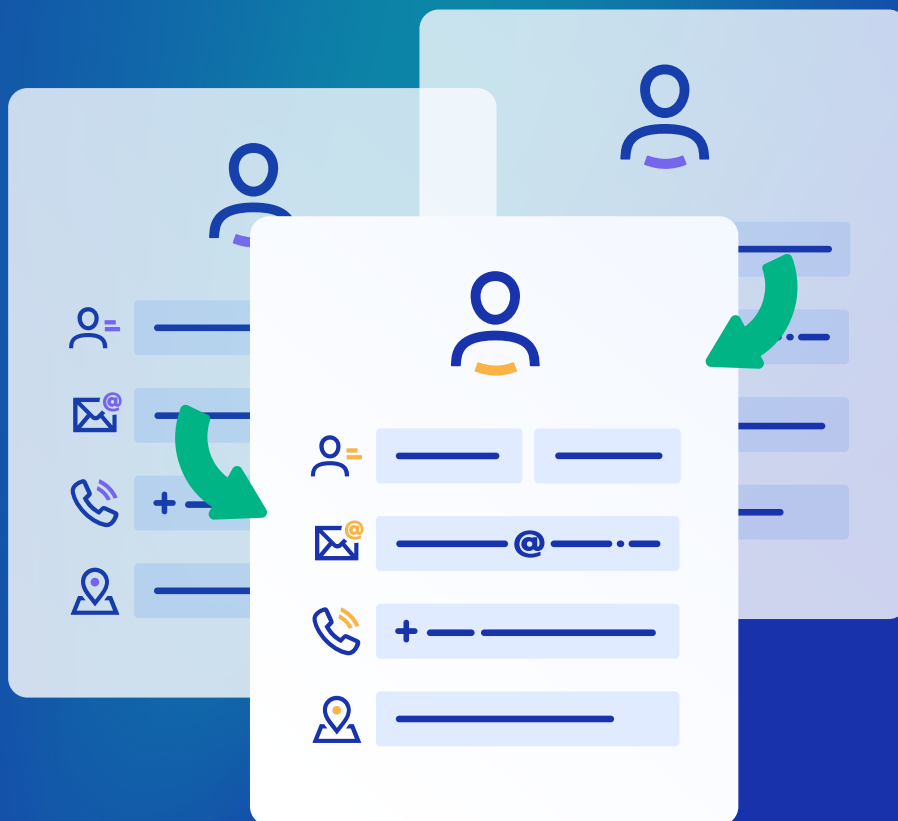


UNIFIER LES DONNÉES CLIENT

Guide des best practices de déduplication et de fusion





INTRODUCTION

Lorsqu'une base de données client compte 10 % de doublons, elle atteint un seuil critique. Il n'est pas rare de voir des entreprises disposer d'une base ayant plus de 20 % de doublons. Beaucoup ne mesurent pas l'ampleur du problème.

La plupart des bases client sont polluées par des doublons ayant des incidences parfois désastreuses sur les efforts des différents services de l'entreprise : les impacts sont majeurs sur le clienteling, la satisfaction ou la rétention, mais aussi dans l'efficacité des services client ou le reporting des services commerciaux.

Aussi, unifier les données clients fait figure de projet stratégique. Pour de nombreuses entreprises, ce défi soulève énormément de questions, voire des réticences généralement liées à la transformation des processus d'alimentation des différents outils.

**Où se situe votre entreprise ? Que dédoubler ?
Quand et quoi fusionner ?**

Aujourd'hui, le Data Quality Management apporte une réponse aux problèmes liés aux doublons. Grâce à l'unification des données, il a vocation à traiter les données client et en assurer la qualité et la fiabilité. Ce livre blanc présente les best practices éprouvées chez les clients de DQE, qui ont permis de mener leur projet d'unification des données client à son terme avec réussite.





1.

Les doublons, un fléau dans les bases client

Toutes les entreprises disposent de leur base de données client. La qualité de cette donnée est un enjeu universel et le traitement des doublons fait intégralement partie du Data Quality Management. Il importe de cerner, en premier lieu, les processus générateurs de doublons.

D'où viennent les doublons ?

