

Qualité des données hydrobiologiques et bancarisation Eléments de validation des données « macrophytes en cours d'eau »

S. Lorient, T. Feret, C. Chauvin

Juillet 2016

Document final

Avec le soutien de

Contexte de programmation et de réalisation

Ce rapport a été réalisé dans le cadre du programme scientifique et technique AQUAREF pour l'année 2014, thème E « Garantir la qualité des données bancarisées ».
Achevé en 2016, il reprend l'évolution du contexte et les éléments de connaissance acquis jusqu'en juin 2016.

Auteurs :

Sandrine Lorient
Irstea Bordeaux

Thibaut Feret
Irstea Bordeaux

Christian Chauvin
Irstea Bordeaux
christian.chauvin@irstea.fr

Vérification du document :

Jean-Philippe Ghestem
BRGM
jp.ghestim@brgm.fr

Les correspondants

Onema : Isabelle Barthe-Franquin, isabelle.barthe-franquin@onema.fr

Irstea : Sandrine Lorient, sandrine.lorient@irstea.fr

Référence du document : Sandrine Lorient, Thibaut Feret, Christian Chauvin - Qualité des données hydrobiologiques et bancarisation - Eléments de validation des données « macrophytes en cours d'eau » - Rapport AQUAREF 2016 - 38 p.

Droits d'usage :	<i>Accès libre</i>
Couverture géographique :	<i>National</i>
Niveau géographique :	<i>National</i>
Niveau de lecture :	<i>Professionnels, experts</i>
Nature de la ressource :	<i>Document</i>

RESUME	5
1 Introduction.....	6
2 Nature des données « macrophytes en cours d'eau »	7
2.1 Données générales sur l'opération de contrôle	7
2.2 Description par unité de relevé	8
2.3 Données stationnelles.....	8
2.4 Données floristiques	9
3 Mode actuel de bancarisation des données « macrophytes en cours d'eau »	9
4 Anomalies constatées.....	12
4.1 Types d'anomalies	12
4.2 Données et anomalies associées	12
4.2.1 Données d'identification et de référencement de l'opération de contrôle.....	12
4.2.2 Données de description de végétation.....	15
4.2.3 Données de description stationnelle.....	18
4.2.4 Observations et schéma	22
4.3 Impacts potentiels des anomalies	23
4.3.1 Anomalies fortement impactantes.....	23
4.3.2 Anomalies impactantes	23
4.3.3 Anomalies potentiellement impactantes	24
5 Eléments de contrôle et de validation	25
5.1 Eléments de contrôle automatisé.....	25
5.2 Eléments de contrôle par expertise.....	27
6 Commentaires, perspectives.....	28
7 DOCUMENTS CITES.....	31
8 ANNEXES	31

QUALITE DES DONNEES HYDROBIOLOGIQUES ET BANCARISATION -
ELEMENTS DE VALIDATION DES DONNEES « MACROPHYTES EN COURS D'EAU »
S. Lorient, T. Feret, C. Chauvin

RESUME

Dans le contexte de la mise en œuvre de la « Directive Cadre Européenne sur l'Eau » ou DCE (2000/60/E), Irstea a été en charge de la bancarisation des données macrophytes en cours d'eau de 2006 à 2016, dans le cadre de la convention d'appui avec l'ONEMA. Irstea utilise ces données pour mener à bien les programmes de développement des méthodes d'évaluation de la qualité écologique des cours d'eau à partir des macrophytes. Fort du retour d'expérience de la bancarisation de plus de 5600 relevés IBMR collectés dans les réseaux DCE depuis 2005, Irstea a développé un protocole de contrôle des données avant bancarisation, pour plusieurs des informations composant les résultats de la « mesure macrophytes en cours d'eau ».

La présente note, issue d'un travail réalisé dans le cadre des actions « qualité des données bancarisées » du programme Aquaref, se propose d'explicitier la structure et la nature des données macrophytes en cours d'eau puis de dresser la liste des erreurs et non-conformités mises en évidence lors des contrôles préalables à la bancarisation réalisés par Irstea. Un bilan de l'impact potentiel de ces anomalies, notamment sur l'évaluation de la qualité écologique des masses d'eau, est établi. Enfin des éléments de contrôle et de validation automatisés qu'il serait envisageable de mettre en place sont proposés. Des préconisations sont formalisées pour l'intégration de protocoles de contrôle dans les prétraitements des données de surveillance collectées par les organismes gestionnaires.

