Pays Basque
Euskal Herria

COMMUNAUTÉ D'AGGLOMÉRATION
HIRIGUNE ELKARGOA
COMUNAUTAT D'AGLOMERACION

Cadre global de la gouvernance des données

Rédacteur du document	Date/Description	Direction émettrice	Version
BREVET Etienne Responsable du pilotage stratégique des données	12/06/2022 – Initialisation	DSI	1.0

Table des matières_Toc136335876

1. But et objectifs 2
2. Périmètre et applicabilité 2
3. Documents de référence et glossaire 3
   3.1 Documents de référence 3
   3.2 Glossaire 3
4. Cycle de vie et intégrité de la donnée 7
   4.1 Cycle de vie de la donnée 7
   4.2 Métadonnées 7
   4.3 Intégrité de la donnée 7
5. Pratiques obligatoires pour la gestion des données 8
   5.1 Organisation de la Gestion des Données 8
   5.2 Processus et Organisation 9
   5.3 Bonnes pratiques 9
6. Evaluation et Suivi de la Qualité des Données 10
   6.1 Cible de qualité des données 10
   6.2 Axes de contrôle de la qualité 10
7. Responsabilités et comitologie 10
   7.1 Comité de Pilotage 11
   7.2 Comité technique 11
8. Evolution du document 11

1. But et objectifs

Afin de répondre aux besoins des usagers, élus et agents de la Communauté d'Agglomération Pays Basque ainsi qu'aux normes et règles imposées par le législateur, **le présent document vise à donner un cadre global à la gestion des données.**

Le premier objectif de ce document est de doter la Communauté d'Agglomération Pays Basque de documents de référence et d'un glossaire partagé sur les sujets liés à la gestion des données.

Le second objectif, recherché à travers la « mise sous contrôle » des données, est d'améliorer la confiance dans les données et finalement de fiabiliser la prise de décision et l'efficacité opérationnelle :

- des données de bonne qualité et disponibles au bon moment permettent de construire des prévisions et des hypothèses plus solides pour bâtir nos politiques publiques ;
- le partage d'un langage commun, de jeux de données validés et accessibles au plus grand nombre, simplifie la création de contenus de qualité, la transversalité entre les différentes parties prenantes de la CAPB et permettent la mise en place de nouveaux services à forte valeur ajoutée.

Avant tout, cette gestion des données s'applique à travers la mise en place d'un langage commun, de procédures formalisées, d'une gouvernance partagée entre toutes les composantes de l'organisation et plus largement avec l'ensemble des acteurs concernés par le cycle de vie de la donnée, de sa production à sa conservation en passant par sa consommation.

Cette gouvernance sera, par la suite, appuyée par des outils, des contrats d'interfaçage et d'utilisation, etc., qui seront référencés au fur et à mesure dans le présent document actualisé.

2. Périmètre et applicabilité

Ce cadre global s'applique à l'ensemble de la Communauté d'Agglomération Pays Basque : toutes ses composantes organisationnelles et tous les agents communautaires mais aussi les partenaires et prestataires qui accompagnent la Communauté d'Agglomération Pays Basque dans ses transformations et ses évolutions. En tant qu'acteurs clés, les communes membres sont des partenaires majeurs, appelés à y adhérer.

Ce cadre s'applique à toutes les données produites, transformées, consommées par les systèmes, les agents, les partenaires… en commençant par les données transversales, utilisées par plusieurs acteurs. Le périmètre évoluera avec le temps pour couvrir l'ensemble des données produites par la Communauté d'Agglomération Pays Basque.

Le cadre global synthétise les éléments structurels de la gouvernance globale. Tous ces éléments devront être développés plus avant dans chaque Direction ou service « métier » en fonction des besoins et des opportunités. Dans certains secteurs (Eau, information territoriale…), des normes plus précises et contraignantes peuvent exister, ce sont alors ces normes externes qui seront « reprises » et trouveront à s'appliquer.

Ce cadre global est applicable dès son adoption par la Communauté d'Agglomération Pays Basque et avec ses mises à jour successives. Il sera de ce fait opposable à l'ensemble des acteurs externes.

