

# Création d'un référentiel d'indicateurs de mesure de la qualité des données CRM

Delphine Clément  
Analyste Qualité des Données  
Hewlett-Packard  
14 rue du Général Caunègre – 40000 MONT DE MARSAN  
delphine\_clement@hp.com

Brigitte Laboisie  
Directeur technique  
A.I.D.  
4 rue Henri Le Sidaner – 78000 VERSAILLES  
[blaboisse@aid.fr](mailto:blaboisse@aid.fr)  
<http://www.aid.fr>

**Résumé :** Le sujet de cet article est de présenter le Référentiel d'indicateurs de mesure de la qualité des données Customer Relationship Management (CRM) publié avec l'Association Française de Normalisation (AFNOR). Il s'agit avant tout d'expliquer pourquoi ce référentiel a été créé, les grands chapitres, pour quels utilisateurs, et de fournir des éléments méthodologiques autour de l'identification du client/prospect. La deuxième partie de ce papier est la présentation d'un cas pratique avec l'implémentation de ce référentiel chez Hewlett-Packard, dans un contexte international.

## 1 Introduction

Analyse Informatique des Données (A.I.D.) est à l'initiative et a collaboré à l'élaboration d'un référentiel d'indicateurs de la mesure de la qualité des données CRM avec l'AFNOR. La première partie de cet article décrit le groupe de travail, la méthode pour constituer ce référentiel et présente les raisons de choix des différents indicateurs. Un focus particulier est fait sur les méthodes d'identification des individus : il est rappelé les méthodes proposées par les différents outils commerciaux, celle retenue dans le référentiel et il est proposé une méthode de comparaison pour utiliser les résultats de la méthode référentiel. A notre connaissance, peu de travaux ont été publiés sur les méthodes de comparaison de dédoublement et ce papier propose quelques évolutions par rapport aux travaux publiés.

La seconde partie de cet article est un exemple de mise en place de ce référentiel dans un contexte international chez Hewlett Packard. Les principaux freins rencontrés sont décrits ainsi que les solutions envisagées.

Enfin, en conclusion, les évolutions envisagées pour ce référentiel sont présentées.

## 2 Contexte

### 2.1 Le métier d'A.I.D.

A.I.D. est une société de services française spécialisée dans les Bases de Données marketing, la qualité de données et l'enrichissement statistique. A.I.D. travaille à l'international de part son appartenance au groupe de communication OMNICOM et les outils, référentiels développés depuis plusieurs années, et ce au niveau mondial.

Le métier d'A.I.D. est donc autour de la donnée, spécifiquement sur les données Marketing, CRM, 'Identification du client/prospect'.

### 2.2 L'absence de norme, de référentiels

Dans le domaine d'intervention d'A.I.D., il n'existe pas, à notre connaissance, de normes, de référentiels. La Poste française a mis en place Visa EV@®<sup>1</sup> et Labeladresse®<sup>2</sup> qui permettent de valoriser les bases d'adresses tenues correctement :

- Avec des adresses conformes à la norme Afnor XPZ 10-011 ou Z 10-011
- Avec des adresses en cours de validité

Cette première initiative, à notre connaissance, n'a pas été poursuivie sur d'autres canaux (téléphone, email), et quand on parle de taux de faux téléphones dans un CRM par exemple, les chiffres qu'on pourra obtenir seront en général :

- L'utilisateur marketing fournira le dernier taux de la campagne remonté par le centre d'appels
- Le centre d'appels inclura les sociétés ou personnes qui ne répondent pas au téléphone
- Le responsable Système d'Information, en général, n'aura pas de réponse.

