

| | |
|-------------|---|
| Titre | Données sur l'eau (4) : la fiabilité des données |
| Description | Note pour l'évaluation de la politique de l'eau |
| Éditeur | Office national de l'eau et des milieux aquatiques |
| Date | 2013-04-05 |
| Version | 1 |

1. Objectif

Ce document a pour objectif de préciser ce qui est communément désigné comme la « fiabilité des données » et de décrire la variété des mécanismes mis en place à cette fin dans le cadre du SIE en présentant des exemples concrets.

2. Qu'est ce que la fiabilité des données ?

La notion de « donnée fiable » n'a en soi aucun sens, alors qu'elle est communément employée.

La fiabilité doit d'abord être examinée au regard d'un *usage*, et comprise comme la *confiance* en la capacité à satisfaire cet usage. La confiance implique nécessairement un *usager* et l'usage comporte une finalité et des exigences particulières, dites *exigences de qualité*. La fiabilité est par nature un *jugement* porté par un usager en fonction des exigences de l'usage visé.

L'ensemble des moyens qui permettent à l'usager de former ce jugement est généralement qualifié d'*assurance qualité* dans le cadre d'un *système de gestion de la qualité*.

Dans le cas d'un système d'information, plutôt que de donnée fiable, on devrait parler d'*information* fiable, en incluant dans ce concept non seulement la donnée elle-même mais aussi son *contexte de création*, sa *généalogie* (les autres informations dont la donnée dérive et le mode de dérivation) et sa *sémantique* (comment sa signification peut être élaborée).

Le contexte de création d'une donnée comporte notamment des informations sur la méthode employée (le contexte méthodologique : un protocole d'observation, une méthode d'analyse, un référentiel d'évaluation, etc.), sur la réalisation de l'opération de création (le contexte opératoire : quand, où, par qui, etc.) et sur l'environnement de cette opération (le contexte environnemental : informations indépendantes de l'opération de création, comme par exemple les conditions hydrologiques d'un prélèvement d'eau).

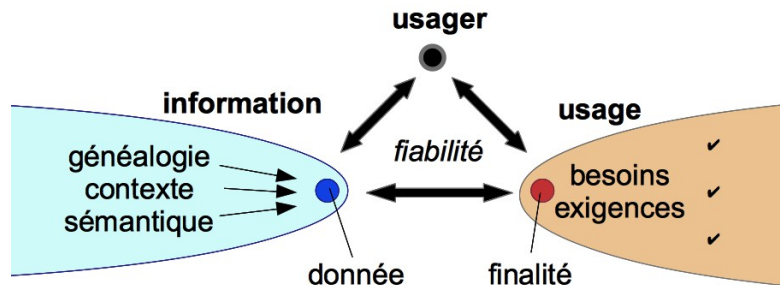
La fiabilité n'est donc pas une caractéristique intrinsèque d'une donnée. Ce n'est pas non plus une notion technique mesurable comme la justesse ou la fidélité d'un appareil de mesure ou comme l'incertitude d'une mesure, bien que ces notions techniques puissent intervenir pour juger qu'une donnée est fiable.

Exemple : « une mesure de pollution est-elle fiable ? »

Une mesure de concentration d'un polluant dans l'eau (une donnée) ne peut pas être jugée fiable, pour conclure à l'existence d'un risque pour le milieu (finalité), tant que la limite de quantification de la méthode employée (contexte méthodologique) n'est pas connue, car la limite de quantification doit être inférieure à la norme de qualité applicable (exigence de qualité) ; en outre, l'exigence sur la limite de quantification sera plus faible pour constater une pollution accidentelle fla-

grante que pour étudier les effets chroniques d'une pollution historique (deux usages différents de la même information). Il n'est donc pas possible de répondre simplement à la question générale « une mesure de pollution est-elle fiable ? ».

On peut résumer la fiabilité, en tant que jugement porté par un usager sur une information, relativement à un usage, par le schéma suivant.



3. Quels sont les rôles des acteurs ?

Outre l'usager, de nombreux autres acteurs jouent un rôle déterminant dans la fiabilité de l'information.

Compte tenu du caractère réglementaire d'une grande partie du cadre méthodologique du SIE (programmes de surveillance, état des lieux, déclarations d'auto-surveillance, rapports du maire sur le prix et la qualité des services, etc.), la responsabilité première de l'adéquation des données aux usages réglementaires revient à l'*autorité administrative*. Elle établit les normes auxquelles doivent satisfaire les dispositifs de production et de gestion des données quand ils sont institués par la loi et arrête la définition de ces dispositifs ; elle assigne des objectifs aux organismes techniques chargés d'élaborer et de mettre en place ces dispositifs et contrôle leur mise en œuvre. Les acteurs concernés sont au niveau national le MEDDE (direction de l'eau et de la biodiversité, direction générale de la prévention des risques), la direction générale de la santé et au niveau territorial les préfets, en particulier les préfets coordonnateurs de bassin.