La source majeure réside dans le cloisonnement des différents systèmes d'informations de l'entreprise, qui démultiplient les bases de données - base CRM marketing / ventes, service client, ERP, points de vente, portail Web... Cette configuration engendre des doublons dans le système d'information : pour chaque point de contact, la collecte de données client a toutes les chances de déclencher une nouvelle fiche client. Inexorablement, les doublons se propagent à l'échelle de l'entreprise. Pour autant, chaque service est conscient de la valeur d'une information de qualité.

D'autres sources de doublons peuvent apparaître. C'est le cas de biais organisationnels, notamment lorsque la création de nouvelles fiches client est incentivée dans une optique de new business. Cette pratique encourage les opérationnels à créer systématiquement une fiche client en double.

Quitte à utiliser des contournements quand un outil opérationnel bloque la saisie multiple d'une même adresse email. Une petite modification à la marge - un point en plus dans une adresse email, une inversion volontaire de lettres - permet ainsi de « forcer la main » au système qui identifiera la fiche comme différente.



Les conséquences des doublons peuvent être désastreuses :



Service client, call-center et points de vente

Les doublons mettent les conseillers sous pression. Lors d'un appel téléphonique, plusieurs fiches client apparaissent à l'écran pour un même client, laquelle choisir ? À laquelle se fier ? Quelle fiche compléter ?

Impossible d'avoir une vue à 360° du client, de ses interactions avec l'entreprise – commandes, appels au SAV, visites en agence ou point de vente, précédentes réclamations... Comment dès lors donner des réponses rapides et pertinentes, la principale mission d'un service clients ? Ces doublons deviennent un frein majeur à l'efficacité du service client.



Le Service Marketing et CRM

Ce service peut difficilement conduire des campagnes avec une segmentation juste de sa clientèle si les données client sont éparpillées sur différents comptes.

Par ailleurs, une base client truffée de doublons peut vite engendrer des envois à répétition aux mêmes personnes, ou des messages inappropriés, par exemple une incitative pour un premier achat envoyée à un client régulier. Quand une entreprise défend une image de marque Premium, ces approximations rédhibitoires pour les clients contredisent le message.



Les forces de vente

Elles subissent la dégradation des chiffres qui s'appuient sur les données client doublonnées. Exemple avec le CA par client : si les données d'achat sont dispersées sur plusieurs fiches pour un même client, impossible de connaître le revenu que génère la personne, et de l'identifier le cas échéant comme client à fort potentiel.

De même, le taux de transformation des opportunités part à la baisse quand il est calculé sur une base client avec 10 % ou plus de doublons. Le reporting, la visibilité, les chiffres : tout est faussé.

Votre entreprise est-elle concernée ?

Vous avez peut-être déjà pris des mesures pour réduire le nombre de doublons dans votre base client, mais le résultat est-il à la hauteur ? Posez la question à vos opérationnels de terrain : ils sont les mieux placés pour confirmer la présence de doublons qui entravent leur efficacité opérationnelle.

Vous avez besoin avant tout de prendre la mesure du problème dans votre entreprise. Pour connaître la part de doublons dans votre base client, confiez à un spécialiste l'audit d'un échantillon de votre base. Le constat peut être amer, mais il justifie de passer à l'action.



2.

Préparer le terrain

Un projet d'unification des données clients concerne toute l'entreprise. Il est donc nécessaire de mettre tous les services au diapason. En plus des bonnes pratiques de conduite de projet et du changement, il est important de travailler plus spécifiquement sur l'unification des données.

Bonnes pratiques



Accepter de décroisonner

La culture de la donnée dans de nombreuses entreprises repose encore sur une réticence à partager les données en interne. Pour chaque service, sa donnée est son pré carré. Quand plusieurs systèmes cohabitent chacun avec leur base, impossible de dédoubler, de fusionner et de constituer un référentiel client unique – la recherche du Golden Record devient un chemin de croix.

Aboutir au Golden Record exige de dédoubler chaque base existante, puis de les réconcilier dans un référentiel unique au sein duquel les données client validées sont fusionnées.



Définir un sponsor

Le sponsor est un rôle déterminant dans la réussite du projet. Ce rôle peut être confié, selon les entreprises, au directeur marketing, au directeur du service clients, au directeur financier... Ce sponsor doit avoir « les coudées franches » et s'assurer que chaque service « joue le jeu » dans la mise à disposition de ses données. En contrepartie, le sponsor assure que la restitution du Golden Record sera accessible à tous.