La conséquence de cette absence de norme, de référence est bien connue : l'utilisateur se sent démuné, sans éléments de comparaison, et va, sur un sentiment, une impression, fournir un jugement sur la qualité des données :

- « Il y a trop de valeurs manquantes »
- « Il y a trop de valeurs erronées »
- « C'est inutilisable »
- « En effet, j'ai comme l'impression ces derniers temps d'une dégradation de la qualité de notre base de données, ce que je ne peux accepter. Je compte donc sur vous pour faire le nécessaire »

---

<sup>1</sup> Visa EV@® est une marque déposée de La Poste.

<sup>2</sup> Labeladresse® est une marque déposée de La Poste.

### **3 Création d'un référentiel d'indicateurs de la mesure qualité des données d'un CRM**

#### **3.1 Le livrable**

A.I.D. a entamé en 2005, en participation avec l'AFNOR, la démarche de mise en place d'un référentiel d'indicateurs qui permettent de mesurer la qualité de données d'un CRM. Le résultat est un document : AC X50-111 disponible à l'achat sur le site de l'AFNOR <http://www.boutique.afnor.org/NEL1AccueilNormeEnLigne.aspx>

Ce document n'est pas une norme, mais plutôt un 'guide de bonnes pratiques'. Il a été rédigé par les auteurs de cet article sous la direction d'un consultant AFNOR. Un groupe de professionnels : statisticiens (Gilbert Saporta, Conservatoire National des Arts et Métiers (CNAM)), responsables marketing (Air Liquide par exemple), responsables CRM (Auchan par exemple), responsables qualité de données (Hewlett Packard par exemple) a permis de valider la pertinence, l'approche opérationnelle des indicateurs retenus afin de mesurer la qualité des données d'un CRM, sur la partie Identification des individus.

#### **3.2 Identification des individus**

L'identification des individus est un vaste sujet, largement abordé dans la littérature Qualité de Données : Madnick et al. (2002) présentent par exemple une méthode d'identification des entreprises prenant en compte la notion de contexte.

##### **3.2.1 Qu'est-ce qu'un individu ?**

On distingue, en général :

Dans une base de données Business to Consumer (B2C), 2 types d'individus :

- La personne physique : Raoul Dupont, numéro de sécurité sociale 1XXXXXXXXXXXXX
- Le foyer : [1] regroupement de personnes vivant à la même adresse. Cela peut inclure les couples mariés, les concubins, leurs ascendants ou descendants vivant à la même adresse.

Dans une base de données Business to Business (B2B), 2 types d'individus :

- La personne physique : Raoul Dupont, numéro de sécurité sociale 1XXXXXXXXXXXXX
- Le site : [1] Etablissement d'une entreprise : nom, adresse d'une entreprise. La notion de site est en général assimilée à une relation bijective (1,1) avec un numéro issu du Système d'Identification du Répertoire des Etablissements (SIRET) (identifiant unique fourni par l'Institut National de la Statistique et des Etudes Economiques (INSEE), mais on peut noter des exceptions : adresse non déclarée à l'INSEE : par exemple, dans le cas d'un entrepôt ou annexe. On aura alors une notion de site sans SIRET.

##### **3.2.2 Que signifie identifier un individu?**

- Qu'il est représenté d'une manière unique dans le système d'information, de manière non ambiguë
- Qu'on peut s'adresser à lui : soit par courrier, soit par téléphone, fax ou courriel
- Qu'on connaît son 'état' : décédé, en cessation d'activité,...

## Création d'un référentiel d'indicateurs de mesure de la qualité des données CRM

Il s'agit de retrouver dans un ensemble d'enregistrements ceux qui correspondent au même individu. Par exemple :

### Exemple 1 :

DUPONT RAOUL	A.I.D.	4 rue le Sidaner	78000	Versailles	0139239345
DUPONT RAOUL	A.I.D.	rue Henri le Sidane	79700	Versailles Cedex	

### Exemple 2 :

DUPONT RAOUL	A.I.D.	4 rue Henri le Sidaner	78000	Versailles	0139239345
DUPONT RAOUL		6 rue des Acacias	78150	Le Chesnay	0139239345

La détection de doubles est différente sur une personne, un foyer ou un site.