Au niveau technique, c'est l'Onema qui assure la coordination du SIE, selon les modalités du schéma national des données sur l'eau (SNDE), approuvé par l'arrêté du 26 juillet 2010, pris en application du décret N°2009-1543 du 11 décembre 2009. Cette coordination technique concerne notamment l'élaboration de trois référentiels techniques :

- ▶ le *référentiel méthodologique* ;
- ▶ le *référentiel des données* ;
- ▶ le *référentiel qualité*, qui porte sur l'ensemble des processus du SIE, de l'élaboration des méthodologies et du référentiel des données à la production, la bancarisation, le traitement, la valorisation et la diffusion des données et qui vise à « assurer la fiabilité et la traçabilité des données et des services du SIE » (§ 12 du SNDE).

Ce référentiel qualité, qui définit le système de gestion de la qualité du SIE, assigne à son tour des rôles aux différents acteurs intervenant dans les différents processus du SIE, notamment la produc-

tion et la bancarisation des données. Un rôle essentiel revient aux nombreux *producteurs de données* :

- ▶ quelques 24 000 autorités organisatrices des services publics d'eau et d'assainissement (prix et performances des services) ;
- ▶ des services déconcentrés de l'État (hydrométrie, piézométrie, hydrobiologie, organisation des services) ;
- ▶ des établissements publics : l'Onema (poissons, hydromorphologie, continuité, température, étiages), le BRGM (piézométrie), l'Ifremer (eaux littorales) ;
- ▶ les secrétariats techniques de bassin (agence, délégation de bassin, Onema), pour les données de la planification (pressions, classification de l'état, objectifs, etc.) et leur rapportage ;
- ▶ des collectivités et d'autres organismes (associations, etc.) à associées à leur demande au SIE ;
- ▶ des bureaux d'études et des laboratoires d'analyses.

Au sein de tous les acteurs ayant à traiter les données (producteurs de données, opérateurs des banques de données, etc.), la fonction d'*administrateur de données* est primordiale en ce qui concerne la fiabilité des données.

4. Comment la fiabilité d'une information peut-elle être établie ?

4.1. La fiabilité d'une information peut être réfutée

Tout usager d'une information peut prouver qu'elle n'est pas fiable pour l'usage qu'il compte en faire, selon les exigences liées à cet usage, en lui appliquant des *contrôles de qualité* et en constatant que certains d'entre eux échouent.

Exemple : en constatant que la limite de quantification n'est pas connue, ou que la localisation géographique d'une mesure est anormale, l'usager peut prouver qu'une information n'est pas fiable (pour l'usage visé).

Une illustration d'un contrôle de localisation d'une mesure, qui devrait se trouver sur un cours d'eau, est présentée sur la figure suivante, avec deux conclusions différentes selon qu'il s'agit d'un petit ou d'un grand cours d'eau.



Les contrôles de qualité s'inscrivent, comme l'assurance qualité, dans un système de gestion de la qualité. Ces contrôles peuvent être plus ou moins poussés, selon le niveau d'exigence de qualité requis par l'usage de l'information.

4.2. La fiabilité d'une information ne peut pas (toujours) être prouvée

Il ne suffit pas que tous ses contrôles de qualité aient réussi, pour que l'utilisateur puisse prouver que l'information est fiable. En effet, l'utilisateur n'est pas nécessairement capable d'énoncer les exigences de qualité propres à l'usage visé ou de traduire ces exigences en contrôles de qualité, lesquels sont toujours partiels ; en outre, il ne maîtrise pas l'ensemble des opérations qui ont conduit à cette information, notamment parce que ces opérations ne sont pas toujours reproductibles.

Exemple : l'utilisateur vérifie la localisation, mais peut-être pas la date de la mesure, il n'a aucun moyen d'exclure la possibilité d'une erreur de lecture d'un instrument de mesure par le producteur de la donnée et même s'il cherchait à renouveler cette lecture, il est normal qu'une mesure répétée à des instants différents dans un milieu naturel donne des résultats différents (le vivant bouge!).

4.3. La fiabilité d'une information repose sur la qualification et la traçabilité des données

Après avoir constaté le succès ou l'échec de ses contrôles de qualité, l'utilisateur peut affecter aux données son estimation de leur fiabilité : données vérifiées, douteuses, etc. S'il s'agit d'un producteur, il peut aussi attester de la conformité à un référentiel de qualité des opérations qu'il a effectuées (par exemple, attester que les analyses sont produites « sous accréditation »). Pour que ces informations soient exploitables par d'autres usagers, il est nécessaire qu'elles soient associées à la donnée, sous forme d'éléments de métadonnées ou de commentaires.