Le constat est cependant fait que pour la validation de certains types de données faisant appel à des compétences en taxinomie et en écologie, la validation ne peut être, en l'état actuel des connaissances et des moyens techniques à disposition, réalisée de façon automatique ; l'avis d'expert reste incontournable pour la qualification des données.

Il est également souligné que ces phases de contrôle des données ne doivent pas être sous-évaluées, tant dans le délai de mise à disposition des données exploitables qu'elles imposent que dans leur importance pour assurer une qualité suffisante des informations stockées dans les bases de données.

1 INTRODUCTION

Dans le contexte de la mise en œuvre de la « Directive Cadre Européenne sur l'Eau » ou DCE (2000/60/E), Irstea a été chargé dès 2005 de la constitution de bases de données qui permettaient de structurer les données recueillies par les opérateurs de la surveillance dont les réseaux se mettaient en place. L'objectif initial de cette base de données multi-éléments PANDORE était principalement de constituer une banque de données contrôlées pour alimenter d'une part les travaux de développement des indicateurs hydrobiologiques « cours d'eau » par les équipes d'Irstea, d'autre part les jeux de données nécessaires à l'exercice européen d'intercalibration.

La majeure partie de ces phases de développement pour la mise en œuvre de la DCE étant maintenant achevée et l'exploitation de ces données étant très principalement orientée vers l'évaluation et le diagnostic par les gestionnaires (Agences de l'eau en particulier), la banque de données PANDORE d'Irstea n'est plus alimentée directement pour la plupart des éléments biologiques. Depuis 2012, pour ce qui concerne les cours d'eau (PANDORE n'inclut pas les données « plans d'eau ») Irstea n'assure plus la bancarisation des données issues des réseaux de surveillance que pour l'élément « macrophytes ». L'équipe CARMA-Irstea Bordeaux poursuivant le développement des méthodes « macrophytes cours d'eau », elle a à ce titre assuré la bancarisation des données pour cet élément biologique jusqu'en 2016. PANDORE intègre les données acquises suivant le protocole normalisé de l'IBMR (Norme AFNOR NF T90-395), complété par les préconisations qui ont été validées par les instances de pilotage des réseaux DCE, et mises en œuvre par les Agences de l'eau, maîtres d'œuvre quasi exclusif de la surveillance « cours d'eau ». Ces données proviennent des mesures réalisées depuis 2005 dans les réseaux DCE : réseau de référence (RCR), réseau de contrôle et de surveillance (RCS), stations des premiers programmes du réseau de contrôle opérationnel (RCO), ainsi que des sites de réseaux antérieurs (RNB). A cela, s'ajoute quelques données issues d'études ponctuelles menées localement par des Agences de l'eau (principalement pour les macrophytes, EQB pour lequel il n'existait pas de données de réseaux antérieurs).

La bancarisation de plus de 5600 relevés IBMR collectés dans les réseaux DCE depuis 2005 sur la base d'un protocole développé par Irstea, ainsi que leur exploitation depuis 2010 pour le développement de la méthode d'évaluation¹, a permis de développer une approche globale de contrôle des données recueillies sur un certain nombre de points. Ces contrôles ont mis en évidence plusieurs types d'erreurs et de non-conformités dans les données remontées par les opérateurs. Un retour d'expérience est donc maintenant disponible sur la qualification des données « macrophytes en cours d'eau ».

Ce travail de formalisation des contrôles de pré-bancarisation appliqués et des pistes possibles pour les améliorer et les compléter a été réalisé dans le cadre des actions du programme Aquaref (thème E : qualité des données bancarisées). Il vise principalement à apporter une aide aux organismes qui reprennent en charge la bancarisation des données « macrophytes cours d'eau ».

Après avoir rappelé la structure et la nature des données macrophytes en cours d'eau et explicité les items qui figurent dans les modèles actuellement utilisés pour la collecte des données de surveillance et sur lesquels s'appuient les critères de contrôle, une liste des erreurs et non-conformités mises en évidence lors des contrôles de pré-bancarisation réalisés par Irstea est dressée. Un bilan sur les impacts de ces anomalies et leur degré est établi. Enfin des éléments de contrôle et de validation complémentaires sont proposés pour contribuer à l'évolution des systèmes de contrôle des données et à l'amélioration de leur qualité.

¹ Chauvin *et al.*, 2013 : Méthode d'évaluation de la qualité écologique des cours d'eau.

Loriot *et al.*, 2014 : travaux sur les traits écologiques des taxons et la recherche de métriques de diagnostic.