### Détection de doubles sur une personne :

Les procédés utilisés consistent en général à 'nettoyer' les données, les confronter à un référentiel afin de pouvoir les comparer ensuite à l'identique entre elles à l'aide de 'match-keys' ou clés de rapprochement. Par exemple, on va comparer les adresses aux référentiels de La Poste (HEXAVIA, HEXAPOSTE) et corriger l'information erronée.

Dans notre exemple 1, le résultat sera :

4 rue le Sidaner	78000 Versailles	→ 4 rue Henri le Sidaner	78000 Versailles
rue Henri le Sidane	79700 Versailles	→ rue Henri le Sidaner	78000 Versailles

Cette correction par un référentiel est en général utilisée pour l'adresse, le prénom. Sur le téléphone ou l'email, il s'agit plus d'un nettoyage de la syntaxe de part le manque de référentiels.

La comparaison à l'identique ensuite sur des clés de rapprochement (par exemple les 4ers caractères du nom, la 1<sup>ère</sup> lettre du prénom, le nom de la rue, et la ville) est motivée par un besoin de rapidité : les volumes concernés sont en général importants.

### Détection de doubles sur un site :

Les méthodes utilisées sur les personnes ne sont pas performantes sur les sites :

- Un nom de site pourra apparaître sous forme de sigle : A.I.D. ou Analyse informatique de Données
- Ce nom pourra inclure des mots creux ou synonymes : Société Anonyme (SA), Mairie ou Commune,...
- On trouvera souvent une adresse physique (la rue) ou une adresse postale (boîte postale)

La comparaison à l'identique sur des clés de rapprochement ne fonctionne pas. Les logiciels du marché (Principaux éditeurs : Fuzzy Logic, Business Objects, Harte&Hanks, ..) consistent à comparer tous les enregistrements sur un sous ensemble de la population (ville ou département par exemple) et à fournir des règles telles que :

Si  $\text{index1}(\text{nom1}, \text{nom2}) > 90$  et  $\text{index2}(\text{adr1}, \text{adr2}) > 70$  alors double

Index1 et Index2 sont des index de proximité entre les 2 noms nom1 et nom2. Les index implémentés dans les logiciels peuvent se baser sur la phonétique, sur le nombre de modifications entre les 2 chaînes par exemple (Berti-Equille (2005) fournit une liste d'algorithmes).

Comparaison de méthodes :

Les méthodes d'identification présentées ci-dessus pourront donner des résultats différents, de part les algorithmes de proximité utilisés, mais également des règles de comparaison (variables retenues, combinaisons, seuils). Comparer les résultats de différents outils, ou du même outil sur des règles différentes est un travail laborieux, souvent fait en manuel sur quelques enregistrements. Des méthodes automatiques Hashemi et Talburt, (2006), Talburt et Hess (2004), Wang (2000) émergent : le principe consiste à comparer 2 partitions A et B et à fournir un index de proximité : nombre d'enregistrements classés ensemble par les 2 méthodes.

Soit l'exemple ci-dessous : 2 dédoublements ont été menés, qui ont permis de classer la population en A et B. Dans notre contexte, par exemple, le dédoublement A est le résultat de la méthode préconisée dans le référentiel et le dédoublement B correspond à un test fait avec les outils usuels du CRM pour corriger les doubles détectés par A. On observe ainsi que la partition B3 a mis ensemble 3 individus, alors que ces derniers se répartissent pour 2 individus dans la partition A3 et pour 1 dans la partition A6. Les pourcentages horizontaux et verticaux fournissent le taux de convergence entre les 2 partitions : 100% de A3 est inclus dans B3, 66,67% de B3 est inclus dans A3. On peut calculer un index de proximité entre les partitions A et B :

Index de proximité =

$$\frac{\text{Nombre d'individus dans les cellules avec un \% vertical et horizontal } \geq \text{seuil}}{\text{Nombre d'individus}}$$

Soit un seuil à 90% : 2 partitions seront considérées comme convergentes si  $A \subset B$  à 90% et réciproquement  $B \subset A$  à 90%.