Ces informations supplémentaires, que chaque producteur (intermédiaire) peut associer à la donnée, constituent une *qualification* de la donnée ; ces qualifications doivent être préservées afin que les usagers puissent retracer, sur toute la généalogie de la donnée, les contrôles de qualité effectués et les référentiels de qualité employés.

L'examen de la fiabilité d'une information doit donc considérer sa généalogie, c'est-à-dire à partir de quelles autres informations et par quelle suite d'opérations elle a été obtenue : le traitement de ces informations de généalogie constitue la *traçabilité* de la donnée.

4.4. Que faire d'une information qui n'est pas jugée fiable ?

Trois cas se présentent quand une information n'est pas jugée fiable par l'utilisateur :

- ▶ elle n'est pas utilisée, ce qui compromet l'usage visé ;
- ▶ elle est utilisée, mais le résultat de son exploitation est assorti de mises en garde sur le caractère « non fiable » de certaines données utilisées, ce qui laisse l'utilisateur du résultat décider ce qu'il en fera ;
- ▶ elle est utilisée, mais le résultat de son exploitation n'est assorti d'aucune mise en garde.

Le 3ème cas est certainement le plus dangereux, non seulement pour l'usage qui en est fait, mais plus généralement pour la crédibilité de l'ensemble du SIE. On ne peut cependant pas empêcher un usager quelconque d'utiliser une donnée qu'il sait non propre à cet usage

4.5. Quels sont les enjeux ?

Les enjeux de la fiabilité des données sont liés aux conséquences négatives que peuvent avoir les deux situations suivantes pour un usage déterminé des données :

- ▶ l'absence ou l'insuffisance de données fiables,
- ▶ l'utilisation de données non fiables sans mise en garde.

Les enjeux dépendent évidemment des usages, dont le périmètre est déterminé par les sept finalités du SIE, telles qu'elles sont fixées par le SNDE (au § 4, qui les désigne comme ses « objectifs ») :

1. *« La vérification de conformité de la mise en œuvre de la législation environnementale sur l'eau, par les autorités nationales et communautaires ;*
2. *L'analyse des pressions dues aux activités humaines, les analyses économiques et l'évaluation de l'état des eaux, par les autorités de bassin et les services locaux, sur la base d'une connaissance objective de l'état des milieux et des usages ;*
3. *L'évaluation de l'efficacité et de l'efficience des politiques publiques, notamment en ce qui concerne la performance des services publics d'eau et d'assainissement, par les responsables de ces politiques et de ces services publics ;*
4. *L'aide à la décision, notamment pour l'exercice de la police de l'eau, la définition des programmes de mesures et la prise en compte des risques liés à l'eau dans les politiques d'urbanisation et d'infrastructures, ou la gestion des crues et des sécheresses ;*
5. *L'information du public ;*
6. *La fourniture des données nécessaires à d'autres systèmes d'information (santé, travail, risques, statistique publique, etc.) ;*
7. *La contribution, en particulier en matière de données patrimoniales, aux programmes de recherche appliquée et à la connaissance rendus de plus en plus nécessaires du fait de la complexité croissante des processus en cause. »*

Ces enjeux sont esquissés dans le tableau suivant, selon chacune des finalités du SIE :

| Finalités | | Conséquences si | |
|-----------|----------------------------|---------------------------------|---|
| | | Insuffisance de données fiables | Utilisation de données non fiables sans mise en garde |
| 1 | Vérification de conformité | Contentieux à court terme | Évaluation faussée, contentieux lourds à moyen terme |
| 2 | État de l'environnement | Ignorance | Mauvaise compréhension des enjeux, hiérarchisation des problèmes inadaptée, |

| | | | |
|---|--|---|--|
| | | | anticipation insuffisante |
| 3 | Évaluation des politiques publiques | Pilotage défectueux, redevabilité inadéquate | Orientations erronées, mauvaises décisions à moyen terme, décrédibilisation de l'action publique |
| 4 | Aide à la décision | Décisions non justifiées, absence de décision, contestations | Mauvaises décisions à court terme, coûts pour la collectivité, dommages sur l'environnement |
| 5 | Information du public | Participation inadéquate aux processus de décision, désintérêt ou préjugés tenaces, redevabilité insuffisante | Mauvaise compréhension des enjeux, défiance envers certains usagers ou acteurs |
| 6 | Fourniture à d'autres systèmes d'information | Difficultés d'intégration des politiques et d'approches interdisciplinaires | Dégradation de ces systèmes d'information |
| 7 | Contribution aux programmes de recherche | Désengagement des acteurs de la recherche | Résultats faussés, anticipations erronées, modèles incorrects, décrédibilisation de la recherche |

Exemple : « 2/3 des masses d'eau de surface doivent être en bon état écologique en 2015 »

