

Fiche d'introduction

Le groupe Qualité des données (QDD) travaille sur la thématique de la qualité des données de surveillance commune aux 3 plateformes.

Les membres du groupe de travail recensent les grands principes et bonnes pratiques applicables au sein des dispositifs de surveillance, tous secteurs confondus, pour optimiser la qualité des données.



Ils sont mis à disposition des professionnels impliqués dans les travaux des 3 plateformes et de l'ensemble des acteurs de la surveillance via la publication d'un guide pratique.



Le groupe QDD a publié un Guide pratique sur la qualité des données de surveillance. Dans ce guide, un panel d'exemples, basés sur des données de surveillance en santé animale, végétale ou surveillance de la chaîne alimentaire, permet d'illustrer concrètement certains facteurs modifiables, susceptibles de compromettre la qualité des données. Des indicateurs d'évaluation de la qualité des données sont par ailleurs proposés.



[Guide qualité des données](#)

Afin de faciliter l'appropriation du guide, des fiches synthétiques ont été réalisées. Il existe 4 fiches synthétiques :

- 1 La qualité des données à l'échelle d'un dispositif de surveillance
- 2 Les bonnes pratiques à chaque étape de la vie des données
- 3 Evaluation de la qualité des données
- 4 L'animation de la qualité des données au sein d'un dispositif de surveillance

DÉFINITIONS :

DONNÉE : Valeur affectée à une variable (un champ), représentation formalisée de l'information, adaptée à l'interprétation, destinée à faciliter le traitement et la communication. La donnée peut être représentée comme un conteneur porteur d'une information ou d'un fragment d'information. Une donnée est une description élémentaire, représentée sous forme codée, d'une réalité (chose, évènement, mesure, transaction, etc.) en vue d'être collectée, enregistrée, traitée, manipulée, transformée, conservée, archivée, échangée, diffusée et communiquée.

VARIABLE : Série de données au format homogène, collectée sur un ensemble d'unités statistiques, renseignant le même attribut (âge, nom, plan, méthode). Une variable peut prendre différentes valeurs, appelées modalités lorsque la variable est qualitative.



Fiche 1

La qualité des données à l'échelle d'un dispositif de surveillance

Durée du dispositif de surveillance



Concerter les parties prenantes de l'ensemble du dispositif

Solliciter une vision extérieure (bonus)



Définir les **objectifs** et le **champ** du projet

Définir les **valorisations** attendues

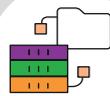
Identifier les **contraintes** et problématiques



Identifier les **données utiles*** à collecter et leur volume

Anticiper le type d'**analyse** de données

Proposer un **outil** de collecte adapté et performant



Plan de gestion de données* intégrant un **plan d'analyse de données**

****donnée utile** = donnée qui apporte une information pertinente/objective, est exploitée et potentiellement diffusée.*

****plan de gestion de données** = document formalisé évolutif qui aide à organiser et anticiper toutes les étapes du cycle de vie de la donnée (collecte, analyse, diffusion).*



Remise en question de la pertinence des données utiles/ ré-adaptation en fonction des objectifs du dispositif