Dans notre exemple =  $(1+1+1+1)/10 = 5/10 = 0.5$

Par rapport aux travaux publiés cités ci-dessus, l'index que nous proposons prend en compte l'effectif dans les cases (pas uniquement 0/1 si la case est remplie) et filtre les intersections selon un pourcentage paramétrable. L'intérêt est d'écarter les différences peu significatives et de prendre en compte les volumes importants de convergence.

## Création d'un référentiel d'indicateurs de mesure de la qualité des données CRM

	B1	B2	B3	B4	B5	B6	B7	Total
A1	1							1
A2			1					1
A3				2				2
A4	1							1
A5					1			1
A6							1	1
A7						1		1
A8							1	1
A9								1
Total	2	1	3	1	1	1	1	11

Fréquences

  

	B1	B2	B3	B4	B5	B6	B7	Total
A1	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
A2	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
A3	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	100,00%
A4	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
A5	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	100,00%
A6	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	100,00%
A7	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	100,00%
A8	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	100,00%
A9	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	100,00%

% horizontaux

  

	B1	B2	B3	B4	B5	B6	B7
A1	50,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
A2	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%
A3	0,00%	0,00%	66,67%	0,00%	0,00%	0,00%	0,00%
A4	50,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
A5	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%
A6	0,00%	0,00%	33,33%	0,00%	0,00%	0,00%	0,00%
A7	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%
A8	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%
A9	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
Total	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

% verticaux

TAB 1 – Tableau de comparaison de 2 dédoublemnages A et B : fréquences, % horizontaux, % verticaux

### 3.3 Mon CRM identifie-t-il correctement les individus ?

La démarche de mesure proposée dans le document AFNOR est volontairement pragmatique et simple. Il ne s'agit pas de trouver l'ensemble des doubles mais bien de mesurer et ce de la même manière, quel que soit le CRM. La méthode de mesure proposée, inspirée des méthodes décrites plus haut, est en 2 étapes :

- Travail manuel sur un échantillon : on recommande d'extraire une zone géographique ou une lettre de l'alphabet selon la complétude de l'information.
- Un calcul complémentaire par clés de dédoublemnage sur l'ensemble de la population est recommandé afin d'essayer de corriger le biais de l'échantillon précédent : le taux de doubles peut varier sensiblement d'une zone géographique à l'autre.

Des recommandations sur le choix de l'échantillon, les clés de dédoublemnage sont fournies ainsi que sur le calcul du taux de doubles.

Il sera possible, en fonction des résultats, de donner ces éléments de contrôle aux responsables d'outils de dédoublemnage afin qu'ils comparent (voir méthode de comparaison ci-dessus) et améliorent leur dédoublemnage en se comparant aux résultats manuels obtenus.

### 3.4 Puis je contacter les individus de mon CRM ?

Puis-je m'adresser aux individus par un des canaux usuels : téléphone, courrier, email ? C'est-à-dire, ai-je au minimum les coordonnées pour un canal, et cette donnée est-elle exacte ? La mesure de l'exactitude est proposée selon 2 méthodes :

- Les indicateurs a priori : on mesure, en général sur des échantillons, la qualité de l'adresse, du téléphone, de l'email. On parle ici d'indicateur a priori dans le sens où l'information est contrôlée avant son utilisation ou de manière indépendante.
- Les indicateurs a posteriori : ils résultent de l'usage observé opérationnel des données : taux de N'habite Pas à l'Adresse Indiquée (NPAI), taux de faux téléphones, taux de faux emails,...

Avantages et inconvénients de chaque type d'indicateurs :

#### Indicateurs a priori

Avantages : Mise en œuvre indépendante de l'usage des données, actions préventives possibles avant l'utilisation des données.

