

L3 MIAGE — TD n°1 bis – Qualité des données (avec correction)

Ce document reprend le sujet précédent et propose des éléments de correction. Les réponses sont données à titre indicatif et peuvent être complétées en fonction du support de cours.

Exercice 1 – Repérer les problèmes de non-qualité

Analyse des anomalies

- **Doublons et orthographes :**
 - Les abonnés AB1, AB3 et AB6 correspondent probablement à la même personne (Marie Dupont) mais l'orthographe du nom varie : « Dupont », « Dupond » et « Dupon » et la date de naissance est saisie une fois avec un zéro devant le mois et une fois sans.
 - Les lignes AB1 et AB6 comportent les mêmes informations mais le code postal ou la commune manque, ce qui empêche de vérifier qu'il s'agit bien d'une seule personne.
 - Il en résulte un risque de comptabiliser trois abonnements au lieu d'un.
- **Noms en majuscules/minuscules :**
 - AB2 (Martin Jean) et AB5 (MARTIN Jeanne) pourraient désigner deux personnes différentes (frère et sœur, par ex.), mais l'écriture du nom en majuscules et l'adresse identique laissent penser qu'il s'agit d'une duplication ou d'une **saisie erronée du prénom**.
 - Sans date de naissance pour AB5 il est impossible de lever l'ambiguïté.
- **Champs manquants :**
 - AB3 ne mentionne pas la commune ; AB5 n'a pas de date de naissance ; AB6 n'a pas de code postal ; certains emprunts (AB6) ne correspondent à aucun abonnement (orthographe « Dupon »).
 - Ces **incomplétudes** empêchent d'appliquer correctement les tarifs (tarif réduit pour les résidents de la commune, tarifs jeunes/seniors) et de contacter les lecteurs.
- **Erreurs de saisie dans les emprunts :**
 - la table *Emprunts* contient un identifiant AB8 (Durand Julien) qui n'existe pas dans la table *Abonnement*, ce qui traduit un **défaut d'intégrité référentielle**.
 - Le livre emprunté par AB5 ne précise pas le nom complet (« À la recherche du temps... ») et AB6 fait référence à « Dupon » sans abonnement.

Conséquences lors de l'interrogation

- La présence de doublons et d'informations manquantes fausse les statistiques (nombre d'abonnés, nombre d'emprunts par personne), entraîne des erreurs de facturation (plusieurs factures pour la même personne), rend difficile la recherche d'un abonné ou l'envoi de messages personnalisés (courriers qui ne parviennent pas, courriels envoyés au mauvais destinataire) et complique le calcul des indicateurs (par exemple le nombre de lecteurs grenoblois).
- De manière générale, **une mauvaise qualité des données se répercute à toutes les étapes de traitement** : si des données erronées entrent dans le système, les résultats obtenus seront eux-mêmes erronés (« garbage in, garbage out »).

Exercice 2 – Circulation de l'information

Q1 : identification des flux

Dans la plateforme e-commerce, les données de commandes et de stocks circulent en continu :

- **Flux ascendants** : les informations saisies lors d'une commande (article, quantité, client, paiement) alimentent la base de données puis sont transmises en temps réel au **service logistique** afin de préparer les colis. Chaque entrée/sortie de stock est remontée vers la **base centrale**. Ces données sont agrégées mensuellement pour les services **achats, finance et marketing** et consolidées semestriellement pour la **direction**.
- **Flux descendants** : la direction définit des objectifs (réduction des stocks, lancement de nouveaux produits, campagnes promotionnelles) et transmet ces directives vers le service achats (planning de réapprovisionnement), le marketing (campagnes ciblées) et le service logistique (niveaux de stock à respecter). Des retours sont ensuite envoyés aux équipes de terrain.

Ainsi, on observe plusieurs niveaux d'échanges :

- Un premier niveau opérationnel entre la plate-forme et la logistique,
- Un deuxième niveau décisionnel entre les services fonctionnels (achats, finance, marketing) et, enfin
- Un niveau stratégique entre ces services et la direction.

