Pourquoi mettre en œuvre une gestion des données de référence ?.....	21

Chapitre 2 – La donnée et ses dimensions de valeur	25
2.1 Données et valeur	25
2.2 Qualité des données	26
2.2.1 Principaux concepts	26
2.2.2 Les critères de la qualité des données	27
2.2.3 Quels objectifs de qualité des données ?	32
2.2.4 Amélioration de la qualité des données	33
2.3 Principales causes d'incohérence ou de non-qualité	35
2.4 Exemples de non-qualité ou d'incohérence	36
2.4.1 Doublons	36
2.4.2 Données incomplètes	37
2.4.3 Données trop longues à générer	38
2.4.4 Données incohérentes	39
2.4.5 Autres exemples de problèmes fréquemment rencontrés	40
2.5 Métadonnées	40
2.5.1 Types de métadonnées	41
2.5.2 Les apports des métadonnées à la valorisation des données	43
Chapitre 3 – Données et processus	45
3.1 Processus	45
3.2 Processus métier et processus référentiels	47
3.3 Cycle de vie métier	50
3.4 Cycle de vie technique	51
3.5 Urbanisme, urbanisation et données	53
3.6 Urbanisme et données en pratique	56

Deuxième partie – Mettre en œuvre : technologies et solutions

Chapitre 4 – Typologies d'architectures	61
4.1 Fondement des architectures	61
4.2 Architecture et chaîne de l'information	64
4.3 Les quatre types d'architecture pour la gestion de données de référence	67

4.4	Architecture de consolidation	69
4.5	Architecture de coopération	70
4.6	Architecture de centralisation	73
4.7	Architecture de répertoire virtuel	74
4.8	Tableau de synthèse entre architecture et cas d'utilisation	76
4.9	Conséquences pour les métiers	78
4.10	Synthèse des critères de choix d'une architecture	79
4.11	Couverture du référentiel	80
4.12	Les modes d'implémentation des référentiels	81
4.13	Chaînes référentielles	83
	Chapitre 5 – Outillage d'une solution référentielle	85
5.1	Typologie des solutions	85
5.2	MDM (Master Data Management)	86
5.3	DQM (Data Quality Management)	89
5.3.1	Zoom sur l'évaluation des sources de données	90
5.3.2	Zoom sur la qualité et la migration des données	91
5.3.3	Zoom sur l'amélioration de la qualité des données	92
5.3.4	Positionnement DQM versus MDM	93
5.3.5	Pourquoi le DQM ne peut-il remplacer le MDM ?	95
5.4	EII (Entreprise Information Integration)	95
5.4.1	Positionnement EII versus MDM	96
5.4.2	Pourquoi un outil EII ne peut-il servir de référentiel ?	96
5.5	Annuaire	97
5.5.1	Positionnement annuaire versus MDM	97
5.5.2	Pourquoi un annuaire ne peut-il servir de référentiel ?	98
5.6	CRM (Customer Relationship Management)	98
5.6.1	Positionnement CRM versus MDM	99
5.6.2	Limites d'un système CRM	99
5.7	PLM (Product Lifecycle Management)	100
5.7.1	Positionnements PLM et MDM	101

5.7.2	<i>Les fonctions du PLM</i>	102
5.8	Synthèse	103
5.9	Complément d'information sur l'EIM	106
Chapitre 6 – Architecture fonctionnelle du MDM		107
6.1	Fonctionnalités d'une solution de gestion de données de référence	107
6.1.1	<i>Acquisition de la donnée et processus</i>	108
6.1.2	<i>Validation et qualité</i>	111
6.1.3	<i>Fonctions de pilotage</i>	120
6.1.4	<i>Modèles de données et métadonnées</i>	121
6.1.5	<i>Fonctions de stockage et journalisation</i>	124
6.1.6	<i>Fonction d'accès et de diffusion des données</i>	127
6.1.7	<i>Administration et maintenance</i>	128
6.2	Les catégories de solutions MDM	131
6.3	Socle référentiel	132
Chapitre 7 – Positionner le référentiel dans le SI		135
7.1	Briques applicatives	135
7.2	Projets classiques et leurs applications	138
7.2.1	<i>Amélioration de la performance des processus</i>	139
7.2.2	<i>Interfaçage du SI avec les tiers (communication B2B)</i>	144
7.2.3	<i>Reporting et analyse</i>	145
7.3	Importance des échanges de données	147
7.3.1	<i>Point à point</i>	148
7.3.2	<i>EAI (Enterprise Application Integration)</i>	148
7.3.3	<i>ETL (Extraction Transformation Loading)</i>	149
7.3.4	<i>Services et SOA</i>	150
7.3.5	<i>MDM et échanges, apports et nécessités</i>	152
7.4	Pour une insertion progressive du MDM dans les échanges du SI	154
Chapitre 8 – Guide de choix des architectures et solutions		159
8.1	Choix d'architecture	159
8.2	Choix de solutions	161

8.3	Mode d'implémentation et éligibilité des solutions de MDM	166
8.4	Solutions mises en place par quelques entreprises	168
8.4.1	<i>Un grand distributeur</i>	168
8.4.2	<i>Un producteur</i>	170
8.4.3	<i>Un fournisseur</i>	171
8.4.4	<i>Conclusion</i>	172
8.5	Bonnes pratiques	173
8.5.1	<i>Métier et urbanisme</i>	174
8.5.2	<i>Architecture</i>	175