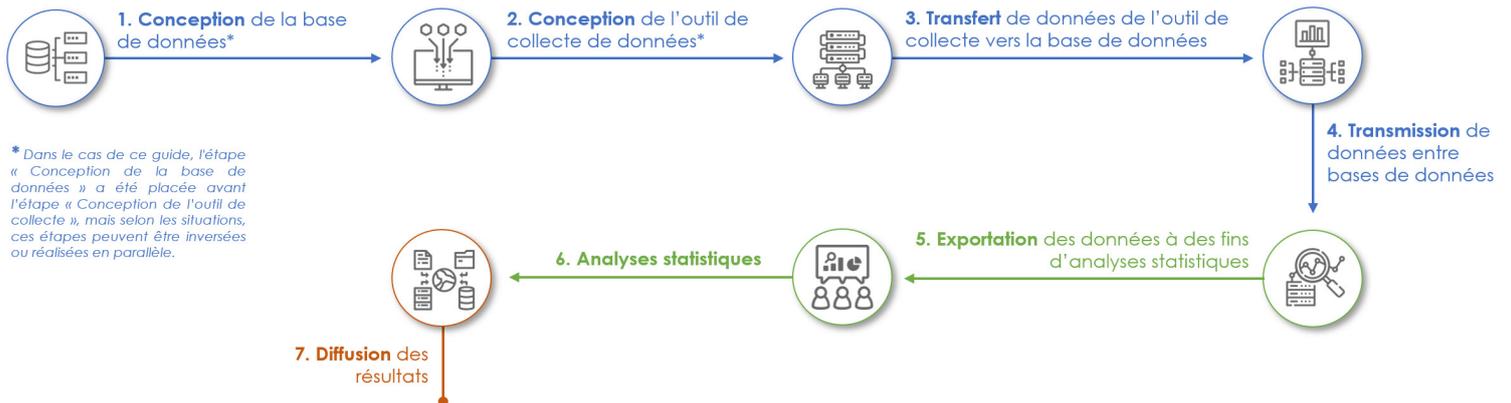
Pour en savoir plus :
[Guide - partie II.A](#)
[Guide complet](#)



Fiche 2

Les bonnes pratiques à chaque étape de la vie des données

Collecte → Analyse → Diffusion



1. Conception et évolution de la base de données

La structure de la base de données :

- Aucun doublon d'enregistrement
- Une identification unique de chaque enregistrement
- Des champs à contraintes obligatoires (format, champ à compléter)
- Une liste de vocabulaire contrôlé (via une liste de choix proposés, de préférence au singulier) se référant de préférence à des standards préétablis

La gestion de la base de données :

- Enregistrer toute modification de la base dans un journal pour assurer la traçabilité
- Réaliser une sauvegarde automatique des données voire des versions des logiciels et/ou packages utilisés
- Réaliser une réplication de la base de données
- Réaliser un archivage

- Homogénéiser les pratiques de nommage et le classement des fichiers en signalant chaque modification par une date ou un versionning
- Formaliser les décisions dans un plan de gestion de données ou un plan d'analyse de données
- Anonymiser les données avec possibilité de ré-identification ultérieure le cas échéant



Fiche 2 Les bonnes pratiques à chaque étape de la vie des données

Indicateurs pour décrire une base de données :

Proportions de variables

- Qualitatives
- Quantitatives
- Standardisées



Se référer à la fiche 3

Indicateurs de remise en question de la base de données et de ses objectifs :

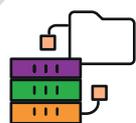


Nombre de réunions de concertation autour de la base de données



Nombre de mises à jour et validation de documents de formalisation

2. Conception de l'outil de collecte des données



- Définir une liste de variables à collecter et leurs contraintes associées (format, type...)
- Tester l'outil de collecte
- Prévoir une procédure de collecte
- Envisager une formation et une assistance

2

3. Renseignement de la donnée

- Éviter les doubles saisies
- Privilégier les champs fermés avec un vocabulaire commun dès que possible
- Identifier les champs obligatoires avec un système de blocage et message d'erreur en cas de non complétion
- Prévoir des modalités d'autocontrôle de la cohérence au moment de la saisie des données dans l'outil de collecte si possible via un message d'alerte



4. Échanges entre les outils de collecte et les bases de données



- Sauvegarder les données brutes
- Conserver la traçabilité des informations issues d'autres bases de données (provenance, dates...)
- Sécuriser les échanges



Pour en savoir plus :
[Guide - partie II](#)
[Guide complet](#)



Fiche 3

Évaluation de la qualité des données

L'évaluation de la qualité des données est indispensable avant toute exploitation de données (analyses statistiques ou autres valorisations). Elle s'effectue par le biais d'indicateurs et peut être définie par différents objectifs :

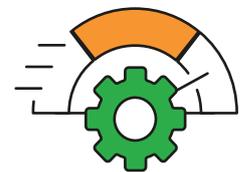
Adapter les analyses de données selon le niveau de qualité des données obtenu