3. Documents de référence et glossaire

3.1 Documents de référence

Les documents de référence sont disponibles dans le répertoire Documents validés (DSI/Gouvernance des Données). Cette liste est vouée à évoluer. A date, elle contient les documents suivants :

Titre	Direction émettrice
Carte de transformation des Usages Numériques	RSS - DSI
Charte d'utilisation des moyens informatiques et des outils numériques	RSS - DSI
Projet de digitalisation des données sur l'eau	ELMN - QFS
Guide de saisie des éléments de métadonnées INSPIRE	STAH - SIG via CNIG

3.2 Glossaire

***Avertissement** : si certaines définitions sont cadrées par des textes officiels ou largement partagées dans la littérature sur la gestion des données (ex : les données personnelles sont cadrées par le RGPD), d'autres peuvent avoir des définitions qui varient d'un organisme ou d'un auteur à l'autre. Le but de ce présent glossaire est de fournir à la Communauté d'Agglomération Pays Basque son propre vocabulaire pour ôter toute ambiguïté lors d'échanges en son sein ou avec des tiers (même si des nuances pourraient apparaître avec des partenaires ou des sources extérieures).*

Donnée	"Ce qui est connu ou admis comme tel, sur lequel on peut fonder un raisonnement, qui sert de point de départ pour une recherche"[1]. C'est une mesure, un nom, un lieu, un prix… des valeurs objectives (numériques ou non) de la réalité observée, produites par chacun et chaque entité (CAPB, usagers, entreprises…) *Exemple : la mesure du trafic (nombre de vélo à un point précis et en un temps défini)*
Information	Ensemble de données contextualisées qui permet de donner du sens. L'information est basée sur un raisonnement et est donc subjective. *Exemple : transformer la donnée « trafic » relevée ci-dessus en indicateur vert/jaune/rouge d'adoption des modes doux. Il s'agit d'une conceptualisation humaine et chacun pourrait déterminer des seuils différents pour chaque niveau*
Référentiel de	Ensemble des données se rapportant à un sujet (Tiers,