Troisième partie – Piloter : méthodes et organisation

Chapitre 9 – Gouvernance des données de référence	179
9.1 Définition de la gouvernance des données.	179
9.2 Modèle de déploiement de la gouvernance	181
9.3 Exemple de cadre synthétique de gouvernance	183
9.3.1 Structure du cadre (objectifs et contraintes)	184
9.3.2 Les leviers du cadre.	188
9.4 Organisation	190
9.4.1 Organisation de la gouvernance au niveau entreprise	191
9.4.2 Rôles et acteurs.	192
9.4.3 Mise en œuvre des rôles en pratique	198
9.5 Règles et procédures	199
9.6 Outils de gouvernance	201
9.6.1 Les outils de pilotage de la gouvernance	201
9.6.2 Les outils technologiques relatifs à la gouvernance	202
9.6.3 Les priorités dans la mise en œuvre	203
9.7 Exemples de mise en place de la gouvernance dans des entreprises	203
9.7.1 Un grand distributeur	203
9.7.2 Un producteur	204
Chapitre 10 – Étapes de déploiement d’un projet de gestion des données de référence	205
10.1 Principes généraux	206

10.2 Tâches spécifiques à la gestion des données	207
10.2.1 Identifier et décrire les données de référence	207
10.2.2 Identifier et décrire les processus référentiels	209
10.2.3 Spécifier les méthodes et règles de gouvernance	210
10.2.4 Définir les principes de migration des données, analyser et assainir les données existantes	210
10.2.5 Définir les règles de qualité.	211
10.2.6 Définir les contrats d'échange, de services	212
10.3 Éléments de méthode	213
10.4 Charges de mise en œuvre	216
10.5 Bonnes pratiques	217
10.5.1 Méthodes	217
10.5.2 Organisation	218
Chapitre 11 – Points clés à retenir	219
11.1 Règles de base	219
11.2 Bonnes pratiques essentielles	224
11.3 Vision prospective et progressive	224
Conclusion	227

Annexes

Annexe A – Modélisation des données	237
A.1 Outils pour créer un modèle de données	237
A.2 Modélisation Merise	238
A.2.1 Modèle conceptuel de données (MCD)	238
A.2.2 Modèle logique de données (MLD)	240
A.2.3 Modèle physique de données (MPD)	240
A.3 UML	241
A.3.1 Notions UML	241
A.3.2 Correspondance Merise UML	242
A.3.3 Modélisation XML	243

A.4 La modélisation des flux de données	243
A.4.1 Problématique.	243
A.4.2 Les besoins	245
A.4.3 Rappels sur l'urbanisme	245
A.4.4 Une méthodologie et des outils communs	246
A.5 L'usage en pratique des modèles de données	249
Annexe B – SOA, services et données	251
B.1 Définitions	251
B.2 Objectifs et enjeux de la SOA	253
B.3 Services	255
B.3.1 Définition d'un service	255
B.3.2 Accès aux services	257
B.3.3 Services et services Web	258
B.4 SOA en pratique.	259
B.5 SOA et intégrations de données	259
Glossaire	263
Bibliographie	281
Index	283

Introduction

Le volume des données ne cesse de croître dans les entreprises et leur bonne gestion devient un problème crucial. En effet, chaque métier (marketing, vente, gestion des ressources humaines, comptabilité, production...) génère, et parfois gère, ses propres données. Dans la majorité des cas, elles sont dispersées en différents îlots, ce qui rend difficile la mise à disposition d'un système efficace de gestion (plus de 90 % des entreprises sont concernées selon une étude de *01 DSI* réalisée en France en 2007).

Notre objet n'est pas de nous arrêter à tel ou tel type de donnée spécifique, mais plutôt de proposer une vision, des outils et une démarche propres à s'adapter à l'ensemble des données à traiter.

Bien entendu les données sont une constante au sein des systèmes d'information (SI) et nous n'avons pas la prétention de réécrire 60 ans d'histoire de l'informatique. Cependant, plusieurs éléments nous encouragent à **établir un état des lieux et à proposer une approche synthétique d'une problématique transversale** :

- l'apparition d'applications capables de traiter spécifiquement de la gestion des données ;
- les nouvelles orientations technologiques au sein des systèmes d'information (SOA, BPM...) ;
- la valorisation et la protection de la donnée au service de l'entreprise ;
- la nécessité de se conformer aux réglementations de plus en plus nombreuses (CNIL, Art. 29, SOX, Bâle II...) ;
- l'émergence des notions de gouvernance des données associées à l'ensemble de ces mouvements.

Historiquement, le système d'information d'une entreprise s'est structuré autour des applications opérationnelles et des applications décisionnelles. La figure Intro.1 illustre cette situation. On trouve dans le système opérationnel de production toutes les applications transactionnelles de l'entreprise, que ce soit en *front*, *back-office* ou support (voir au chapitre 1 ces notions). Toutes ces applications génèrent la grande masse des données qui sont ensuite utilisées, notamment dans les applications décisionnelles, lesquelles ont besoin de données fiables et à jour pour produire les rapports et les analyses indispensables au pilotage de l'entreprise.

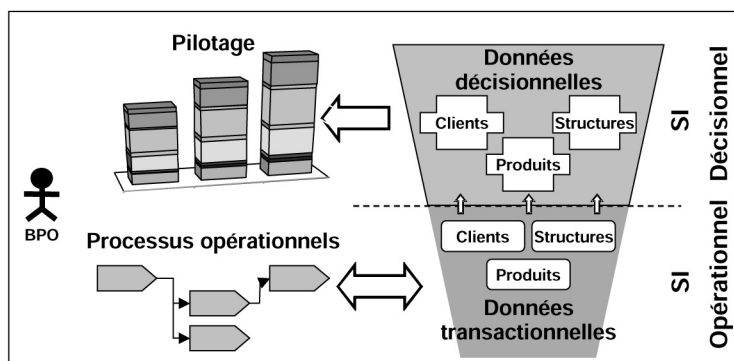


Figure Intro.1 – Système d'Information, modèle traditionnel

Dans cette vision traditionnelle, seuls les processus opérationnels métier et le pilotage associé ont une réelle valeur pour l'entreprise. **Les données sont dispersées dans les différentes applications** et il est souvent bien difficile de déterminer quelle application constitue la référence pour telle ou telle donnée métier. De même les problèmes d'incohérence entre applications, voire parfois entre sites d'hébergement d'une même application, sont difficilement maîtrisés.