Cet objectif a été fixé par le Grenelle de l'environnement en octobre 2007. À cette époque, on ne disposait que des données des réseaux pré-existants à la DCE, dont la représentativité privilégiait les pressions d'origine domestique, d'un système d'évaluation incomplet, d'une délimitation des plus grandes masses d'eau et d'aucune base de données économiques. Les programmes de surveillance commençaient à être mis en place, leurs données n'étaient pas disponibles, la délimitation des masses d'eau, dont le nombre allait être multiplié par trois, n'était pas achevée, le système d'évaluation de l'état des eaux n'a été finalisé qu'en 2009, la faisabilité technique, les délais de réponse des milieux et le coût des mesures envisagées n'étaient pas suffisamment connus pour justifier les exemptions à l'objectif de bon état. Les données disponibles, qui n'étaient pas adaptées aux exigences de la DCE, n'ont pas pu être utilisées pour étayer cet engagement du Grenelle de l'environnement, qui allait pourtant être traduit dans la loi en 2009.

Il y a ainsi un enjeu majeur à ce que les données utilisées dans le cadre de la politique de l'eau soient « fiables », non seulement pour garantir la crédibilité des résultats obtenus, mais également pour étayer les orientations de cette politique ou pour les rectifier. La logique interne des directives européennes (DERU, DCE, Nitrates, DCSMM, etc.), qui servent aujourd'hui de fondement à la politique de l'eau nationale, requiert explicitement la production et l'usage de données pour orienter leur mise en œuvre et établit pour certaines d'entre elles des exigences de qualité.

5. Le système de gestion de la qualité

L'ensemble des moyens contribuant à la fiabilité des données constitue le système de gestion de la qualité du SIE. Ce système est déployé tout au long de la chaîne de production, de gestion et de diffusion des données.

5.1. Le référentiel méthodologique

Au delà des textes réglementaires qui prescrivent à un moment donné l'emploi de certaines méthodologies (protocoles, méthodes d'analyse, définitions d'indicateurs, etc.), le référentiel méthodologique traduit un état de connaissances qui est appelé à évoluer (nouveaux protocoles, nouvelles problématiques, progrès scientifique, etc.).

Des évolutions trop fréquentes du référentiel méthodologique, non contrôlées et mal préparées sont une source de défiance et d'inefficacité, de même qu'une absence d'évolution, qui implique un retard croissant de l'action sur l'état des connaissances.

Exemple : la campagne exploratoire sur les micropolluants dans les eaux de surface

En complément des programmes de surveillance courants, mis en œuvre par des laboratoires agréés, une campagne exploratoire a été conduite en 2011-2012, en recourant aux « meilleures pratiques » analytiques disponibles dans les laboratoires de recherche, pour mesurer la présence de nouvelles substances polluantes non surveillées jusqu'à maintenant. Les résultats de cette campagne apporteront de nouvelles connaissances qui permettront d'orienter les prochains programmes de surveillance, qui s'appliqueront à partir de 2015.

L'organisation de cette évolution du référentiel méthodologique est une part importante de la confiance accordée aux données : par qui, comment et quand sont validées et décidées les évolutions. La qualité de ces processus de décision, à la fois technique et politique, contribue ainsi à la fiabilité des données.

5.2. Le référentiel des données

Compte tenu du caractère central du référentiel des données pour la cohérence du SIE et l'interopérabilité de ses composantes, tout défaut de qualité a de lourdes conséquences sur la fiabilité de l'ensemble du SIE.

C'est pourquoi la codification des jeux de données de référence à administration centralisée et l'élaboration des dictionnaires et des scénarios d'échange par le secrétariat technique du Sandre sont passées sous certification ISO 9001.

Exemple : l'abandon des codes provisoires

Le Sandre a décidé de ne plus affecter des « codes provisoires » aux éléments du référentiel, qui avaient l'inconvénient de perdurer dans les bases de données et de causer des erreurs de synonymie (un même objet désigné par plusieurs codes). Cette nouvelle règle de gestion du référentiel des données améliorera sensiblement la qualité des données, non seulement des données de référence, mais de l'ensemble du SIE. La contrepartie de cette règle est l'engagement à répondre dans un délai réduit à toute demande de création d'un nouveau code, délai qui est maintenant d'une quinzaine de jours.

La mise en place prochaine d'une plateforme d'administration répartie des référentiels (pour les interlocuteurs, les points de prélèvements, etc.) devra aussi être accompagnée d'une démarche qualité.

5.3. La production de données

Les différents dispositifs de production des données (réseaux de surveillance, déclarations, enquêtes, etc.) comportent leurs propres démarches d'amélioration de la qualité.

5.3.1. Données de surveillance des milieux

La fiabilité des données produites nécessite la maîtrise de la totalité de la chaîne d'acquisition depuis l'échantillonnage jusqu'à la restitution des résultats.