Ajuster les attentes dans le projet

En termes de déduplication, la perfection n'existe pas. Autrement dit, viser l'identification de l'intégralité des doublons dans une base n'est pas un objectif réaliste.

En effet, l'identification de doublons est une question de compromis entre la tolérance aux fautes d'une part, et le pourcentage de proximité entre les fiches d'autre part. L'accompagnement d'un expert est ici impératif pour savoir où votre entreprise doit placer le curseur.



3.

Identifier les doublons

Qu'est-ce qui constitue un doublon pour votre entreprise ?

La réflexion doit commencer par cette question primordiale. Il n'existe pas de réponse universelle, car tout dépend du métier de l'entreprise. Il faut avant tout identifier les champs de données les plus pertinents pour effectuer la comparaison. Ces champs vont ensuite servir de clés de rapprochement pour identifier les doublons.

Bonnes pratiques



Définir, au minimum, trois champs d'identification des doublons

Définir au minimum 3 champs d'identification des doublons pour aboutir à une déduplication efficace. Si les champs tels que nom, prénom, adresse email font partie des éléments couramment utilisés, les entreprises peuvent aussi en définir d'autres, tout aussi importants selon leur métier. Numéro de sécurité sociale dans le secteur de la santé, numéro d'immatriculation d'un véhicule dans le secteur automobile, Siret et informations légales d'entreprise dans les bases BtoB... Les champs ainsi retenus doivent avoir une importance prioritaire pour identifier un client unique.

Pour identifier un maximum de doublons, mieux vaut passer la base au crible de chaque clé de rapprochement définie. Autrement dit, il s'agit d'utiliser trois filtres consécutifs pour resserrer les mailles d'identification des doublons.



Introduire la notion de « flou »

Il n'y a aucun intérêt à rechercher des doublons sur des critères de similitude exacte entre fiches. Seuls quelques doublons ressortiront tout au plus. D'où la notion de flou, soit l'exploration des approximations de saisie à cause desquelles les fiches d'un même client sont identifiées comme différentes par les systèmes métier type CRM ou ERP. Prénom sous forme d'initiale puis saisi en toutes lettres à une autre occasion, numéros fixe et mobile inversés entre différentes fiches, coquille dans un nom, adresses de différentes résidences... Le flou prend de nombreuses formes ! Il est nécessaire de l'introduire pour détecter davantage de doublons, ce qui exige également l'accompagnement d'un spécialiste. Pour identifier un maximum de doublons, mieux vaut passer la base au crible de chaque clé de rapprochement définie. Autrement dit, il s'agit d'utiliser trois filtres consécutifs pour resserrer les mailles d'identification des doublons.



4.

Fusionner

Deux questions guident la logique de fusion : que fusionner et quand le faire ?

Que fusionner ?

Parmi les données incluses dans les doublons d'un même client, lesquelles sont prioritaires et vont « écraser » les autres dans la fiche parent unifiée ?

Bonnes pratiques



Garder le doublon le plus ancien

Garder le doublon le plus ancien comme fiche de base, car il est en général le plus complet.

Des autres fiches en double, ne garder que les données les plus récentes, dont les coordonnées email, adresse, téléphone, et les éléments associés tels que les commandes, les appels au SAV ou les rendez-vous, par exemple.



Tenir compte de la provenance des données

Des données saisies en selfcare par le client sur son compte en ligne seront bien plus fiables que celles issues d'un listing marketing.



Écarter les mots « bruit »

Écarter les mots « bruit » qui, à l'inverse des clés de rapprochement, n'offrent aucune utilité pour la fusion. Par exemple, dans une base BtoB d'entreprises de transport, nombreuses seront celles qui incluent les termes « Transports » ou « Logistique » dans leur raison sociale. Le champ de la raison sociale doit donc être écarté, puisqu'il ne peut pas être un critère pertinent pour fusionner des fiches client.



Quand fusionner ?

Faut-il fusionner à la saisie en temps réel, ou en mode batch en traitements réguliers sur la base où des doublons peuvent avoir été créés ?

Bonnes pratiques



Adapter à la configuration du système d'information

Plus une entreprise a décloisonné son SI et interconnecté ses différentes solutions, mieux elle peut opter pour la solution la plus qualitative : la fusion des données client en temps réel à la collecte.