Parmi toutes les données, certaines sont plus critiques pour l'activité métier et le système d'information car elles sont structurantes et largement partagées entre plusieurs applications : nous les appellerons **les données de référence**. Elles sont souvent disséminées, sans recherche de fertilisation croisée entre les différents processus et activités. Cela génère éventuellement des problèmes d'homogénéité et de rapprochements (dans les métiers et/ou dans le système d'information) et pénalise les gisements de valeurs liés à la donnée.

Prenons, comme exemple a priori fictif (voir figure Intro.2), le cas d'une entreprise qui utilise plusieurs canaux de communication avec ses clients : serveur web pour une application de vente en ligne, courrier électronique pour l'envoi des factures, téléphone pour la gestion des contacts de type demandes et réclamations. Si ces trois applications offrent leur propre vision client sans souci de partager une même référence, le client constatera vite l'incohérence du système d'information de cette entreprise. Il sera probablement perplexe, irrité, et doutera sérieusement de la capacité de la dite entreprise à gérer correctement ses clients !

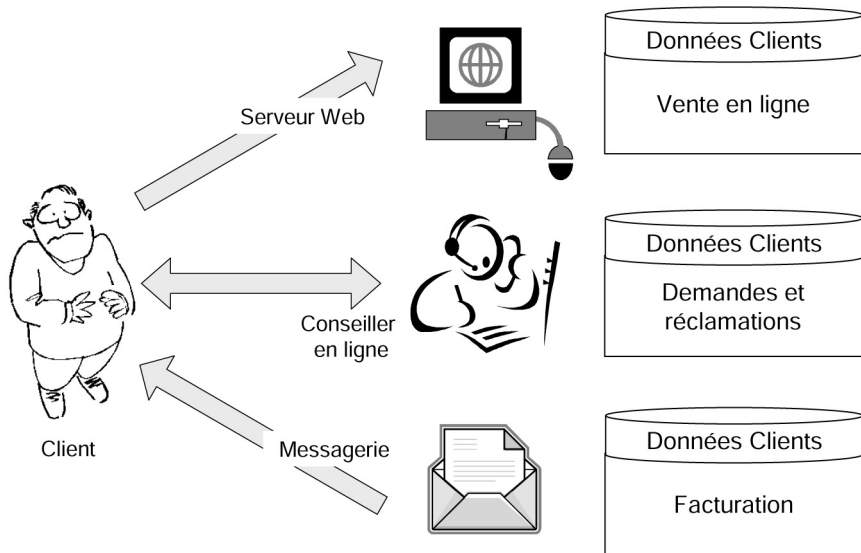


Figure Intro.2 – Un exemple de mauvaise gestion de données de référence

Il est ainsi évident que la notion de **MDM (Master Data Management)**, à savoir les disciplines nécessaires à « la gestion des données de référence métier », nécessite une rationalisation d'approches bien souvent disparates et hétérogènes. Il s'agit de dresser le bilan des dernières évolutions dans la prise en compte et le traitement des données de référence par les systèmes d'information bien sûr, mais peut-être encore plus par les métiers de l'entreprise.

C'est ce que nous allons présenter dans cet ouvrage, qui se décline comme suit :

- La première partie de l'ouvrage s'attache à la **compréhension des notions**. Il s'agit en particulier de définir le vocabulaire et de montrer la valorisation possible des données.
- La deuxième partie présente la **mise en œuvre d'architectures et de produits** pour améliorer la gestion des données. Le MDM (*Master Data Management*) y est en particulier détaillé.
- La dernière partie propose des **méthodes et des organisations pour le pilotage**, avec le concept-clé de gouvernance des données.

1

La gestion des données de référence

Objectif

Ce chapitre vise à répondre aux questions fondamentales relatives aux données de référence. Qu'est-ce qu'une donnée de référence ? Quels en sont les différents types ? Quels liens entretiennent-elles avec les autres notions relatives aux données ?

Il s'attache également aux questions relatives à leur gestion. Pourquoi constituent-elles un domaine de gestion à part entière ? Quels sont les besoins et les enjeux qui s'y rattachent, tant métier que **système d'information** ?

Il s'agit donc ici de fixer le vocabulaire de la gestion des données, de présenter les problèmes potentiels et d'évoquer les premières améliorations possibles.

1.1 PRINCIPES ET NOTIONS ÉLÉMENTAIRES

Dans les technologies de l'information, une **donnée** est une description élémentaire, souvent codée, d'un objet, d'une transaction d'affaire, d'un événement...

Les données ont une importance fondamentale pour toutes les activités de l'entreprise. Ces activités ou fonctions métier peuvent être de différents types : cœur de métier, *front*, *back* ou encore support. Ce sont, par exemple :

- La vente, qui est une activité *front* reposant sur les données produits et clients.

- La facturation en *back* et la comptabilité d'entreprise en *support* qui s'appuient sur des données telles que les prix et les comptes.
- Le pilotage de l'ensemble qui effectue la consolidation de toutes les données mobilisées lors des transactions.

La figure 1.1 schématise ces activités.