2 NATURE DES DONNÉES « MACROPHYTES EN COURS D'EAU »

La collecte des données « macrophytes en cours d'eau » issues des réseaux DCE résulte de la mise en application de la norme AFNOR NF T 90-395 (Octobre 2003) et les documents d'application pour les réseaux DCE. Elles comprennent des données générales sur l'opération de contrôle², des données floristiques et des données soutenant la biologie. Une des caractéristiques des relevés « macrophytes en cours d'eau », et ce suivant les recommandations de la norme, réside dans le fait que les données floristiques et stationnelles sont à relever en distinguant d'une part les faciès les plus lents, d'autres parts les faciès les plus rapides. Les fiches sont conçues pour que les données soient saisies par regroupement, appelés « unité de relevé », de ces faciès les plus lents et les plus rapides.

Les documents de référence concernant les modalités techniques et procédurales de la collecte des données hydrobiologiques « macrophytes en cours d'eau » sont donc :

- la norme AFNOR T90-395 « Qualité de l'eau – Détermination de l'indice biologique macrophytique en rivière (IBMR) »
- l'arrêté de 2011 portant modalités de l'agrément des laboratoires,
- les modèles de relevé stationnel et de transmission des données préconisés par l'arrêté « surveillance » dès sa version de janvier 2010, diffusés via le site web <http://hydrobio-dce.irstea.fr>

Lors des réflexions menées dans le cadre de l'élaboration du projet de base de données nationale Naïades, l'ensemble des items renseignés dans les fiches de terrain puis reportés dans les fichiers de transmission des résultats standardisés a fait l'objet d'une homogénéisation avec les référentiels du SANDRE, afin de correspondre à ces dictionnaires et d'aligner le type d'information sur les autres méthodes hydrobiologiques utilisant les mêmes descripteurs, lorsque cela était pertinent.

2.1 Données générales sur l'opération de contrôle

L'opérateur doit renseigner les informations générales qui permettent d'identifier la station et l'opération de contrôle qu'il a effectué :

- le nom du cours d'eau,
- le nom de la station,
- le code Sandre³ de la station (il s'agit d'une référence essentielle qui identifie de façon unique le couple cours d'eau / station),
- la date
- les coordonnées géographiques du point de prélèvement (limite amont de la section de relevé de 100 m),
- le nom de l'opérateur ayant réalisé la mesure, en complément de l'identification de l'organisme opérateur (au titre des obligations liées à l'agrément).

Parmi ces éléments, la date et le code SANDRE de la station sont des paramètres obligatoires, car ils constituent le couple d'identification de l'opération de contrôle dans la base de données.

Les coordonnées géographiques sont également obligatoires, à la fois pour contrôler l'emplacement du relevé et pour assurer le suivi des mesures par le gestionnaire de la station sur le long terme.

² L'« opération de contrôle » est le terme validé dans le dictionnaire SANDRE pour désigner un relevé. Il est caractérisé par un code station et une date. C'est la clé primaire utilisée dans la base de données PANDORE.

³ Le Service d'Administration Nationale des Données et Référentiels sur l'Eau, SANDRE, établit et met à disposition le référentiel des données sur l'eau du Système d'Information sur l'Eau : dictionnaires de données, scénarios d'échange, données de référence, etc. www.sandre.eaufrance.fr

2.2 Description par unité de relevé

La norme AFNOR précise que le relevé doit être réalisé selon deux entités, appelées Unités de relevés (UR). L'UR 1 est définie comme le regroupement des faciès d'écoulement les plus rapides de la station (ou comme le faciès unique s'il n'y en a qu'un seul), l'UR 2 comme celui regroupant les faciès les plus lents. La norme actuelle n'impose pas de relevé floristique distinct mais le préconise pour la richesse d'informations apportée pour l'interprétation des résultats et la possibilité de calculer des métriques pour les deux types d'habitats. Les préconisations pour la surveillance DCE reprennent ces recommandations en les rendant obligatoires. Il importe donc de renseigner le nombre d'unités de relevé observées sur le point de prélèvement : de 2 par défaut, il peut être réduit à 1 si la morphologie est très homogène et ne permet pas de distinguer 2 unités de relevés.

2.3 Données stationnelles

La norme elle-même n'impose pas le relevé mésologique. Cependant, l'application de la méthode IBMR dans les réseaux nationaux de mesure DCE impose la description des paramètres environnementaux « soutenant la biologie ». La description du point de prélèvement d'un point de vue hydromorphologique est ainsi préconisée, selon un protocole précisé par les fiches de relevé, afin de disposer des informations nécessaires à l'interprétation écologique des résultats d'évaluation.