Inconvénients : Difficile d'avoir une vue complète de la qualité des données : la validation d'un téléphone passe à ce jour par un appel pour une meilleure efficacité. Les référentiels ne sont pas assez exhaustifs pour être efficaces. Méthode de mesure souvent mal reconnue par les utilisateurs

#### Indicateurs a posteriori

Avantages : Vue réelle de la qualité des données à l'utilisation. Méthode de mesure bien acceptée par les utilisateurs

Inconvénients : Vue partielle sur la qualité de données : périmètre utilisé peut être faible.

Difficulté importante à avoir les retours d'une manière exhaustive et fiable (voir ci-dessous)

Le référentiel décrit l'ensemble de ces indicateurs, avec une recommandation sur la mise en place. Les freins, problèmes rencontrés, les biais sont décrits avec des solutions palliatives autant que faire ce peut.

## 4 Mise en œuvre opérationnelle : cas pratique

Dans le cadre de la gestion de la relation clients sur le périmètre des petites et moyennes entreprises en Europe, Moyen-Orient et Afrique (EMEA), Hewlett-Packard pilote, depuis un an, la mesure de quatre indicateurs a posteriori que sont le taux de NPAI, le taux de « bounced » emails ou emails non aboutis, le taux de faux téléphones et le taux de contacts obsolètes. Cette mesure est produite à la fin de chaque trimestre, avec un décalage d'un à deux mois par rapport à la fin du trimestre, afin d'assurer la récupération d'un maximum de NPAI.

### 4.1 Méthodes de collecte des informations non abouties et difficultés rencontrées

#### NPAI

Chaque pays de la région a sa propre méthode de traitement des NPAI. Nous pouvons regrouper les méthodes en 3 catégories principales :

## Création d'un référentiel d'indicateurs de mesure de la qualité des données CRM

1. Les pays qui traitent les NPAI localement et produisent eux-mêmes leur taux de NPAI. Ces pays se contentent de nous fournir un taux déclaratif ; nous n'avons malheureusement pas accès à l'enregistrement site et contact en retour, ce qui nous empêche d'analyser les caractéristiques de cet enregistrement, d'investiguer sur les causes du retour et sur les possibles corrélations avec d'autres facteurs. Cependant, à défaut, la mesure de l'évolution dans le temps de ce taux est toujours intéressante.
2. Les pays qui sous-traitent les NPAI à une agence locale. Pour ces pays, il nous est possible de récupérer les enregistrements en retour ainsi que le code campagne auquel ils se rattachent. Ceci étant, chaque pays ayant sa propre agence, les fichiers récupérés sont tous de différents formats, le niveau d'information retourné varie également d'un pays sur l'autre ; pour certains pays, nous arrivons à obtenir le motif de non-distribution indiqué sur l'enveloppe par exemple, alors que pour d'autres, nous n'avons que l'identifiant de l'enregistrement, à nous de faire ensuite un croisement avec le système d'information pour récupérer les données. De plus, même si nous attendons de un à deux mois après la fin du trimestre avant de produire la mesure, nous ne sommes jamais certains de récupérer l'exhaustivité des NPAI du trimestre.
3. Les pays qui passent en gestion centralisée de leur NPAI. La Suisse, les Pays-Bas et la Suède ont expérimenté cette méthode. Cependant, l'équipe centralisée étant en Inde, il est extrêmement onéreux d'y envoyer des caisses entières de catalogues en NPAI ; les frais engagés, ainsi que les problèmes de douane rendent cette solution non viable. Dès lors, une telle méthode doit être combinée avec une gestion locale (au sein du CRM ou au sein d'une agence) pour pallier ce problème; ce qui ne facilite pas, pour les analystes qualité, la collecte des enregistrements en retour et l'analyse.