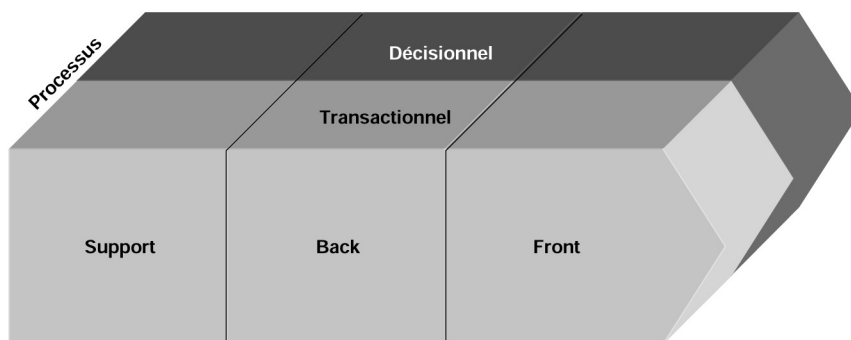


Figure 1.1 — Structure des activités de l'entreprise

Les activités de l'entreprise génèrent donc des données métier. Tous les processus métier utilisent des données comme éléments de base. Parmi ces données, certaines sont particulières car elles sont utilisées par :

- un grand nombre de processus métier ;
- plusieurs entités organisationnelles ;
- l'entreprise mais aussi ses partenaires.

Ces données sont généralement les **données de référence**. La première caractéristique de ces données est leur partage, leur utilisation par plusieurs applications du système d'information. Client, fournisseur, employé, article, contrat, offre sont des données de référence. Ces données sont les éléments de base de la gestion de toute entreprise. Elles sont indispensables à toute transaction commerciale, à tout contrat d'achat, à toute commande d'approvisionnement.

Dans la suite de l'ouvrage on utilisera les termes suivants que l'on considérera équivalents : donnée de référence, donnée maître, *master data*.

Tout ce que nous évoquerons à propos des données dites de référence peut être en général étendu à toute donnée partagée entre plusieurs processus et/ou applications.

La figure 1.2 présente quelques données de référence dans un système d'information. On voit que ces données, comme d'ailleurs l'ensemble des données du système d'information, sont dispersées et souvent dupliquées entre différents processus et applications.

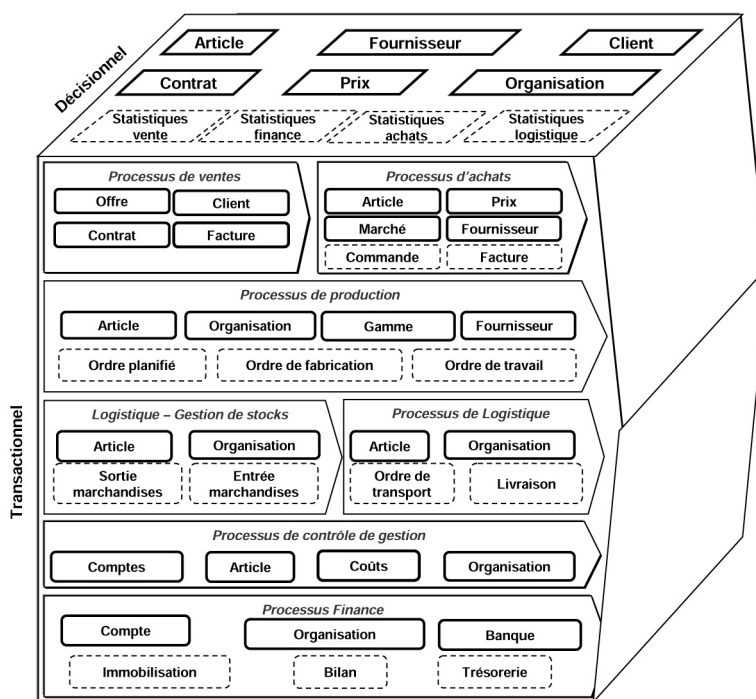


Figure 1.2 — Quelques données de référence dans un système d'information

1.1.1 Caractéristiques des données de référence

Selon que l'on se situe côté métier ou côté système d'information, différentes définitions peuvent être attribuées aux données de référence :

- Pour le manager, une donnée de référence est une information structurante de l'activité. Sa maîtrise offre un avantage à l'entreprise par une meilleure connaissance de son environnement et permet aux différents acteurs (internes ou partenaires) de travailler sur des bases communes.
- Pour le responsable de projet métier, une donnée de référence est une donnée des processus métier, partagée par plusieurs d'entre eux. Cette donnée revêt différentes dimensions en fonction de l'activité supportée par tel ou tel processus.
- Pour le responsable système d'information, une donnée de référence est une donnée **du système d'information, partagée et utilisée sans modifications**

par ses applications. Cette dernière caractéristique est essentielle pour identifier une donnée de référence.

La création ou la modification d'une donnée de référence est réalisée au travers de processus spécifiques. On nommera « processus référentiel métier » et par abus de langage « **processus référentiel** » les processus spécifiques de création ou de modification d'une donnée de référence. On nommera « processus opérationnel métier » ou « **processus métier** », les processus consommateurs structurés par les données de référence (voir chapitre 3 pour plus de détails sur ces notions).

Par essence, les données de référence ont, en général, **une relative stabilité** (relativement aux processus qui les utilisent). Une donnée de référence vit au-delà des instances de processus qui la mettent en œuvre. Par exemple, un client peut passer plusieurs commandes.

Les caractéristiques des données de référence peuvent évoluer sans remettre en cause leur nature ou leur « **unicité** ». Par exemple, un client peut déménager et donc changer d'adresse, mais il sera toujours identifié de la même façon.

Enfin, la donnée de référence doit être de **qualité** et respecter des **règles** communes à l'ensemble des entités utilisatrices.