Les paramètres soutenant la biologie considérés pour les relevés « macrophytes en cours d'eau » comprennent :

- les conditions météorologiques et hydrologiques dans lesquelles a été effectué le relevé (influence sur la qualité et la complétude du relevé),
- une description des faciès hydromorphologiques identifiés (profondeurs, vitesses d'écoulement, l'éclairement du lit et nature des substrats),
- un schéma d'ensemble du point de prélèvement mentionnant la répartition des unités de relevés, les principales composantes des populations végétales, les caractéristiques particulières, l'occupation des berges, les points de repérages des limites du point de prélèvement, l'orientation du cours d'eau et tout repère le long du point de prélèvement qui permet de se repérer. Le schéma peut également comporter des observations liées aux conditions hydrologiques, à la présence d'embâcles, d'espèces végétales exotiques envahissantes, à un aménagement etc.

Cette dernière partie cartographique de la description stationnelle est destinée à la fois à la validation du relevé (part des UR, localisation, etc.) ainsi qu'au suivi des conditions de réalisation des relevés sur le long terme. Elle fournit également des éléments explicatifs de certaines modifications dans les relevés (évolution des faciès d'écoulement selon le débit, par exemple). Son format informatique est généralement de type « image ». Le schéma n'est pas bancarisé dans les bases de données uniquement alphanumériques, mais conservé et utilisé par le « responsable de la données » (généralement les Agences de l'eau » ou le gestionnaire de la station de surveillance (DREAL, par exemple).

2.4 Données floristiques

Conformément au protocole normalisé, le relevé doit prendre en compte l'ensemble des végétaux macrophytiques aquatiques : Phanérogames, Ptéridophytes, Bryophytes (mousses et hépatiques), Lichens, Hétérotrophes (bactéries et champignons en colonies filamenteuses), les Algues macroscopiques ou en colonies macroscopiques. Les ligneux (arbustes et arbres) ne correspondent pas à la définition méthodologique et normative de « macrophytes » qui est considérée dans le protocole IBMR, même s'ils sont observés à l'intérieur des limites de la surface en eau.

On considère comme aquatique toute plante croissant de façon naturelle dans l'eau ayant au moins la base de son appareil végétatif immergée dans des conditions d'étiage (définies comme les conditions de débit proches du QMNA₅⁴). Tous les taxons macrophytiques rencontrés dans l'eau sont inventoriés, qu'ils soient ou non contributifs à l'IBMR. A la fin du relevé, la liste des taxons doit être impérativement complète même s'il reste des déterminations à préciser ou à confirmer au laboratoire.

Le niveau de détermination pour les Phanérogames, les Bryophytes, les Lichens et les algues Characées est celui de l'espèce. Pour les organismes unicellulaires (Algues et Hétérotrophes), le niveau retenu est celui du genre ; ceci sauf exception d'un taxon contributif à un niveau de détermination différent comme par exemple l'algue *Stigeoclonium tenue* ou le phanérogame *Ranunculus penicillatus* var. *calcareus*.

Le recouvrement de chaque taxon doit être estimé en distinguant les unités de relevé. Bien que la norme actuelle considère l'évaluation du recouvrement soit en classes (pour le calcul direct de l'IBMR) soit en pourcentage de la surface totale en eau de l'UR, c'est la deuxième option (pourcentage) qui est imposée dans les mesures en réseaux DCE.

3 MODE ACTUEL DE BANCARISATION DES DONNÉES « MACROPHYTES EN COURS D'EAU »

Rappelons que, jusqu'en 2015 et pour ce qui concerne les macrophytes en cours d'eau, c'est la base de données Irstea-PANDORE qui alimentait le SEEE (Système d'évaluation de l'état écologique des eaux, géré par l'ONEMA) utilisés pour l'évaluation et le rapportage DCE.

Pour les données « macrophytes en cours d'eau », Irstea a développé des applications sous MS-Excel permettant de standardiser les données collectées (fiche terrain), de donner un cadre à leur saisie (fiches de saisie) et d'appliquer des premiers contrôles à la saisie. L'utilisation de menus déroulants permet de renseigner des champs de données qualitatifs (*i.e.* conditions météorologiques, classe d'abondance, nom des taxons). Des macro-commandes programmées en Visual Basic (VBA) permettent le calcul de la note IBMR et réalisent divers contrôles de saisie et des validations. L'utilisation de ces formulaires lors de la transmission des relevés par les opérateurs permet de faciliter l'intégration des données dans les bases d'Irstea et de réduire, dans une certaine mesure, les anomalies dans les données.