Partant du constat que les concepts d'assurance qualité sont encore peu appliqués à la phase d'échantillonnage, l'Onema a engagé avec Aquaref des actions d'amélioration de la qualité des *pré-lèvements*. Ces travaux ont conduit à la production et à la diffusion de guides méthodologiques intégrant des recommandations en termes de contrôles de qualité. La réalisation de *formations* sur l'échantillonnage (déjà effective pour les cours d'eau et les eaux souterraines) a pour but de garantir une homogénéité des pratiques des préleveurs et de ce fait une comparabilité des résultats. Ces formations ont vocation à devenir qualifiantes.

L'amélioration de la qualité de la phase d'analyse à proprement parler fait appel à un ensemble de dispositifs. Les analyses doivent être réalisées par des *laboratoires agréés* par le ministère chargé de l'environnement. Un arrêté d'octobre 2011 précise les modalités d'agrément qui doivent en premier lieu posséder une accréditation délivrée par le Cofrac dans leur domaine de compétence. L'inspection des dossiers d'agrément, après vérification des critères d'agrément par des auditeurs Cofrac, est réalisée par l'Onema. Les performances techniques attendues, en matière d'*incertitude des mesures* et à de *limite de quantification* des méthodes sont arrêtées par la DEB sur avis d'Aquaref conformément à la directive « QA/QC¹ » de juillet 2009. L'agrément prend en compte les phases amont et aval de l'analyse en imposant d'une part l'accréditation sur la phase d'échantillonnage et en exigeant d'autre part la connaissance et l'utilisation du format d'échange EDILABO pour la réception des demandes d'analyses et la fourniture des résultats.

L'expertise d'Aquaref est mise à contribution de façon croissante pour l'amélioration des analyses à travers la mise à disposition de méthodes et d'outils : développement de protocoles, participation à la normalisation, mise à disposition de « fiches méthodes », développement d'essais inter-laboratoires (EIL)...

Les concepts d'assurance qualité et de contrôle qualité (QA/QC) bien développés pour la chimie sont de plus en plus pris en compte dans le domaine complexe de l'hydrobiologie. Les éléments de qualité biologiques suivis dans le cadre de la DCE sont ainsi en grande majorité couverts par l'accréditation et l'agrément, à l'exception du compartiment piscicole. Les travaux d'Aquaref permettent d'améliorer progressivement les connaissances en termes de sources d'incertitudes des différents protocoles appliqués ou encore de définir ce que peuvent être des EIL en hydrobiologie.

Exemple : les dénombrements de taxons

En biologie, des études menées en 2011 (Onema-Aquaref) ont montré que le dénombrement d'espèces observées dépend significativement de l'opérateur (au moins 10% d'écart pour un opérateur sur deux, au moins 25% pour un opérateur sur trois) ; des taxons peuvent être oubliés par certains opérateurs, et jusqu'à 30% des taxons oubliés expriment une sensibilité forte à très forte aux pollutions, ce qui conduit à un indicateur sous-évalué par ces opérateurs.

¹ Directive 2009/90/CE de la Commission du 31 juillet 2009 établissant, conformément à la directive 2000/60/CE du Parlement européen et du Conseil, des spécifications techniques pour l'analyse chimique et la surveillance de l'état des eaux

Pour assurer la cohérence de ces différentes étapes de la production de données Aquaref et plus particulièrement le LNE, laboratoire national de métrologie et d'essais, travaille à la définition d'un schéma de traçabilité métrologique des données permettant, par un suivi de l'ensemble du processus d'acquisition, de connaître le niveau de fiabilité des données finales.

Il y a une incertitude attachée à la mesure, en chimie comme en biologie, qui n'est pas encore bien caractérisée : choix du point de mesure (variabilité spatiale), choix de la date de mesure (variabilité temporelle, due notamment aux conditions climatiques), choix de l'opérateur (préleveur, laboratoire).

Plusieurs types de données sont encore hors du champ de la procédure d'agrément, en l'absence d'accréditation. C'est le cas des données de surveillance piscicoles et hydromorphologiques.

L'acquisition de données piscicoles bénéficie d'un protocole normalisé, bien connu et appliqué en routine, de sessions de formation et de journées techniques d'intercalibration dont l'objectif est l'amélioration et l'harmonisation des pratiques des opérateurs (agents de l'Onema, bureaux d'étude prestataires). Les agences de l'eau font par ailleurs appel à l'Onema pour le contrôle de la qualité sur le terrain des pêches électriques réalisées par leurs prestataires dans le cadre des contrôles opérationnels de la DCE. C'est pourquoi la démarche qualité dans laquelle l'Onema s'est engagé depuis 2009 s'applique progressivement à la production de ces données ; elle devra couvrir à terme toutes les phases de l'acquisition des données sur les peuplements de poissons, qu'elles soient produites par des agents de l'Onema ou par ses prestataires (mise en place d'audits internes, obtention d'une certification ISO 9001 pour l'externalisation de sa production de données...). Il s'agit également de développer un agrément pour le compartiment poisson. L'étape préalable étant la mise en place d'une accréditation, l'Onema assurera une expertise technique auprès du Cofrac pour la définition des exigences relatives à cette accréditation sur le protocole d'échantillonnage des poissons à l'électricité en cours d'eau.