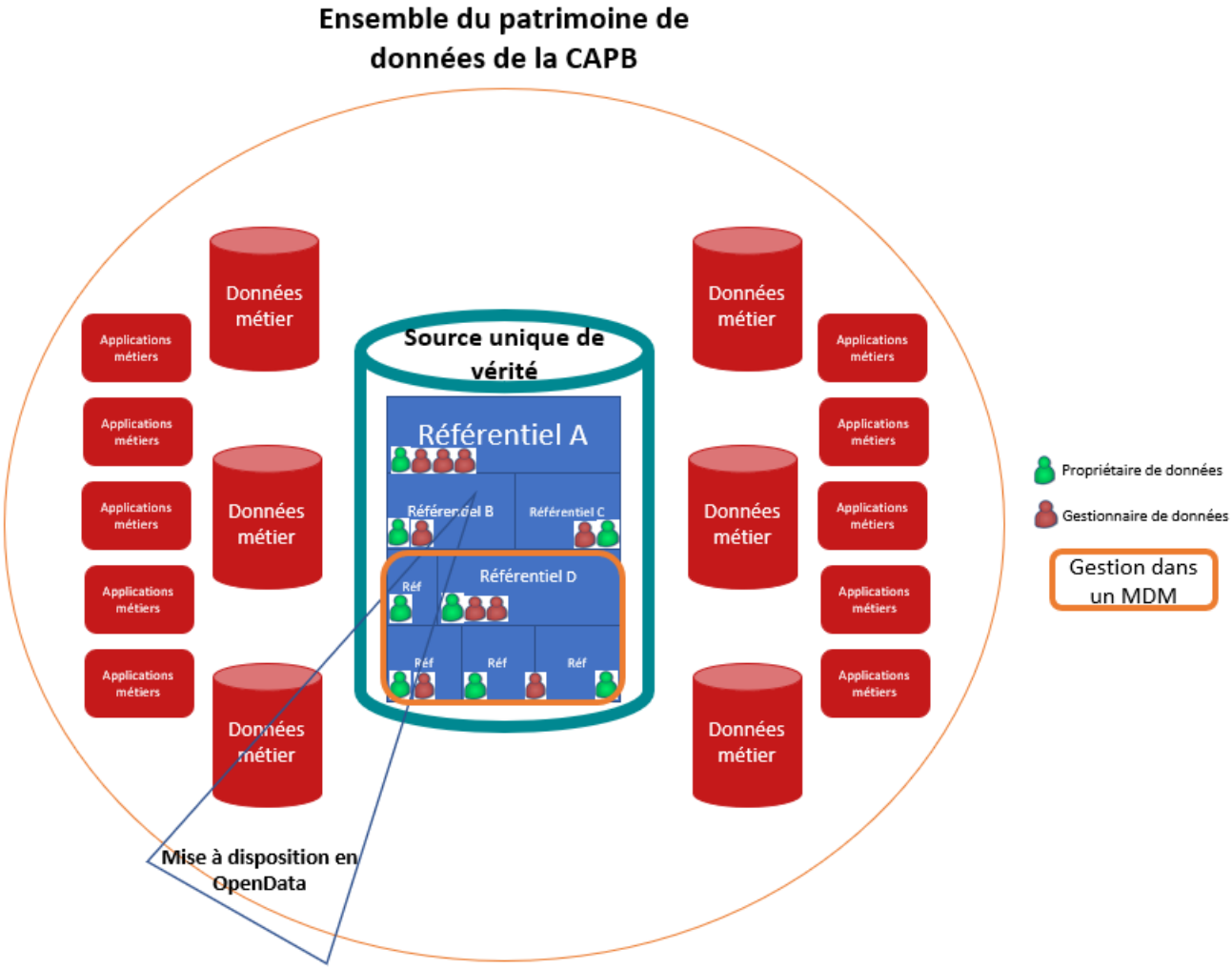
[1] https://www.larousse.fr/dictionnaires/francais/donn%C3%A9e/26436

données	Patrimoine bâti et non bâti…), organisées et structurées en attributs selon un modèle de données pour répondre aux besoins fonctionnels. Les processus de création, de validation, d'utilisation et de diffusion de la donnée sont clairs et maîtrisés. Il ne peut y avoir qu'un référentiel pour un sujet considéré, la référence. *Exemple :* *- la Liste du Patrimoine bâti et non bâti segmentée attributs tels que « Adresse », « Coordonnées GPS », « Superficie », « Héberge du personnel Oui/Non » …*
Propriétaire de données	Au sein de sa direction, le propriétaires de données a pour mission de gérer un périmètre de données clairement défini. Il est responsable de l'atteinte des objectifs de qualité des données jusqu'au niveau déterminé, pour supporter la criticité des activités métier. Il rédige les procédures de gestion des données liées à son périmètre et s'assure de leur bonne application. *Exemple : le Propriétaire des données liées aux emplacements et horaires de transports en commun a pour mission de s'assurer de la bonne application de la procédure de création d'un nouvel arrêt ou de l'évolution des horaires dans les systèmes et de faire corriger les déviations éventuelles. Il fait évoluer les standards et procédures le cas échéant. Il communique également les données mises à jour vers l'extérieur.*
Gestionnaire de données / *Data Steward*	Sous la responsabilité du propriétaire de données, le gestionnaire de données est responsable des activités opérationnelles autour de la donnée et veille au respect des procédures et nomenclatures définies. Il saisit dans les systèmes, contrôle leur cohérence lors de remontées par des capteurs et s'assure de leur disponibilité dans les systèmes. Un gestionnaire de données a un rôle fonctionnel qui peut être mutualisé entre plusieurs services ou directions. *Exemple : les Gestionnaire des données liées aux emplacements et horaires de transports en commun ont la charge de collecter et de renseigner dans les systèmes l'ensemble des données définies dans la procédure de création d'un nouvel arrêt ou d'évolution des horaires, selon les standards de temps et de qualité définis.*
Données référentielles	Données partagées par tout ou partie des processus qui soutiennent l'activité courante de la Communauté d'Agglomération Pays Basque. Ce sont des données qui évoluent très peu avec le temps. *Exemple : Les coordonnées géographique d'une crèche sont uniques et n'évolueront pas avec le temps. Son adresse pourrait évoluer en cas de changement du nom de la rue, chose qui n'arrive pas fréquemment . Ces deux données sont des données dites référentielles.*
Données transactionnelles	Données générées par diverses applications lors d'une « transaction », un évènement. Il peut s'agir d'une transaction financière, mais pas uniquement : validation d'un formulaire, incidents/tickets, commande, relevé de mesure… Ces données sont caractérisées par :