⁴ QMNA₅ : débit moyen mensuel minimum de fréquence quinquennale

Pour la saisie des données « macrophytes en cours d'eau », deux fiches de saisie sont à compléter. Une première fiche permet de saisir les données floristiques et de calculer la note IBMR. Une seconde fiche permet de renseigner les données stationnelles c'est à dire un ensemble de paramètres environnementaux. Ces deux fiches comprennent un en-tête commun dans lequel sont à saisir les identifiants de l'opération de contrôle.

Les fonctions de calcul des indicateurs sont proposées dans ces modèles comme aide à l'opérateur. Seules les données brutes (liste floristique et recouvrement des taxons par UR) sont extraites pour être bancarisées dans PANDORE.

Pour la bancarisation sous PANDORE par Irstea des données « macrophytes cours d'eau », la collecte des données a évolué :

- De 2006 à 2014 : fichiers Excel adressés directement à Irstea par les producteurs de données ou les opérateurs ;
- De 2014 à 2015 : fichiers versés dans un « entrepôt » SEEE (ftp) géré par l'ONEMA, dans lequel Irstea les récupérait,
- A partir de 2016 : les données ne sont plus ni transmises à Irstea ni versées dans un ftp de l'ONEMA. Les Agences récoltent et gèrent les données de leurs marchés dans leurs propres bases de données.

Jusqu'en 2016 (ou partiellement en 2015), les données collectées par Irstea étaient celles qui étaient transmises par les « responsables des données » (principalement les donneurs d'ordre de la surveillance, c'est-à-dire les Agences de l'eau). Elles étaient donc réputées déjà validées, soit par les opérateurs eux-mêmes dans le cadre de leurs procédures Qualité, soit par les Agences de l'eau en tant que contrôleurs du résultat des marchés.

Lorsque Irstea entame le processus de bancarisation des données dans sa base Pandore, un contrôle dit de pré-bancarisation est effectué afin de détecter les éventuelles anomalies résiduelles c'est-à-dire les données manquantes, les non-conformités entre la fiche stationnelle et la fiche floristique et enfin les données aberrantes. Un rapport de contrôle pré-bancarisation est alors produit et transmis aux Agences afin de les informer que la bancarisation de certaines données requiert des corrections, précisions ou compléments. Les Agences consultent alors leurs opérateurs. Les données revues sont retournées à Irstea par les Agences ou directement par les opérateurs. Le contrôle de pré-bancarisation est de nouveau effectué. Si nécessaire, un second retour vers les Agences est effectué. Dans certains cas, rares mais réguliers, les demandes de corrections et de compléments restent sans réponse. Irstea informe alors les producteurs et opérateurs concernés que dans l'état actuel, les relevés concernés ne sont pas bancarisés faute de compléments.

Ce processus de validation des données par Irstea en collaboration avec les opérateurs via les Agences est long et augmente fortement (de l'ordre de 300%) le temps consacré à la bancarisation des données. La figure 1 présente le cheminement des données depuis leur transmission à Irstea jusqu'à leur bancarisation.

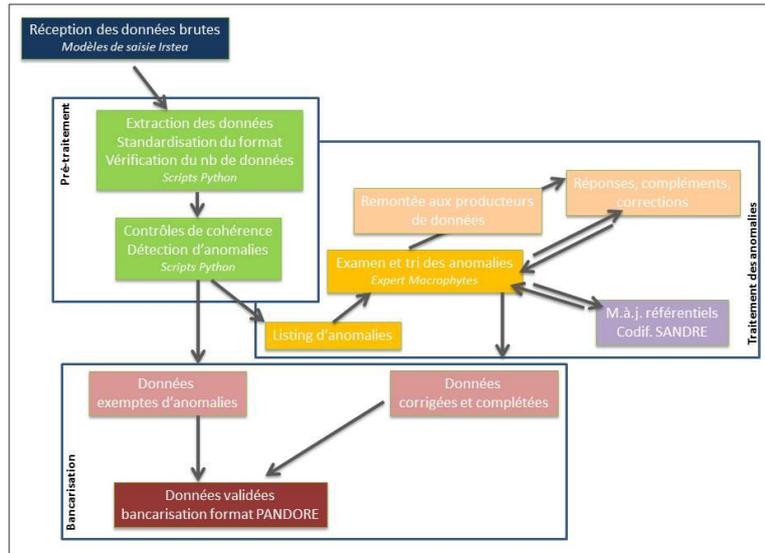


Figure 1 : circuit de validation des données (Irstea, 2007 à 2016).