### Emails non aboutis (Hard Bounce)

Les campagnes d'emailing de Hewlett-Packard sont gérées centralement par une agence externe. De ce fait, il nous est assez facile de récupérer le fichier de campagne que cette agence publie, chaque mois, pour tous les pays, sur le site intranet de Hewlett-Packard. Ce fichier de campagne comprend l'identifiant du correspondant contacté, son nom, son prénom, sa fonction, la société pour laquelle il travaille ainsi que son opt-in, c'est-à-dire son accord à recevoir des communications marketing de la part de Hewlett-Packard sur sa boîte email. Ce fichier comprend aussi le code campagne, et surtout le statut de l'envoi. C'est en analysant ce statut que nous séparons les emails non aboutis en « soft bounces » et « hard bounces » (voir lexique ci-dessous).

Cependant, il est à noter que le statut « hard bounce » n'est pas nécessairement fiable. Par exemple, nous avons pu observer, sur une période d'étude de 3 mois, que des contacts non aboutis « hard bounce » peuvent très bien être aboutis le mois suivant. Ce qui nous a incités à créer un compteur de « hard bounces ». Ainsi, un contact doit être comptabilisé trois fois de suite (ou plus ou moins) en statut « hard bounce » pour être vraiment considéré comme tel.

### Faux téléphones

Les campagnes de télémarketing de Hewlett-Packard sont gérées différemment selon les pays et selon le type de client ou le type de campagne. Parfois, ce sont les centres d'appels internes qui sont utilisés, parfois ce sont des centres d'appels locaux ou régionaux. Cette diversité est source de problèmes dans la collecte des informations concernant les faux téléphones. En effet, chaque centre d'appel a son propre script d'appel, sa propre codification et dès lors, il est très difficile d'interpréter avec certitude la signification du statut « faux

téléphone ». Certains centres d'appel vont en effet qualifier comme tel un téléphone non abouti au bout de 3 à 5 tentatives, alors que d'autres ne vont utiliser ce même statut que pour un téléphone ayant une sonnerie erronée; d'autres enfin vont faire la distinction entre « société ayant déménagé – le téléphone n'est donc plus le bon » et « téléphone à la sonnerie erronée ». Cette absence de standards quant aux différents codes ou statuts entrave la bonne analyse et la bonne interprétation. L'harmonisation des codes retour est donc clairement une piste d'amélioration pour l'avenir.

#### Contacts Obsolètes

Les contacts obsolètes sont dans notre cas collectés lors des campagnes de télémarketing, c'est à dire d'appels sortants. Ainsi que mentionné précédemment, dans la rubrique faux téléphones, notre principal souci est le manque de cohérence entre les codes retour et les statuts « contacts obsolètes ». D'aucuns y mettront un contact qui a quitté la société dans le cadre d'une retraite, d'un licenciement, d'un décès ou d'une démission et d'autres y mettront aussi un contact qui a quitté la fonction pour laquelle il est enregistré dans la base de données mais qui a été promu à une autre fonction au sein de la même société.

Lors de la mesure du dernier trimestre, nous avons exploré une piste qui consiste à croiser les informations avec des informations provenant de bases externes. Certains fournisseurs sont en effet en mesure de fournir, plus ou moins précisément, des informations concernant les contacts obsolètes.

Enfin, nous explorons également l'idée d'analyser une population de contacts obsolètes pour voir si certains facteurs expliqueraient l'obsolescence du contact, ce qui nous permettrait d'aboutir à un modèle de prédiction du taux de contacts obsolètes dans une population définie.

## **4.2 Observations de nos mesures**

Avec le recul d'un an de mesure de ces 4 indicateurs a posteriori, nous constatons que notre principal problème se pose au niveau du taux de faux téléphones. Nous étudions actuellement diverses pistes pour améliorer la qualité du téléphone, car nous savons par ailleurs qu'un mauvais téléphone coûte très cher à Hewlett-Packard, non pas tant en coûts directs, mais davantage en perte d'opportunités, en perte de revenu.