En revanche, dans les entreprises où les différentes solutions sont cloisonnées, le mode batch s'avère plus approprié, à condition de prévoir un interfaçage de la solution de déduplication/unification avec les différents systèmes. Dans ce cas, un traitement régulier de la base permet de la nettoyer de ses doublons selon une fréquence choisie.



Et quand ne pas fusionner !

Certains cas d'usage nécessitent de maintenir plusieurs fiches pour un même client. C'est le cas d'un groupe multimarque dont les clients fréquentent ses différentes enseignes - une fiche doit être maintenue pour chacune. De même, lorsque la clientèle dispose de plusieurs résidences - dans le secteur du Luxe en particulier - chacune est une information importante. Or, le format des fiches client de l'outil CRM n'a peut-être pas prévu assez de champs Adresse pour toutes les renseigner. Là encore, reste à maintenir différentes fiches pour un même client.

L'astuce pour tolérer les doublons sans créer de confusion pour les utilisateurs : le Golden ID, soit un identifiant propre à chaque client et commun à ses différentes fiches.





Oser fusionner !

Certaines entreprises savent qu'elles doivent unifier leurs données client, mais hésitent lors de l'étape de fusion. Leur crainte : risquer de perdre des données client – autrement dit, fusionner en une même fiche deux clients pourtant distincts. Or, la fusion est l'étape ultime à laquelle aboutit le projet d'unification des données client.

Votre entreprise peut compter sur plusieurs filets de sécurité dans l'art de fusionner les données client :



La fusion manuelle

Pour les entreprises qui préfèrent ne pas tout automatiser et laisser la main aux métiers, il est toujours possible de maintenir l'option d'une fusion manuelle. Ainsi, une fois les doublons identifiés, la fusion dépend d'une intervention humaine, ce que certaines entreprises trouvent plus rassurant.



Le contrôle de cohérence

Une fois les doublons identifiés, le contrôle de cohérence contribue à définir quelles données conserver sur la fiche parent, et lesquelles supprimer car erronées.

Un exemple : les éléments précédant l'@ dans une adresse email contrôlée comme valides sont souvent le nom et le prénom du client. Ils offrent un critère de comparaison pertinent pour vérifier l'orthographe exacte du nom et du prénom, quand ces deux informations présentent des orthographe à géométrie variable entre doublons. Le contrôle de cohérence avec l'adresse email permet de confirmer l'orthographe exacte du nom et du prénom à conserver sur la fiche parent.



Le calcul de score

Plus il est haut, plus ce score confirme les cas de doublons. Calculé selon les différents champs utilisés comme clés de rapprochement, le calcul de score implique de définir le seuil auquel une entreprise estime qu'il y a présence de doublons. Cet indicateur exige de bien maîtriser le poids des clés de rapprochement retenues. Par exemple, un score élevé peut cacher un seul écart qui disqualifie d'office la présence de doublons s'il porte sur un champ de données d'importance prioritaire dans votre activité. L'accompagnement de votre expert, une fois encore, est indispensable pour définir les règles appropriées.



Le pourcentage de proximité

Le pourcentage de proximité entre les différentes fiches définit le seuil de similarité pour lequel on décide de fusionner les doublons.



Unifier les données client : jamais sans un expert !