Améliorer la qualité de ces données (ex : par des actions correctrices)

Contrôler le processus de gestion des données et mettre en évidence d'éventuels biais (ex : outil de collecte non adapté)



Les indicateurs de qualité des données



Les principaux indicateurs :

A l'échelle d'une variable :

Le taux de complétude = $\frac{\text{Nb données renseignées}}{\text{Nb données attendues}}$

Le taux de validité = $\frac{\text{Nb données valides (correspondant à la réalité)}}{\text{Nb données renseignées}}$

Ex de données non valides : 35/05/2021

Le taux de données au bon format = $\frac{\text{Nb données respectant le bon format}}{\text{Nb données renseignées}}$

Ex de format : JJ/MM/AAAA pour une date

La précision = $\frac{\text{Nb données respectant la précision souhaitée}}{\text{Nb données renseignées}}$

Ex de précision au sein de la nomenclature végétale : genre, espèce, sous-espèce, variété

A l'échelle de plusieurs variables :

Le taux de cohérence = $\frac{\text{Nb données cohérentes entre les variables testées}}{\text{Nb données renseignées}}$

Ex cohérence entre 2 dates : date de prélèvement d'un échantillon (d1) et date d'analyse de cet échantillon (d2) : $d1 \leq d2$



Fiche 3 Évaluation de la qualité des données

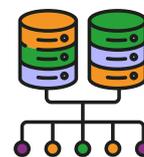
Les indicateurs synthétiques : pour une évaluation plus globale des données :

À l'échelle d'une dimension de la qualité de données

Macro-indicateur

Nb données respectant un critère de qualité donné (ex : complétude)

Nb données renseignées et concernées



Super-indicateur

Nb données respectant l'ensemble des critères de qualité fixés

Nb données renseignées

Les indicateurs secondaires : pour compléter l'évaluation :

La pertinence d'une variable

Utilité de la variable par rapport aux besoins des utilisateurs et des objectifs du dispositif de surveillance

La crédibilité

Perception de l'exactitude des données par l'utilisateur

L'accessibilité

Facilité d'accès et d'utilisation des données

L'objectivité

Niveau d'impartialité lors du renseignement des données

La confidentialité

Accès aux données brutes et/ou valorisées seulement aux personnes autorisées

2



Pour en savoir plus :
[Guide - partie II.E](#)
[Guide complet](#)



Fiche 3 Évaluation de la qualité des données

Les actions pour améliorer la qualité des données



Préventives : Permettent d'éviter l'apparition d'un défaut

Ex : communication auprès des collecteurs et utilisateurs de données suite à la mise à jour de modalités (apparition / suppression / renommage)

Curatives : Éliminent un défaut

Ex: correction d'une donnée (faute d'orthographe)