## **Conclusion**

La publication de ce référentiel avec l'AFNOR est une première marche vers une norme. Notre objectif est maintenant :

- Que ce référentiel soit connu, reconnu à la fois par les professionnels du CRM, les utilisateurs et utilisé. Parvenir à un langage commun, à des indicateurs communs correspond maintenant à un travail de communication de ce référentiel, et de co-optation.
- L'étape suivante est de publier des niveaux de bonne conduite : avec le recul d'utilisation de ces indicateurs, nous avons pu voir apparaître des tendances, des niveaux de bons résultats et à contrario des taux qui sont très mauvais. La difficulté

## Création d'un référentiel d'indicateurs de mesure de la qualité des données CRM

est d'avoir suffisamment d'expérimentations, de cadres d'utilisation pour produire des niveaux qui aient un sens : nous avons pu observer que :

- Ces taux varient beaucoup d'un pays à l'autre : par exemple, en B2B, les fichiers disponibles sur le marché sont de qualité très inégale. En Italie ou en Espagne, sur des fichiers généralistes, un taux de 25% de faux téléphones est courant. En revanche, en Allemagne, un taux de 10% est plus usuel.
- Les taux varient également selon le segment d'entreprises : des grandes entreprises, mieux suivies, avec plus de valeur et de potentiel, auront des taux qualité bien meilleurs que des Petites et Moyennes Entreprises (PME).
- Enfin, dans le cadre de l'application chez Hewlett-Packard, une autre démarche a été initiée. Chaque faux téléphone, adresse fautive a des conséquences financières : directes par les coûts supplémentaires engendrés (courriers en trop, tarifs postaux supplémentaires,...), mais également par des opportunités ratées. Il s'agit de valoriser ce coût de la non qualité et de pouvoir déclencher des actions de correction, enrichissement selon leur rentabilité prévue.

## Lexique

Customer Data Integrity : Intégrité des données Client

Customer Knowledge Management and Data Stewardship : Gestion de la connaissance client et services sur les données

Internet and Marketing Services : Services Internet et Marketing

B2B : Business to Business : Vente aux entreprises

B2C : Business to Consumer : Vente au grand public

Benchmark : Comparaison

Major evolutions : Evolutions majeures

Bounced Emails: on distingue les soft bounces, non aboutis temporaires (la boîte email du correspondant est pleine par exemple) des hard bounces, non aboutis permanents (l'adresse email du correspondant est fautive par exemple) ; c'est en fait la population des emails non aboutis

Emailing : campagne marketing par envoi d'email

NPAI : N'habite Pas à l'Adresse Indiquée

## Références

AFNOR, A.I.D. (2006), *Mesure de la qualité des données CRM – Référentiel d'indicateurs CRM*. Accord AC X50-111

Berti-Equille, L. (2005), *Journées CRM & Qualité des Données au CNAM – Qualité des données multi-sources : un aperçu des techniques issues du monde académique*.

Hashemi Ray R., Talburt John R., (2006) *Significance Test for the Talburt-Wang Similarity Index* – The 11<sup>th</sup> ICIQ (ICIQ-06, MIT IQ) Conference Program

Madnick S., Wang R. et Zhang W. (2002), *A Framework for Corporate Householding* pp. 36-46. Proceedings of the 7<sup>th</sup> International Conference on Information Quality.

Talbur John R., Hess Kimberly, Wang Richard, Kuo Emily, (2004), *An Algebraic Approach to Quality Metrics for Customer Recognition Systems* – ICIQ-04 (MIT IQ) Conference Program.

Wang, R. (2000), *Data Quality*. Kluwer Academic Publishers

## **Summary**

This article published by AFNOR, presents the A.I.D. dictionary of CRM data quality measurement indicators. It explains why the dictionary was created and for which users, and gives the underlying methodology to help find clients and prospects. The second part of the article is a practical case study based on Hewlett Packard in an international context.