1.1.2 Typologie des données de référence

On distingue trois grands types de données de référence qui appellent différents types de gouvernance et de socle technique (voir dans la suite ce que l'on entend par ces termes) :

- Les données « **maître** » sont en général les **objets métier principaux** (« **cœur de métier** ») **d'un domaine fonctionnel**. Ces données sont donc au cœur du **système d'information** et structurent les principales applications. En général, elles sont donc référencées dans de nombreuses applications.
Exemples : client, article, fournisseur...
- Les données « **constitutives** » sont des **données constituées elles-mêmes d'attributs, qui caractérisent en général des données maître**, mais aussi d'autres objets métier.
Exemple : adresse. Elle peut caractériser des données maître comme client, fournisseur... mais aussi d'autres objets métier comme interlocuteur (adresse postale) ou bon de livraison (adresse de livraison).
- Les données « **paramètre** » sont des **tables de valeurs** ou des **nomenclatures**.
Exemples : codes postaux, codes devises, taux des taxes des communes. Par essence, ce sont les données les plus partagées au sein du **système d'information** et donc celles qui doivent faire l'objet d'une attention particulière.

La figure 1.3 montre un exemple de ces différents types de données de référence et des liens associés.

On notera que la nuance entre donnée « maître » et « constitutive » dépend du sujet de l'analyse. En effet, une donnée maître dans un certain périmètre peut s'avérer être constitutive au sein d'une vision plus large.

Les données de référence peuvent être, par ailleurs, communes à divers domaines fonctionnels ou spécifiques à un domaine fonctionnel particulier.

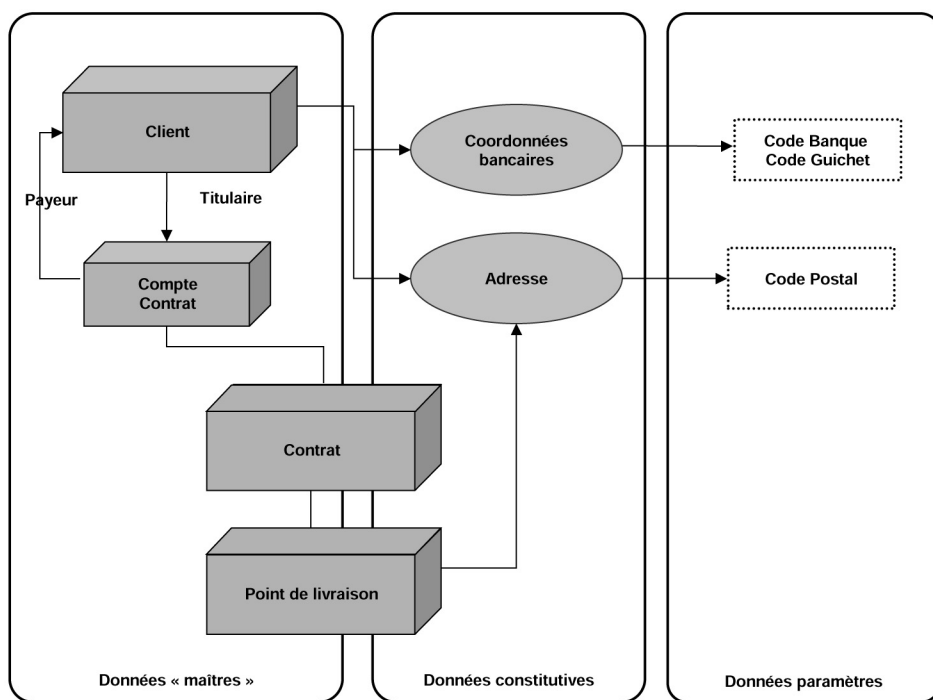


Figure 1.3 — Un exemple de différents types de données de référence.

Notons que cette distinction entre données « maître », « constitutive » et « paramètre » permet de mieux comprendre les différentes présentations de la suite de l'ouvrage mais n'est pas indispensable dans une étude informatique ou un projet.

1.1.3 La notion de famille de données

Si les différentes données de référence semblent pouvoir être analysées isolément, elles sont en fait liées entre elles, ce qui nous permet de parler de « famille de données ».

Ainsi, un référentiel « article » est lié à « fournisseur », un référentiel « client » est associé à « contrat » et pourtant les référentiels ne se substituent pas les uns aux autres.

On peut donc constituer des familles de données qui seront principalement dévolues à une activité de l'entreprise. Ceci aura notamment une incidence sur nos préconisations d'organisation et de gouvernance des données par la suite.

La figure 1.4 indique quelques exemples de données de référence et de regroupements propres à la structure d'une société de type « holding ».

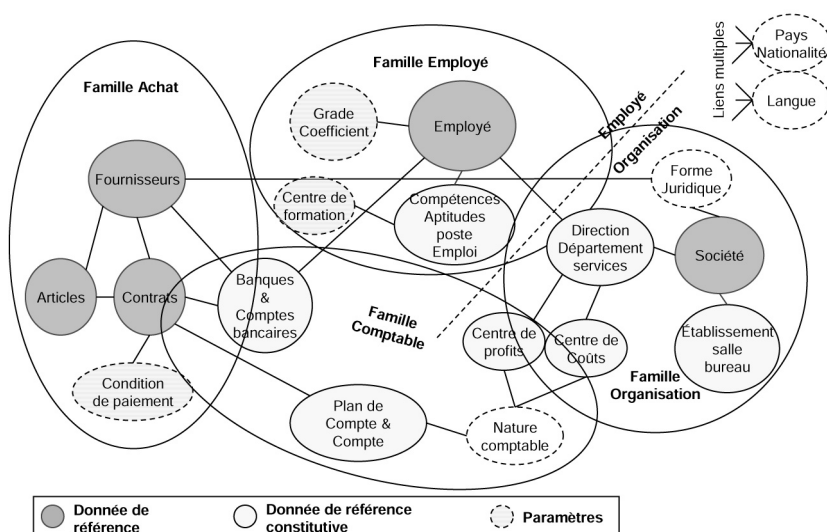


Figure 1.4 — Quelques exemples de regroupement en famille

1.1.4 Positionnement par rapport à d'autres notions

La figure 1.5 positionne les principaux concepts liés à la gestion des données de référence ¹. Ces concepts sont ensuite décrits en détail.