L'Onema est par ailleurs investi dans la définition et le déploiement progressif de protocoles de caractérisation des habitats et des risques d'altération hydromorphologique. Ces suivis hydromorphologiques, plus récents, ne sont encore que peu couverts par la démarche qualité mais ont vocation à l'être au fur et à mesure de leur application.

Les DREAL disposent, lorsqu'elles sont productrices des données d'hydrométrie sur les cours d'eau, d'une démarche qualité commune dont le SCHAPI garantit l'application. Concernant les eaux souterraines, le réseau de suivi piézométrique national est opéré par le BRGM, qui dispose d'un plan qualité sous contrôle des DREAL.

Les agences de l'eau sont pour leur part engagées dans la mise en œuvre de démarches qualité reconnues par l'obtention de la certification ISO 9001 de tout ou partie de leurs activités (dont les activités d'acquisition de valorisation et de mise à disposition des données sur l'eau). D'autres établissements publics comme Ifremer, le BRGM et l'Ineris, appliquent également une démarche qualité dans leurs activités.

5.3.2. Données déclaratives

Les dispositifs de déclaration, même obligatoires, ne permettent pas d'imposer aux producteurs de données une démarche qualité. Il revient à l'opérateur du dispositif de mettre en place des contrôles de qualité.

Exemple : les contrôles de qualité sur les données de SISPEA
Les DDT effectuent des contrôles de qualité sur les données rapportées par les collectivités (autorités organisatrices des services) afin de les qualifier (vérifiées, douteuses, etc.) avant leur publication ; les collectivités peuvent alors les corriger puis les publier. Les analyses statistiques ne sont réalisées que sur des données vérifiées.

5.3.3. Données élaborées

Les données élaborées sont produites à partir d'autres données par des producteurs intermédiaires. C'est notamment le cas de toutes les données produites dans le cadre des plans de gestion des bassins : données de caractérisation des pressions, de classification de l'état des eaux, objectifs, coûts des mesures, etc. Dans chaque cas, ces producteurs intermédiaires sont amenés à contrôler les données qu'ils utilisent et à qualifier celles qu'ils produisent.

Exemple : le niveau de confiance de l'état écologique
Le rapportage des plans de gestion demande que l'état écologique soit accompagné d'un « niveau de confiance », qui est une forme de qualification de cet état ; trois niveaux sont prévus par la Commission (haut, moyen, bas – ou inconnu) mais il revient à chaque État membre de décider comment il est attribué. Pour la France, ce niveau de confiance est attribué à l'état d'une masse d'eau selon le mode de détermination de l'état (calculé à partir de données de surveillance suffisantes sur cette masse d'eau ou extrapolé à partir d'autres données de surveillance et de données de pressions, selon que ces différentes données concordent ou non).
Ce niveau de confiance est présenté dans les documents de planification, mais sa prise en compte est souvent négligée, au vu des conclusions hâtives qui en sont tirées, sachant que l'état écologique de la majorité des masses d'eau a un niveau de confiance « faible » (64%) ou « moyen » (23%).

5.3.4. Données de rapportage

Chacune des directives relatives à l'eau, notamment la DCE, requiert des rapportages formalisés (incluant des jeux de données répondant à des spécifications techniques établies au niveau européen) à échéances régulières. Ces rapportages sont effectués par les administrations centrales (DEB, DGPR ou DGS) et s'appuient sur des dispositifs techniques propres à chaque directive.

Dans le cas de la DCE, l'Onema intervient dans le cadrage méthodologique (guides), l'organisation du rapportage interne, à partir des secrétariats techniques des bassins, la collecte des données et le contrôle de leur cohérence.

Ces rapportages impliquent plusieurs types de contrôles, dont certains ont été automatisés et peuvent être effectués par les producteurs des données :

- ▶ contrôles visant à vérifier la conformité aux spécifications techniques du rapportage : contrôle du format, de la complétude (les informations obligatoires sont-elles fournies?), des codifications utilisées, de cohérence technique (les points de surveillance sont-ils sur le territoire français, les valeurs annoncées sont-elles possibles, la topologie est-elle correcte ?), etc.

- ▶ contrôles visant à vérifier la cohérence des informations relatives à différents domaines (descriptifs des masses d'eau, des programmes de surveillance, des programmes de mesure, etc.), aux rapportages précédents et à ceux réalisés au titre d'autres directives, à d'autres sources de données, ou encore aux différents bassins ;
- ▶ des contrôles métier qui visent à détecter des risques de non-conformité à la directive.

Si tous les contrôles sont passés avec succès, les jeux de données sont transmis à la DEB avec la liste des contrôles effectués.