	- une temporalité, la transaction a lieu à un moment précis (relevé de capteur, validation du formulaire…) - une valeur numérique : un prix, un volume, un nombre… *Exemple : la mesure du nombre de vélo qui passent en un point précis évolue en permanence. Ce nombre est différent d'un heure ou d'un jour à l'autre. Il s'agit d'une donnée transactionnelle.*
Donnée à caractère personnel	Toute information relative à une personne physique susceptible d'être identifiée, directement ou indirectement. *Exemple : le « Nom », « Numéro de matricule » ou même le numéro de téléphone d'un agent peuvent permettre d'identifier un agent et sont donc des données à caractère personnel.* *L'horaire de son bus, si rien ne permet de les lier, n'en est pas une.*
Donnée à caractère légal (souveraine, opposable…)	Une donnée est souveraine si sa disponibilité conditionne la possibilité même de la décision publique, c'est-à-dire qu'elle présente pour la puissance publique une véritable criticité. Elle permet d'établir une décision publique opposable, liée à l'État de droit. *Exemple : les règlements d'urbanisme et environnementaux, les servitudes… liés aux données géographiques sont des données dont un citoyen peut se saisir pour faire valoir ses droits.*
Base de données ouverte / *Open Data*	Mise à disposition de données dont l'accès et l'usage sont laissés libres. Ces données sont produites par des organisations (généralement publiques comme la CAPB) et diffusées sous licence ouverte via une plateforme d'échange. Les données, structurées et accompagnées de métadonnées (définitions, date de collecte…), doivent être anonymisées le cas échéant. L'objectif de la licence ouverte est de favoriser leur usage par le plus grand nombre, sans contrainte d'emploi, et ce dans un souci de transparence. *Exemple : - Zabal est la plateforme Open Data de la CAPB pour les données agricoles. Chaque citoyen peut consulter la base de données et réutiliser les données sans frais.* *- Au niveau national, Vite ma dose est un exemple de service proposé grâce à la mise à disposition en Open Data des données du Ministère de la Santé.* On peut aussi parler d'*Open Data interne* quand les données sont stockées sur une plateforme interne et facilement accessibles pour les différents services de l'organisation. *Exemple : Le GéoSocle de GéoBasque met à disposition, en interne, un certain nombre de jeux de données accessible au plus grand nombre.*
Source unique de vérité (objectif)	Système rassemblant et stockant l'ensemble des jeux de données transversaux validés de la CAPB en un seul et même lieu, afin de favoriser leur partage. Cette source permet de donner les accès aux seules personnes et outils autorisés et d'utiliser les données les plus fiables (validées et à jour). Elle a pour objectif de mettre fin à la duplication inutile des données, à leurs possibles incohérences et de fait d'arbitrer les conflits sur « quelle est la bonne donnée ? ». Elle répond aux besoins de transversalité de l'organisation et