1. Voir le glossaire en fin d'ouvrage pour une définition complète de ces notions.

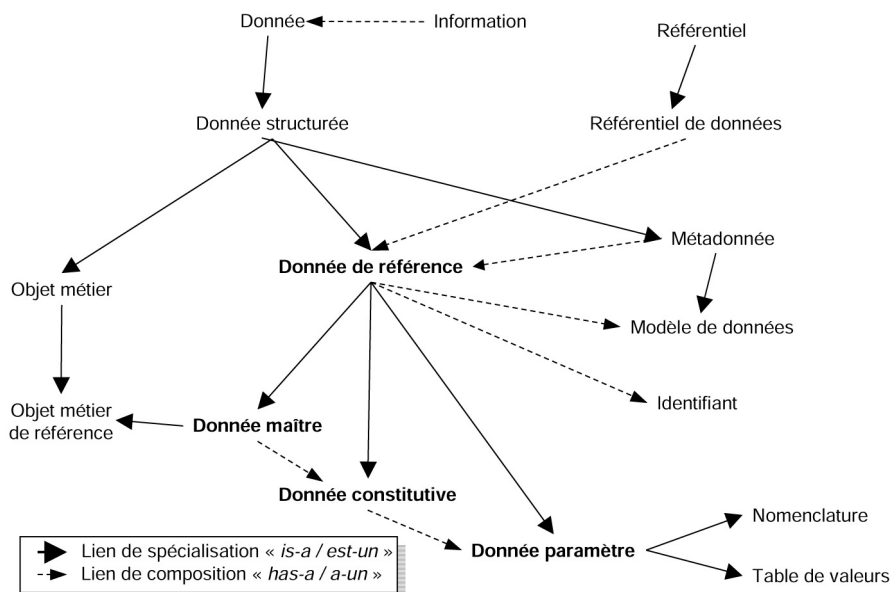


Figure 1.5 — Ontologie des données de référence

Donnée

Description élémentaire de nature numérique ou alphanumérique, représentée sous forme codée en vue d'être enregistrée, traitée, conservée et communiquée. Considérée individuellement, elle ne présente que peu d'intérêt humain et n'est donc utile que pour la machine.

Donnée structurée

Donnée dont on a établi fonctionnellement le sens de manière détaillée ainsi que les règles de création (dont les valeurs possibles dans le cas de listes) et enfin le moyen technique de représentation.

→ Une donnée de référence est une catégorie de donnée structurée.

Information

Données agrégées en vue d'une utilisation par l'homme (par exemple, le résultat d'une requête décisionnelle qui somme des données individuelles est une information). On parle aussi d'élément de connaissance susceptible d'être représenté à l'aide de conventions pour être conservé, traité ou communiqué (image, texte, donnée structurée).

La donnée sert à constituer l'information, cette dernière étant elle-même un élément de la connaissance (figure 1.6).

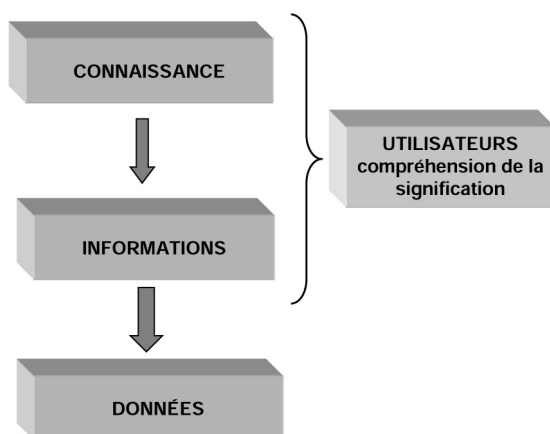


Figure 1.6 — Données, informations et connaissance

Objet métier

Ensemble d'informations homogènes du point de vue métier, manipulées dans le cadre des processus. C'est ce sur quoi porte une activité : l'objet métier matérialise le résultat du travail effectué par un acteur de l'entreprise. Ce sont des structures de données conçues pour représenter les livrables et les connaissances manipulées et gérées par les processus métier. Une structure de données est un ensemble organisé de données ayant quelque chose en commun et que l'on a groupées pour leur traitement (par exemple, la structure Identité est composée de civilité, nom, prénom, âge, profession).

→ Un objet métier peut être un objet de référence (donnée maître) s'il est structurant et partagé. Un objet métier peut lui-même inclure des données de référence (constitutive, paramètre).

Métadonnée

Littéralement, une donnée qui décrit une donnée. Nous reviendrons plus en détail sur cette notion dans les chapitres qui suivent.

→ Une métadonnée est une description métier et/ou technique d'une donnée.

Modèle de donnée

Représentation de l'ensemble des métadonnées ainsi que de leurs structures de regroupement et de relations sémantiques.

On distingue le modèle conceptuel du modèle logique et physique associé à un système de gestion de base de données (tables, attributs, clés). On y associera aussi la notion de modèle de flux (cf. annexe « Modélisation des données »).

Un modèle de données est donc une collection des descriptions des structures de données et des champs (attributs).

→ Dans le cadre de notre sujet, le modèle de données va servir à décrire précisément les données de référence, et en particulier leurs contextes d'utilisation avec l'ensemble de leurs données et relations.

Identifiant

Une donnée spécifique qui repère de manière unique un objet ou une information dans un système.

→ Un identifiant sert à repérer aussi les données de référence.

Exemple : le numéro SIRET identifie les établissements d'une entreprise.

Référentiel

Un référentiel s'apparente à un ensemble d'éléments dans lequel l'entreprise documente et agrège des règles de fonctionnement, techniques ou fonctionnelles.

Référentiel de données

Un référentiel de données est un « répertoire » clairement identifié (qui « fait référence »), qui stocke et permet la gestion et l'utilisation de données dans plusieurs traitements.

→ Un référentiel de données contient toutes sortes de données, de référence ou pas.