Le tableau ci-dessous illustre la circulation de l'information entre les différents niveaux du système d'information (opérationnel, décisionnel et stratégique).

Pour chaque niveau, on précise les fonctions impliquées et les principaux flux entrants et sortants.

Niveau du système / fonctions impliquées	Flux entrants (données reçues)	Flux sortants (données transmises)
Système stratégique : direction	<ul style="list-style-type: none"> • Tableaux de bord consolidés (indicateurs semestriels de ventes, stocks, performances clients) 	<ul style="list-style-type: none"> • Orientations stratégiques (politique de stocks, lancement de nouveaux produits, budgets) transmises aux services décisionnels et opérationnels
Système décisionnel : Achats, finance, marketing	<ul style="list-style-type: none"> • Synthèses mensuelles des ventes et des stocks transmises par le système opérationnel 	<ul style="list-style-type: none"> • Ordres de réapprovisionnement (quantité, calendrier) ; • Factures et budgets ; • Campagnes marketing ciblées
Système opérationnel : logistique (préparation des commandes, livraison)	<ul style="list-style-type: none"> • Commandes des clients (articles, quantités, paiement) ; • Mouvements de stock (entrées/sorties) 	<ul style="list-style-type: none"> • Statut d'avancement des commandes ; • Mise à jour du niveau de stock dans la base centrale

Ce schéma simplifié met en évidence la hiérarchie des flux :

Les données brutes remontent du terrain vers les systèmes décisionnels puis stratégique, qui les transforment en instructions descendantes.

Une bonne qualité et circulation de l'information à chaque niveau est indispensable pour la prise de décision.

Q2 : outils du système d'information

Pour assurer ces flux, plusieurs outils peuvent être mobilisés :

- **SGBD / entrepôt de données** : pour stocker et structurer les données de commandes, de clients et de stocks.
- **ERP (Enterprise Resource Planning)** : pour intégrer les modules achats, logistique et comptabilité et synchroniser les informations en temps réel.
- **CRM (Customer Relationship Management)** : pour gérer les informations clients et personnaliser les actions marketing.
- **ETL (Extract–Transform–Load) et outils de Business Intelligence** : pour extraire les données, les transformer et produire des tableaux de bord à destination de la direction.
- **APIs et plateformes de cloud** : pour échanger des données avec des prestataires externes (transporteurs, fournisseurs) et garantir la disponibilité.

Q3 : risques et impacts

Le tableau suivant propose des exemples de risques et d'impacts associés à chaque phase du cycle de vie des données. Les réponses sont succinctes afin de faciliter la lecture ; dans un devoir, elles peuvent être détaillées.

Phase / technique	Risques (exemples)	Impact (exemples)
Acquisition des données		
– Saisie manuelle par un opérateur	Typos, oublis, doublons	Données erronées, facturation fausse
– Lecture automatisée (code-barres, API)	Code illisible, panne de scanner	Rupture de stock non détectée, blocage des flux
Stockage des données		
– Base de données interne (serveurs locaux)	Panne matérielle, incendie, sauvegarde incomplète	Perte de données, arrêt de l'activité
– Base de données externalisée (cloud)	Fuite de données, indisponibilité du fournisseur	Atteinte à la confidentialité, dépendance forte
Traitement des données		
– Logiciels internes développés en interne	Mauvaise conception, absence de tests	Résultats incohérents, décisions inappropriées
– Logiciels ou services externes (SaaS)	Manque de maîtrise, paramétrage opaque	Non-traçabilité, perte de compétences internes
Diffusion des données		
– Réseau interne (intranet)	Accès non autorisé, version obsolète	Prises de décision erronées, fuite d'informations
– Partage par courriel ou export de fichiers	Envoi au mauvais destinataire, perte de fichier	Violation de la loi RGPD, réputation entachée
– Accès via API externe	Piratage, indisponibilité de l'API	Interruption du service, fuite de données sensibles