	d'interopérabilité des données, tout en simplifiant l'identification des sources pour de nouveaux tableaux de bord ou applications. *Exemple : Une fois entièrement déployé, GéoBasque sera la source unique de vérité, le point central de consommation des données géographiques liées aux compétences communautaires.*
API (*Application Programming Interface)*	Moyen de connecter des applications entre elles pour échanger des services et des données via des langages et des méthodes largement standardisés. C'est en quelque sorte un point de connexion standardisé entre deux systèmes, qu'ils soient internes ou externes. *Exemple : GéoBasque est interfacé via API avec la Base Adresse Nationale (BAN) ce qui permet d'échanger des données de façon standardisée avec un service extérieur à la CAPB.*
Système MDM (*Master Data Management*)	Outil ou ensemble d'outils permettant de stocker, gérer et diffuser les données transversales au sein d'une organisation. Il permet de suivre le cycle de vie de la donnée, les flux de validation, la traçabilité (date et auteur des modifications) … en se calquant sur les processus opérationnels. *Pas d'exemple de système MDM à la CAPB.*

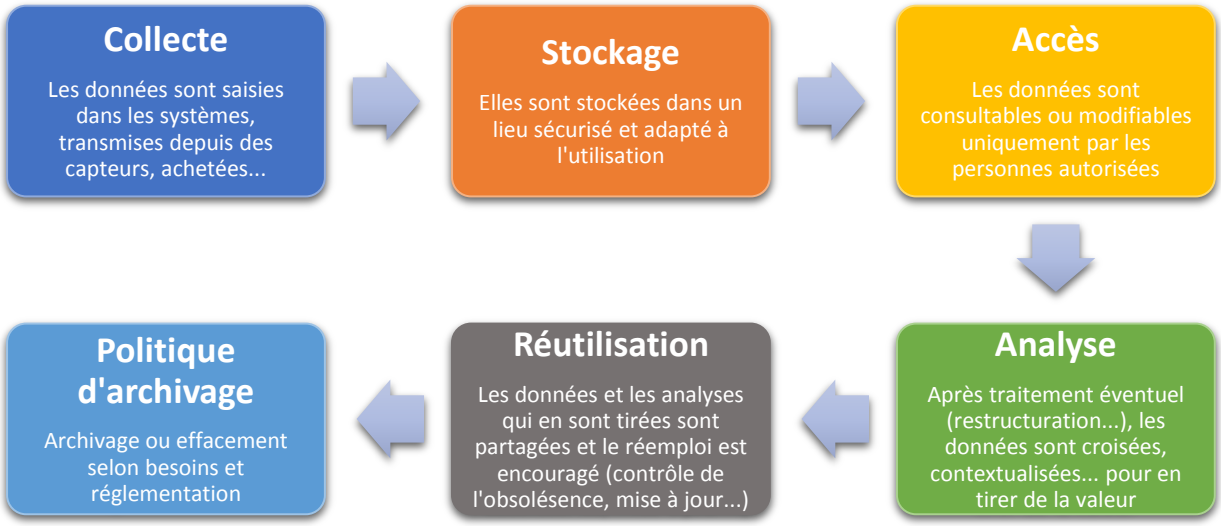
Illustration des définitions



4. Cycle de vie et intégrité de la donnée

Deux notions importantes dans le domaine de la gestion de la donnée concernent le cycle de vie de la donnée et son intégrité.

4.1 Cycle de vie de la donnée



4.2 Métadonnées

Les métadonnées sont d'une importance capitale dans la bonne gestion des données. Il s'agit d'informations décrivant le contenu de ces données : définition de cette donnée, date de création/modification, personne ou capteur l'ayant saisie, coordonnées GPS, type de licence…. Elles fournissent le contexte – et éventuellement les limites – permettant leur bonne utilisation.

Les objectifs et les critères des métadonnées sont donnés par Etalab (administration publique).
Elles ont pour objectif de répondre principalement aux deux problématiques suivantes :

- d'une part, à améliorer la confiance dans les jeux de données ; les utilisateurs de données peinent à identifier les jeux de données de qualité et à évaluer si tel ou tel jeu de données est digne d'intérêt ;
- d'autre part, inciter les producteurs de données à améliorer la qualité de leurs données et à identifier les points d'amélioration.