Par abus de langage, on confond souvent gestion des données de référence, référentiel et référentiel de données.

Table de valeurs (ou domaine de valeurs)

Liste des valeurs possibles d'une donnée.

Exemples : liste de segments marketing.

Nomenclature

Table de valeurs codifiées d'une donnée. Il s'agit d'une donnée de référence de type paramètre.

Exemple : la donnée Activité principale de l'entreprise (ou code APE) est exprimée selon la Nomenclature des activités françaises (NAF).

1.1.5 Gestion des données de référence (GDR)

Nous introduisons la notion de gestion de données de référence ou GDR (ou encore MDM pour *Master Data Management*). Il s'agit uniquement ici de positionner ce concept par rapport à d'autres plus connus. Nous y reviendrons bien entendu très largement tout au long de l'ouvrage.

Gestion des données de référence (GDR) : contexte général

La GDR (gestion des données de référence, *Reference Data Management*), ou MDM (*Master Data Management*), est une discipline des technologies de l'information qui concerne les données de référence. Elle est le garant de la cohérence entre diverses architectures de systèmes et fonctions métier.

La GDR ou MDM inclut d'autres disciplines afin d'atteindre ses objectifs.

Ainsi le DQM (*Data Quality Management*) est une des technologies nécessaires aux traitements relatifs à la qualité des données. Par extension, car la GDR repose sur l'ensemble des couches d'infrastructure d'échange et d'intermédiation du **système d'information**, on pourra aussi inclure dans le champ d'étude les notions d'EAI (*Enterprise Application Integration*), ETL (*Extract Transform Load*), EII (*Enterprise Information Integration*) et ESB (*Enterprise Service Bus*), etc.

En outre, on considère généralement la GDR ou MDM comme une discipline nécessaire à la mise en œuvre d'une SOA (*Service Oriented Architecture*). En effet, une vision métier de la SOA, et en particulier la mise en œuvre de processus transverses via le BPM (*Business Process Management*), repose sur la fourniture de services de données.

Enfin, la GDR ou MDM n'est-elle pas une sous-division d'une discipline plus large qui décline les bonnes pratiques et outille l'ensemble des aspects « données et information » au sein du **système d'information** ? Cette discipline, qui se nomme EIM (*Enterprise Information Management*), reste cependant à ce jour plus un trigramme de cabinet d'analyse qu'une fonction d'entreprise.

La figure 1.7 présente cette typologie.

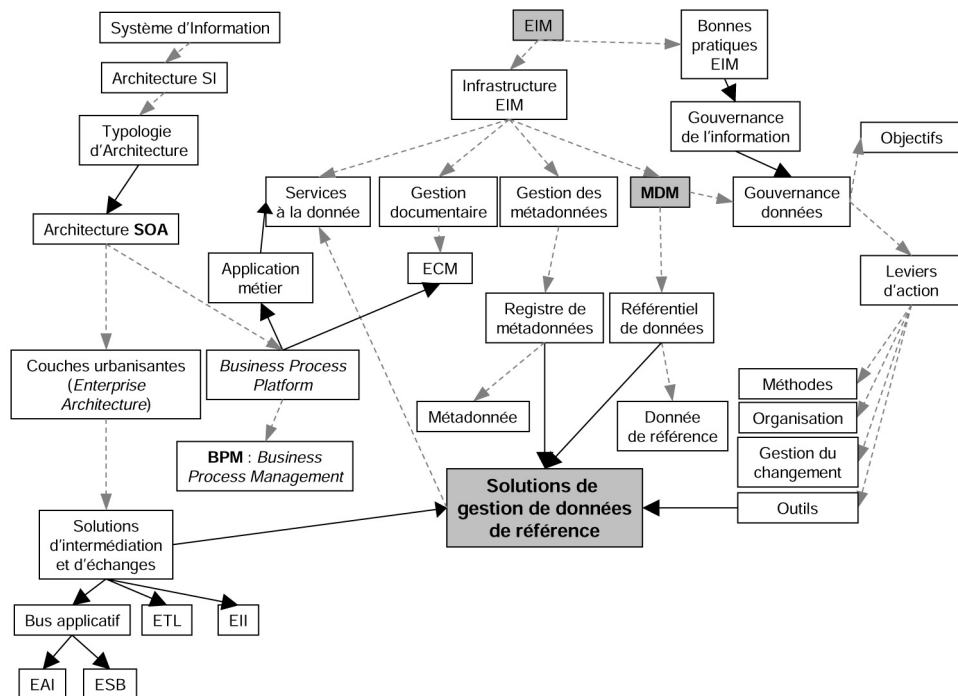


Figure 1.7 — Positionnement des termes

Les grandes entreprises et administrations disposent en général de systèmes informatiques qui outillent l'ensemble de leurs activités ou fonctions métier (recherche et développement, marketing, ventes, budget, finances...) et qui s'étendent au-delà des frontières, notamment pour les grands groupes. Ces différents systèmes ont souvent besoin de partager des données clés indispensables à la société mère (par exemple, les produits, les clients et fournisseurs). La problématique est la même quand deux entreprises ou plus veulent partager des données au-delà de leur propre périmètre. **La cohérence des données partagées entre différents systèmes informatiques est un besoin critique pour ces entreprises.**

Ainsi, la multiplicité des métiers, des pays, des lignes de produits, de services, des canaux de distribution, etc., contribuant à segmenter les activités de l'entreprise, génère une dissémination des référentiels (voir figure 1.8 l'exemple des référentiels marchands dans la grande distribution).