Exemple : le contrôle des zones vulnérables

Pour le rapportage des zones vulnérables au titre de la directive nitrates, le secrétariat technique du Sandre a réalisé une série de contrôles au regard du dictionnaire de données de l'Agence européenne de l'environnement et des arrêtés de bassin classant les communes en zone vulnérable. L'ensemble des contrôles étant réussis, les données ont été transmises à la DEB avec un rapport décrivant ces contrôles et concluant : « La donnée finale transmise NVZ_DES_GIS_2012.shp est conforme au modèle défini par le Data dictionary et les arrêtés des bassins ».

Le rapport signale également les six communes pour lesquelles les arrêtés de bassin ne sont pas conformes aux derniers changements de nom figurant dans le code officiel géographique publié par l'INSEE, sans que cela aboutisse à une ambiguïté sur les communes concernées ; il n'appartient cependant pas au Sandre de vérifier les arrêtés, ceux-ci étant considérés comme une référence à partir de laquelle les jeux de données rapportés sont établis.

Dans le cas contraire, les producteurs de données (secrétariats techniques de bassin, DEB, etc.) sont alertés des contrôles ayant échoué et il leur est demandé d'apporter soit des corrections, soit des justifications. Les corrections sont toujours réalisées par les producteurs de données.

Exemple : le rapportage des plans de gestion et le principe de non-dégradation

Lors du rapportage sur les plans de gestion de mars 2010, l'un des contrôles métiers a consisté à vérifier que chaque masse d'eau, classée en bon état en 2009, avait aussi un objectif de bon état en 2015 (c'est-à-dire, ne faisait pas l'objet d'une dérogation à l'objectif de bon état) ; il se trouve que ce contrôle a échoué pour plusieurs bassins, ce qui pouvait être une cause de non-conformité, l'un des objectifs de la DCE étant la non-dégradation de l'état.

Après avoir signalé le résultat de ce contrôle, les bassins concernés ont fourni des données corrigées (affectant environ 4% des masses d'eau) et un rapportage correctif a été effectué fin 2010 ; c'est ce rapportage correctif qui a été pris en compte par la Commission européenne pour l'évaluation de nos plans de gestion.

Par ailleurs, la Commission européenne effectue des contrôles similaires et adresse aux États membres ses remarques ; ceux-ci doivent alors apporter soit des corrections, soit des justifications. Pour la France, le ministère de l'écologie demande à l'Onema d'analyser ces remarques, qui sont transmises aux bassins pour correction ou justification. Des rapportages correctifs sont ensuite réalisés.

5.4. Gestion des données

La gestion des données inclut toutes les opérations effectuées sur les données, depuis l'enregistrement dans un système d'information jusqu'à leur diffusion. Pour permettre à un usager des données d'apprécier leur fiabilité, il est impératif que les dispositifs de gestion (banques de données) four-

nissent le moyen de qualifier les données et de tracer les opérations qui sont effectuées (en particulier, de garder la trace du producteur de la donnée).

Il faut d'abord que les dictionnaires de données et les scénarios d'échanges spécifient les éléments appropriés à la conservation et à la transmission de ces informations (par exemple, des attributs pour indiquer le producteur de la donnée et le niveau de qualification), qui dépendent du domaine considéré. Le niveau de qualification prend en compte le niveau des contrôles effectués.

Exemple : les données piézométriques

Le dictionnaire de données permet d'attribuer aux mesures piézométriques un niveau de qualification, qui peut être : « non définissable » (qualification impossible, par manque d'information), « correcte » (tous les contrôles sont réussis), « incorrecte » (certains contrôles ont échoué), « incertaine » (tous les contrôles n'ont pas encore été effectués) et « non qualifiée » (aucun contrôle n'a encore été effectué).

En outre, cette qualification est relative à un état d'avancement de la validation, qui permet de préciser le niveau des contrôles de qualité effectués : « brute » (avant tout contrôle), « contrôlée niveau 1 », « contrôlée niveau 2 » (contrôles plus poussés), « mise en perspective » (donnée utilisée dans un rapport ou valorisée, ce qui constitue une meilleure validation).

Il est ensuite nécessaire de vérifier que les banques de données sont conformes à ces dictionnaires et que les jeux de données les alimentant sont également conformes aux scénarios d'échange. À cette fin, le secrétariat technique du Sandre effectue, à la demande des usagers, des contrôles de conformité (certains étant automatisés) qui conduisent à la délivrance d'un certificat de conformité.

La mise en œuvre du système de gestion de la qualité, en matière de gestion des données, relève des *administrateurs de données* qui, en travaillant avec les experts métier, sont les garants de leur fiabilité. Il est à noter que de plus en plus de routines automatisées permettent d'émettre des alertes (valeur hors champs, valeur déjà existante, champ manquant, etc.) dès la saisie des données.