4.3 Intégrité de la donnée

La gouvernance des données vise notamment à garantir leur intégrité tout au long de leur cycle de vie. C'est ce qui garantit qu'une donnée de qualité est mise à disposition des utilisateurs. Elle repose sur plusieurs axes :

- l'**exactitude** : la donnée décrit de façon juste et adéquate ce qu'elle doit mesurer. Cela veut dire qu'elle a été saisie par une personne habilitée et formée, ou un capteur adapté au besoin, et si nécessaire, qu'elle est passée par un flux de validation.
- la **complétude** : la donnée a été saisie dans son entièreté (unité de mesure), les métadonnées contextuelles sont renseignées (définitions, limites éventuelles).
- la **cohérence** : la même donnée est partagée partout dans l'écosystème et il n'existe pas de données concurrentes.
- la **contemporanéité** : la donnée est saisie au moment de sa mesure et est disponible quand nécessaire pour la bonne application des processus opérationnels ou décisionnels.
- la **lisibilité** : la donnée est lisible et intelligible par les humains et les machines pour son exploitation : les formats sont connus et normalisés, le langage utilisé est le même partout (nomenclatures).
- l'**originalité** : la donnée est celle qui a été saisie par les personnes habilitées et n'a pas été altérée. Elle provient directement du système défini comme la source unique de vérité et pas d'un « rebond » depuis un autre système.
- la **traçabilité** : toutes les étapes du cycle de vie de la donnée ont été documentées, les métadonnées (date et heure de modification, traitements, personne ou système responsable de la modification…) liées à sa collecte puis ses modifications sont connues et partagées.
- la **fiabilité** : réalisation d'Analyse des Modes de Défaillances, de leurs Effets et de leurs Criticités (AMDEC), traitement des retours qualité dans une boucle d'amélioration continue, contrôle manuel régulier d'un échantillon de données pour éviter l'obsolescence...

5. Pratiques obligatoires pour la gestion des données

La gestion des données requiert :

- des rôles et responsabilités définis pour gérer les activités opérationnelles et garantir les niveaux appropriés de qualité des données ;
- des processus clairs et formalisés, des méthodologies standardisées, ainsi que des pratiques « de qualité dès la conception » ;
- l'allocation des ressources nécessaires (processus, systèmes et compétences) au niveau de la Direction des systèmes d'information de la CAPB ainsi que dans les directions métiers.

5.1 Organisation de la Gestion des Données

La direction qui saisit une donnée la plus en amont dans son cycle de vie est (généralement) considérée comme propriétaire de cette donnée et doit en être à l'origine. Les directions désignent un propriétaire de données sur un périmètre précis, secondé par un ou des gestionnaires de données.

La matrice RACI ci-dessous détaille les rôles et responsabilités.

	Propriétaire de données (**PdD**)	Gestionnaire de données (**GdD**)	Responsable du pilotage des données	Communauté de la donnée	Direction (ou Manager du **PdD**)
Nomme le PdD et lui fournit les moyens d'accomplir la mission (incluant un GdD le cas échéant)	**C**	**I**	**C**	**I**	**A/R**
Défini les objectifs de qualité	**A/R**	**I**	**C**	**C**	**I**
Rédige et fait appliquer la procédure	**A/R**	**C**	**C**	**I**	**I**
Evolutions et changements (ajouts d'attributs, de valeurs d'attributs…)	**A/R**	**C**	**C**	**C**	**I**
Gère les activités quotidiennes	**A**	**R**	**I**	**I**	**I**

R : Responsable (réalise) A : Approbateur (autorité) C : Consulté I : Informé

5.2 Processus et Organisation

Les responsables hiérarchiques des propriétaires et gestionnaires de données portent la responsabilité :

- d'intégrer les activités liées aux données dans les fiches de postes et de les cadrer par des objectifs – il s'agit du levier le plus efficace pour améliorer et maintenir la qualité des données ;
- de nommer un suppléant – appuyé par des procédures et modes opératoires – disponible et opérationnel pour les activités liées aux données, selon leur niveau de criticité ;
- d'établir un plan de succession à moyen terme décrivant les compétences et ressources nécessaires pour gérer les responsabilités associées aux données.