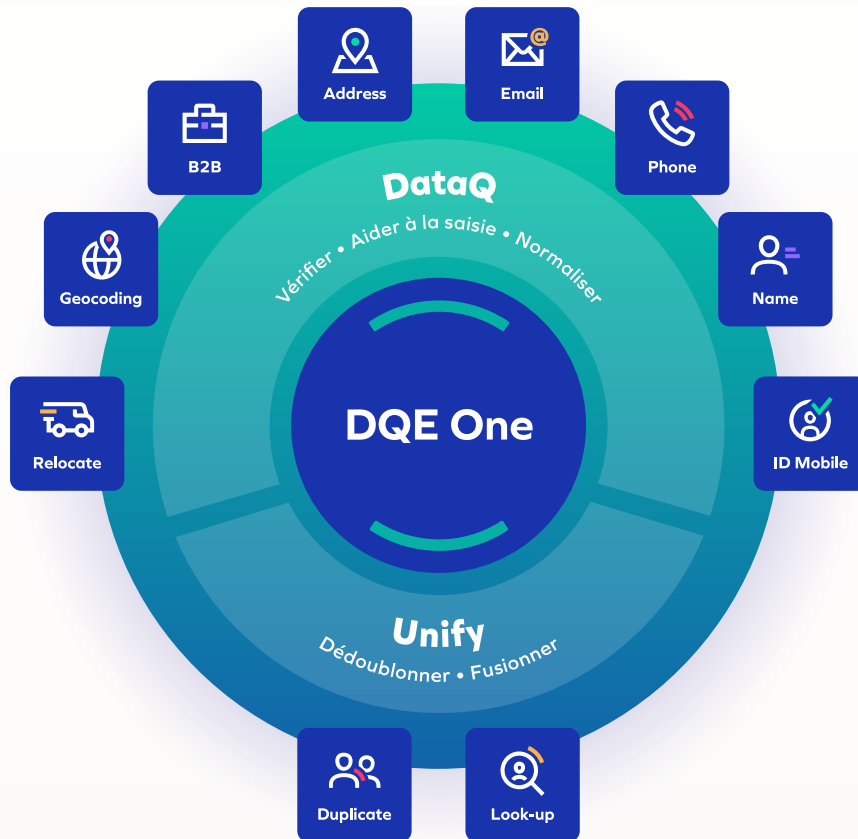
Croire que l'on peut développer seul une solution pour unifier les données client est une chausse-trape à éviter ! Ce sont généralement des projets de longue haleine avec des résultats très incertains voire faux.

Notre expérience auprès de plusieurs entreprises le confirme

Toutes celles qui ont tenté de dédupliquer et fusionner leurs données client par leurs propres moyens y ont perdu des sommes colossales en développement, un temps infini, pour un résultat bien en deçà des exigences en la matière. Il n'est pas rare de constater des temps de traitement de plusieurs semaines avec des solutions maison, quand quelques heures suffisent avec la solution d'un spécialiste. De plus, l'identification des doublons laisse à désirer avec une solution trop générique, ou qui n'utilise pas de clés de rapprochement assez fines pour trouver un maximum de doublons dans une base. Enfin, reste le cas des entreprises qui arrêtent la démarche au moment de fusionner par manque de maîtrise et crainte de perdre des données client.

Pour mener votre projet d'unification des données client à son terme et à la réussite, faites-vous accompagner par un expert tel que DQE

DQE déploie dans son offre de Data Quality 15 ans d'expérience, des technologies éprouvées et sans cesse optimisées, avec en prime, la connaissance fine des bonnes pratiques, des points d'attention, des écueils à éviter. Vous avez ainsi l'assurance d'unifier vos données en toute connaissance, au bénéfice de votre efficacité et ultimement, vos résultats.



À propos de DQE

Créée en 2008 en France, DQE Software (DQE) est l'éditeur de DQE One, solution modulaire de Data Quality Management.

Nos solutions sont spécialisées dans la donnée client de contactabilité et d'identité BtoC et BtoB. DQE One contrôle la qualité des données sur les informations de type nom, prénom, e-mail, téléphone, adresse et données légales. DQE One regroupe deux gammes de produits : DataQ, solution de mise en qualité des données contacts, et Unify, solution de dédoublonnage.

Dans tous les secteurs, DQE aide les organisations à garantir que leurs données client sont fiables, accessibles à tous, à chaque instant. DQE One se distingue comme la solution de Data Quality la plus performante du marché, capable de gérer des bases de plusieurs dizaines de millions de contacts. Chaque année, nos traitements représentent plus de 3 milliards de requêtes dont le temps moyen de réponse est de 150 millisecondes.

DQE One s'appuie sur des moteurs et algorithmes puissants qui bénéficient de plus de 10 années de retour d'expérience. La solution s'interface avec plus de 240 bases internationales pour la partie adresses postales. Nativement, DQE One propose des connecteurs avec les solutions telles que Salesforce, Microsoft/Dynamics, Cegid, Magento, etc. Nos algorithmes fonctionnent sur l'ensemble des clouds de nos clients : Heroku, Azure, AWS ou GCP. DQE One est utilisée par de nombreux métiers (équipes CRM, Chief Data Officers, Chief Digital Officers, DSI, responsables de la relation clients), chez plus de 400 clients. Parmi eux figurent plusieurs sociétés du CAC 40 et des marques de renom telles que Groupama, EDF, BUT ou Belambra.

Pour de plus amples informations, rendez-vous sur www.dqe.tech et suivez-nous sur 