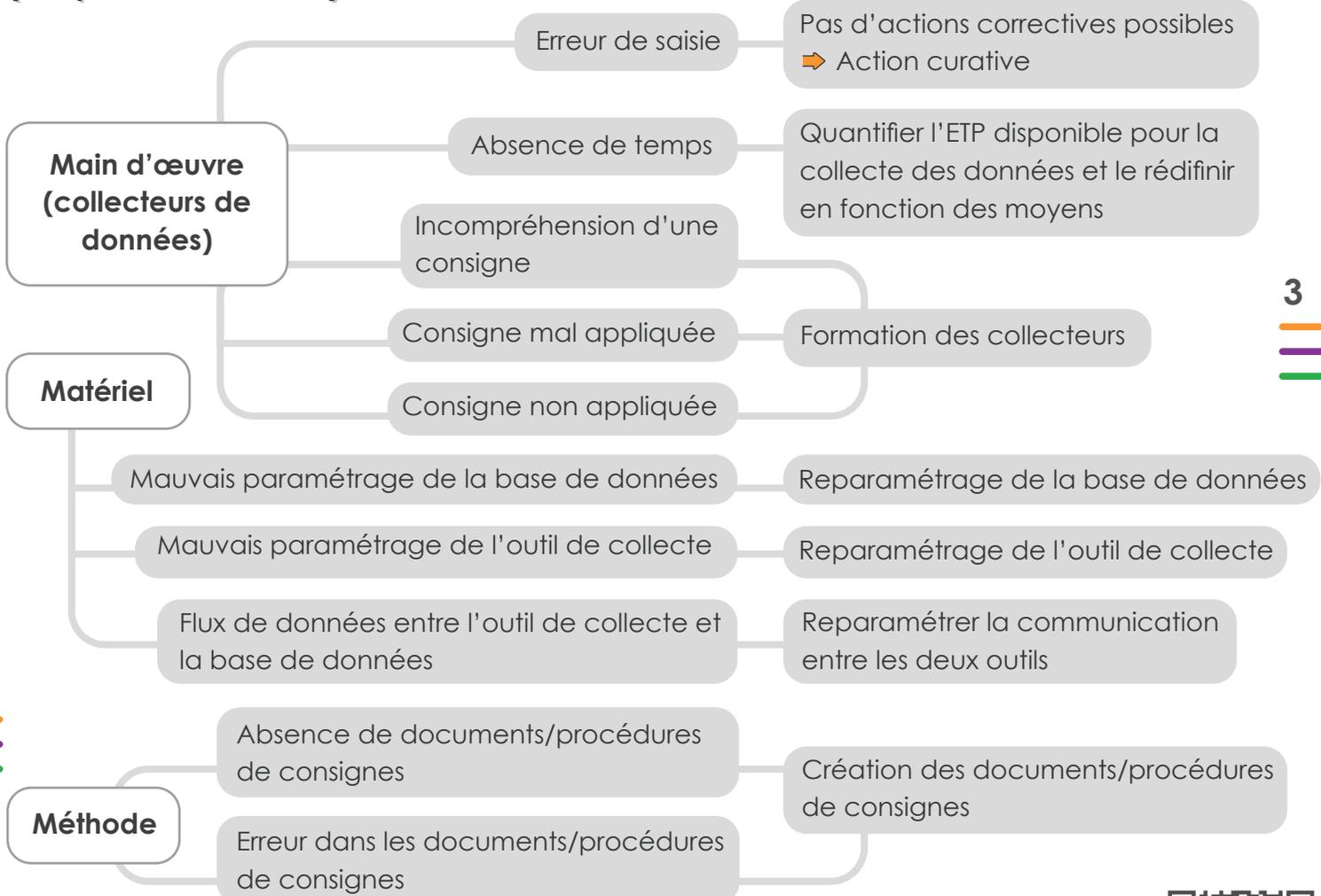


Correctives : Ciblent les causes d'apparition du défaut. Des outils comme la **Méthode des 5M** peuvent aider à la mise en place de ces actions

Diagramme des 5M (adaptée aux données)

Causes

Actions correctives possibles



Fiche 4

L'animation de la qualité des données au sein d'un dispositif de surveillance

L'animation de la **Qualité Des Données (QDD)** a pour objectif de tendre ou de maintenir un niveau de QDD satisfaisant dans le temps. Les indicateurs de qualité des données permettent d'évaluer la QDD, l'animation permet d'améliorer cette qualité en fonction des objectifs fixés. Elle est donc impérative, continue et implique l'ensemble des acteurs du dispositif de surveillance.



1. Implication des partenaires et définition des objectifs :

- Réunir les acteurs régulièrement
- Formaliser, documenter et officialiser

1

2. Formation et maintien des compétences :

- Travailler ses supports de formation et les rendre accessibles
- Former régulièrement
- Adapter les formations à destination et besoins des acteurs