5.3 Bonnes pratiques

Le but de la gouvernance des données est de tendre vers une source unique de vérité, un système où l'ensemble des données transversales et validées est à la disposition des systèmes et des utilisateurs. En ce sens, certaines bonnes pratiques sont à mettre en avant :

- aucun autre système ne devra redistribuer les données, tous les usages devront se sourcer directement sur la source unique de vérité ; les directions métiers pourront en revanche, dans leur système métier, enrichir ces données par des attributs qui leurs sont propres ;
- aucun processus opérationnel ni aucun système ne pourra créer ou dupliquer, utiliser ou maintenir un référentiel de données séparé ;
- une qualité de données minimale sera attendue avant l'intégration dans la source de vérité ;

- les systèmes sources (qui peuvent alimenter la source de vérité) devront utiliser une approche de « qualité par le design » afin de limiter les possibilités de saisies de données erronées ou de simplifier leurs détections via des contrôle de cohérence.

6. Evaluation et Suivi de la Qualité des Données

Le propriétaire de données est comptable d'assurer le respect du niveau de qualité défini dans la procédure associée à son domaine. Cette responsabilité se traduit par l'intégration d'objectifs liés à la qualité dans ses objectifs de performance individuels et annuels.

6.1 Cible de qualité des données

La montée en qualité des données puis le maintien du niveau adéquat suppose de disposer de ressources humaines et techniques dédiées au sujet. Il est important de distinguer le niveau de qualité nécessaire à la bonne exploitation des données.

Quatre grandes cibles de qualité peuvent être identifiées :

- qualité **réglementaire** : niveau maximal fixé par des standards extérieurs ou par la « criticité » des données comme les données souveraines (opposables), les données sanitaires (qualité de l'eau…), les données personnelles (encadrées par le RGPD)…
- qualité **décisionnelle** : niveau de qualité qui permet aux élus de décider et d'arbitrer, au comité de direction de déterminer les grandes orientations stratégiques…
- Qualité **opérationnelle** : niveau de qualité qui sert à l'exécution des processus opérationnels courants ;
- Qualité **informationnelle** : niveau de qualité minimum pour l'analyse de tendance, la mise en place de tableaux de bord…

6.2 Axes de contrôle de la qualité

La qualité de la donnée se décline selon quatre axes :

- **Gouvernance**, concernant l'application de l'organisation de la gouvernance ;
- **Disponibilité**, concernant la mise à disposition de la donnée dans un système ;
- **Fiabilité**, concernant la ponctualité d'une donnée dans un système ainsi que les contrôles de saisie ou de recherche d'erreurs automatiques mis en place ;
- **Exactitude**, concernant la justesse d'une donnée via la vérification manuelle d'échantillons de données.

7. Responsabilités et comitologie

7.1 Comité de Pilotage

Le comité de pilotage a la charge de donner les grandes orientations sur les sujets de gouvernance et de mise à disposition des données. Il aura aussi la charge de suivre la bonne exécution des politiques qu'il aura définies, en s'assurant de l'adéquation entre les ambitions et les moyens mis à disposition des différentes parties prenantes (DSI, DGA…).

Il est composé du Comité de Direction Générale et du DSI , présidé par l'Elu en charge du Numérique et animé par le Responsable du Pilotage Stratégique des Données.

7.2 Comité technique

Le comité technique est composé d'experts et de référents sur les données qui se fixent la tâche collective de définir la gouvernance de la donnée et d'améliorer sa qualité au sein de la Communauté d'Agglomération Pays basque.

Ce comité aussi appelé « Communauté de la donnée » a pour objectif de favoriser le partage entre agents et directions volontaires:

- échange de bonnes pratiques et d'initiatives,
- mutualisation de ressources et de compétences sur des sujets transversaux,
- élaboration et partage de standards au sein de l'organisation

8. Evolution du document

Le présent document a pour vocation d'évoluer en fonction des besoins, des normes… ou de l'organisation de la CAPB etc. Il est donc prévu de le revoir *a minima* tous les ans afin de s'assurer de son adéquation avec le contexte du moment.