L'existence d'un administrateur de données compétent, la formation des utilisateurs aux mécanismes de gestion de la qualité et une assistance sont des conditions importantes pour la fiabilité des données. Pour les banques de données existantes, et lors de la mise en place, par l'Onema, de nouvelles bases de données, des formations sont toujours prévues pour leurs utilisateurs.

Les tâches de qualification des données sont essentielles et elles mobilisent des moyens souvent sous-estimés qui rallongent la durée entre la production de la donnée et sa mise à disposition. Il faut ainsi choisir entre mettre à disposition rapidement des données non qualifiées, dont l'utilisateur sera incapable d'apprécier la fiabilité, ou les mettre à disposition au terme d'une étape de qualification, ce qui prend du temps. Le SIE fait le choix de ne diffuser que des données qualifiées.

6. Comparaisons internationales

Les reportages ainsi que les transmissions de données aux réseaux internationaux, comme celui de l'agence européenne de l'environnement (AEE), permettent de disposer de comparaisons européennes sur la fiabilité des données du SIE. La France est généralement bien classée dans ces comparaisons.

Exemple : le partage des données du SIE à l'échelle européenne

Sur les données européennes relatives aux substances dangereuses, dont le bilan vient d'être publié², la France a 86% de données exploitables (c'est-à-dire non exclues par les contrôles de qualité), alors que l'Allemagne n'en a que 35 %, l'Espagne 39%, l'Italie 68% et le Royaume-Uni 69% ; les Pays-Bas font un peu mieux avec 88% de valeurs exploitables, mais c'est la France qui fournit la moitié des données sur les substances dangereuses de toute l'Union européenne.

L'AEE teste depuis peu des remontées de données sur la qualité hydrobiologique (macro-invertébrés, diatomées) des cours d'eau : sur 4688 points de mesure collectés, le tiers provient de la France, qui alimente à hauteur de 20% la base de données.

Plus globalement, le score de la France attribué par l'Agence européenne de l'environnement pour les transmissions de données a doublé en une dizaine d'années : de 47% en 2000 à 94% en 2011.

Les données sont collectées et contrôlées par le Service de l'observation et des statistiques du Commissariat général au développement durable. La structuration des données dans les bassins a nettement progressé dans le cadre du SIE, ce qui a facilité la collecte des données d'année en année. Pour la qualité des cours d'eau, toutes les données de surveillance sont totalement accessibles sans restriction pour tous les bassins de France métropolitaine à l'exception de Seine-Normandie (en cours) et en outre-mer, pour La Réunion, Martinique et Guadeloupe. Des exports multi-données, multi-stations sont de plus possibles en ligne pour les bassins Loire-Bretagne, Adour-Garonne et Rhône-Méditerranée et Corse. Toutes les autres demandes directes de données ont été satisfaites. Toutes les données concernant la qualité des eaux souterraines sont accessibles via le portail ADES.

Les données disponibles ont ainsi permis de maintenir un niveau élevé et de progresser dans les taux de réponse aux collectes de données eau menées par l'agence européenne de l'environnement dans le cadre de WISE (système d'information européen sur l'eau), comme en atteste l'évaluation par l'AEE de la contribution de la France, classée chaque année en quatre niveaux, de un 😞 à trois 😊, de la façon suivante (source : <http://www.eionet.europa.eu/dataflows>) :

| Data flow name | Progress 2000 | Progress 2001 | Progress 2002 | Progress 2003 | Progress 2004 | Progress 2005 | Progress 2006 | Progress 2007 | Progress 2008 | Progress 2009 | Progress 2010 | Progress 2011 |
|-------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| EWN1: River quality | 😊😊😊 | 😊😊😊 | 😊😊😊 | 😊😊 | 😊😊 | 😊😊😊 | 😊😊😊 | 😊😊😊 | 😊😊😊 | 😊😊😊 | 😊😊😊 | 😊😊😊 |
| EWN2: Lake quality | 😊 | 😊 | 😊 | 😊 | 😊 | 😊 | 😞 | 😊 | 😊 | 😊 | 😊😊 | 😊😊😊 |
| EWN3: Groundwater quality | 😞😞 | 😞😞 | 😊😊 | 😊 | 😊😊😊 | 😊😊😊 | 😊😊😊 | 😊😊😊 | 😊😊😊 | 😊😊😊 | 😊😊😊 | 😊😊😊 |
| EWN4: Water quantity | N/A | N/A | N/A | N/A | N/A | 😞 | 😞 | N/A | N/A | N/A | 😊 | 😊 |
| ME1: Marine data | 😊 | 😊 | 😞 | 😊😊😊 | 😞 | 😊😊😊 | 😊😊😊 | 😊😊 | 😊😊😊 | 😊😊😊 | 😊😊😊 | 😊😊😊 |
| WISE1: Water emission quality | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 😊😊 | 😊😊 | 😊😊 |