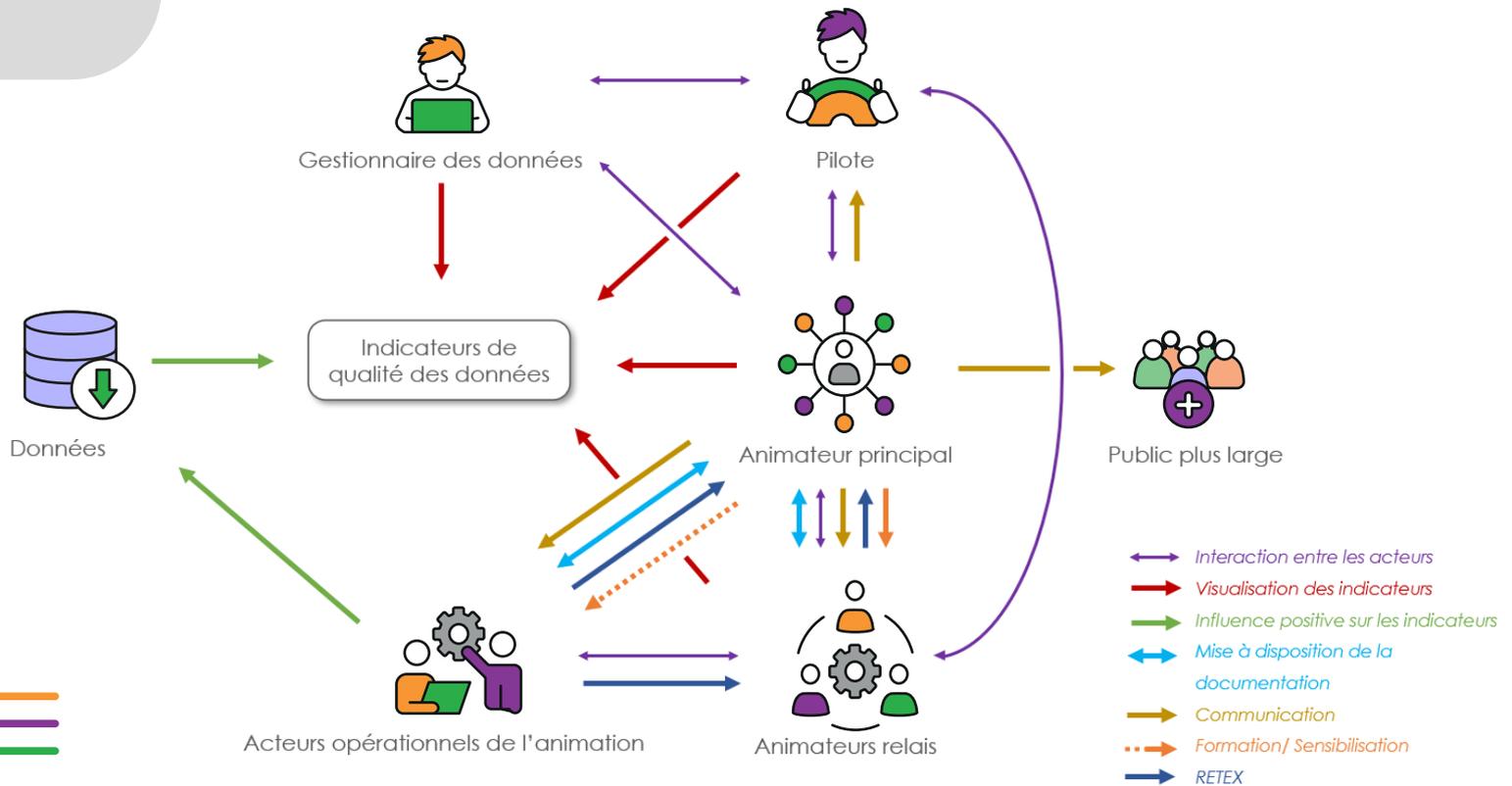


3. Communication et valorisation :

- Concevoir avec rigueur et intelligence la visualisation des indicateurs de qualité
- Profiter des réunions annuelles du dispositif pour sensibiliser les acteurs
- Publier un bilan annuel aux acteurs et communiquer plus largement
- Organiser des séances de retour d'expérience (RETEX) avec les acteurs du dispositif



Fiche 4 L'animation de la qualité des données au sein d'un dispositif de surveillance



Possibilité d'une auto-évaluation régulière de l'animation de la qualité des données

L'animation de la QDD permet aussi au sein du dispositif de :