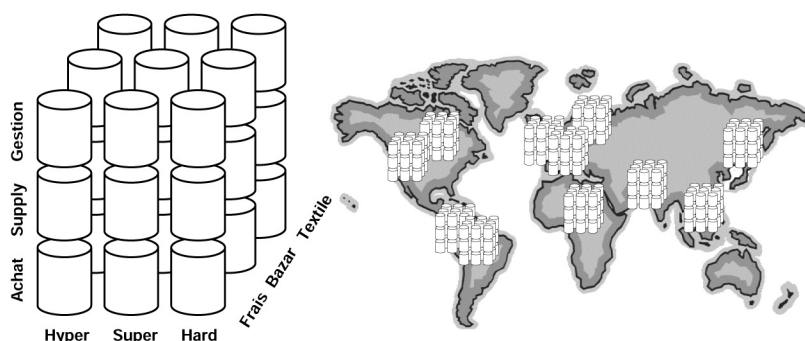


Figure 1.8 — La dissémination des référentiels marchands dans la distribution

Dans un modèle classique de système d'information, la gestion des données de référence est nécessaire pour coordonner entre eux les différents systèmes ERP, les systèmes légitimes, les autres progiciels et le décisionnel qui supportent cette dissémination. La gestion des données est donc une discipline essentielle pour l'agilité et la cohérence des métiers (fusion, acquisition, international, fondement des processus) et du **système d'information**.

1.1.6 Enjeux et besoins

On distingue ici les enjeux (ce qui va permettre de réaliser des gains) des besoins (ce qui est nécessaire pour satisfaire un manque).

Les enjeux et besoins listés ci-dessous sont génériques et à resituer dans le contexte particulier de chaque entreprise. Nous verrons plus loin comment le cadre de gouvernance des données peut aider chacune d'entre elles à identifier ses propres enjeux et besoins.

Index

A

accessibilité 31
acquisition 108
acteur 192
activités 46
actualité 30
administration de la preuve 43
alignement stratégique 54
amont 62
annuaire 97, 138
architecte de données 195, 198, 221
architecture 159, 175, 189
 d'entreprise 184, 189
 de centralisation 73
 de consolidation 69
 de coopération 70
 de répertoire virtuel 74
archivage 51
auditabilité 183
aval 62, 63

B

BDO 192, 196
bonnes pratiques 173, 217, 224

BPO 196
Business Data Owner 192, 196

C

cadre de la gouvernance 181
cartographie 155
CDI 131
centralisation 68
chaîne de l'information 64
chaîne référentielle 83
cleansing 187
cohérence 29, 89
complétude 28, 89
compréhensibilité 31
conduite du changement 184, 190
confident management 120
conformité 29, 89
conformité réglementaire 183, 186
consolidation 67
consommation 51
contextualisation 124
contraintes réglementaires 17
contrôle de synchronisation 130
coopération 67
coordinateur 194, 197

couverture du référentiel 80
critères de choix d'une architecture 79
CRM 86, 98
Customer Data Integration 131
Customer Relationship Management 86, 98
cycle de vie métier 48, 50
cycle de vie technique 51

D

data cleansing 34
data steward 194
decay management 120
document 246
donnée 9
 constitutive 6
 de référence 4, 5
 maître 4, 6
 paramètre 6
 structurée 9
DQM 86, 89, 210

E

EAI 148, 153
échange 136, 147, 152, 246
EII 86, 95
EIM 12, 229
ETL 149, 153
exactitude 28, 89

F

famille de données 7
filtre 127
flux 246
fusion 51

G

GDR 11
générateur d'identifiant unique 112
gestion de données de référence 11, 107
gestion de l'obsolescence 121
gestion de la confiance 121
gestion des droits 129
gestion des métadonnées 124
gouvernance 183, 201
gouvernance des données 87, 148, 179
groupe 110

H

hiérarchie 109
historisation 51, 125

I

identifiant 11
identifiant unique 112
identification 112
indicateur 120
information 9
intégrité 30, 89
intendant 194, 197
intermédiation 88, 147, 153

J

journalisation 126

K

key mapper 131
key mapping 115, 131

L

leviers 184
lien 109

M

master data 4
Master Data Management 11
matrice de maturité 202
MDM 11, 86, 108, 228
métadonnée 10, 40, 121
 administrative 42
 descriptive 42
 métier 41
 structurelle 42
 technique 41
migration 210
mise à jour 51
mode d'implémentation 81
modèle de document 246
modèle de donnée 10, 56, 121, 237
modèle de flux 148
multipropriété 193

N

nomenclature 6, 11, 109

O

objet métier 10
ontologie 9, 230
orchestration 136
organisation 188, 190

P

pertinence 31
pilotage 137
PIM 131
plan de transformation 157
PLM 86, 100
point à point 62, 148
point d'acquisition 62, 67
point de vérité 61, 67, 221
politique 185, 188
pré-référentiel 82
procédure 199
processus 45
 métier 6, 48
 référentiel 6, 48, 128, 209
Product Information Management 131
Product Lifecycle Management 86
profiling 90, 187, 210
propriétaire 192, 196, 220

Q

qualité 211
qualité des données 26, 33, 183, 186
quality assesment 90

R

recherche 127
référentiel 86
 analytique 82
 d'harmonisation 82
 de données 11
 de paramètres 114
 de tables 114

 de zone 82
 principal d'entreprise 81
règle 111, 199
 de cohérence 111
 de gestion 111
 syntaxique 111
répertoire virtuel 68
rôle 192, 198

S

saisie 108
sécurité 138, 183, 185
service 150
SOA 150, 251
socle référentiel 132
solution référentielle 157
solutions 161
solutions d'intermédiation 88
source de vérité 61
sponsor 192
stratégie 183, 184
suppression logique 52
suppression physique 52
synchronisation 223
système d'enregistrement distribué 82

T

table de valeurs 6, 11
tableaux de bord 120
transcodification 113

U

unicité 6, 28, 89
urbanisation 53
urbanisme 53, 174, 189

V

valeur d'une donnée 26
validation 111
versionning 130

W

Web services 258