Pour autant, ce travail est nécessaire pour garantir la qualification des données et donc la robustesse de l'évaluation de la qualité écologique par les données « macrophytes en cours d'eau ». De plus, seules des données complètes et qualifiées peuvent être utilisées pour le développement de méthodes comme pour l'évaluation destinée au rapportage ou au diagnostic.

Ce sont ainsi des données complètes, conformes à la norme AFNOR et à ses documents d'application qui sont bancairisées par Irstea après un processus de qualification qui se veut le plus collaboratif et transparent possible auprès des partenaires, Agences, DREAL et bureaux d'Etude.

Le calendrier de ce cycle annuel de collecte (fig. 2), contrôle et bancairisation s'étend globalement sur 18 mois. Ces délais sont principalement conditionnés par le temps nécessaire pour la collecte des données, très variable selon l'organisation et l'implication des acteurs. Des disparités notables sont observées, avec un envoi de l'ensemble des données validées dès le mois de mars N+1 pour certaines Agences, jusqu'à des échanges encore effectifs en septembre N+1 pour d'autres.

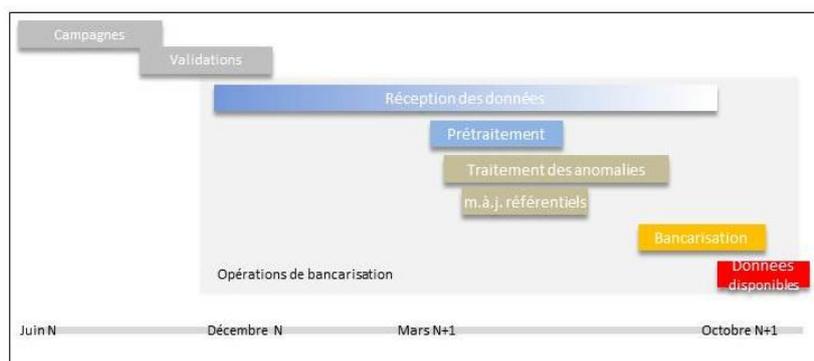


Figure 2 : Planning général du cycle annuel de collecte, contrôle et bancairisation des données de surveillance « macrophytes cours d'eau ». Les délais sont ceux qui ont été constatés sur plusieurs années, pour les actions de bancairisation menées par Irstea. Ils recouvrent une réalité très variable selon les sources (responsables et producteurs de données).

La deuxième phase conditionnant fortement ces délais est constituée par les échanges nécessaires pour corriger et compléter les données, sur la base des listings d'anomalies. Dans certains cas, ces échanges peuvent s'étendre au-delà de la bancairisation de l'ensemble des lots, et nécessite ensuite

un complément de bancarisation. Tous les ans, quelques données ne peuvent d'ailleurs pas être complétées, car les échanges et relances des opérateurs ne débouchent pas sur des réponses utilisables. Ces résultats constituent un « résidu » non bancarisable, car non conforme aux critères obligatoires.

4 ANOMALIES CONSTATÉES

4.1 Types d'anomalies

Les anomalies mises en évidence par les contrôles de pré-bancarisation peuvent être classées en cinq catégories :

- les erreurs de saisie : fautes de frappe dans la saisie des noms et des chiffres ;
- les incohérences entre fiche stationnelle et fiche floristique ;
- les informations manquantes : champs non remplis volontairement car l'opérateur a jugé que l'information n'était pas importante ou bien champs dont la saisie a été oubliée ;
- les aberrations : des données qui ne sont mathématiquement ou écologiquement pas possibles ;
- les données non-conformes : des données saisies qui ne sont pas correctes eut égard aux référentiels « Station de mesure » et « Appellation de taxon » du Sandre et au référentiel taxinomique d'Irstea.

Rappelons que le type d'information attendu pour chaque item des fiches de description et de relevé est explicité dans la notice incluse dans le modèle de saisie Excel et dans la notice d'utilisation des fiches de terrain.

4.2 Données et anomalies associées

Les données et les anomalies associées sont présentées ci-dessous en distinguant les données à renseigner conformément aux différents textes d'encadrement de la production des données de surveillance (norme NF T90-395, GA provisoire, arrêté « agrément des laboratoires », arrêté « Surveillance »).

Les intitulés des champs à remplir dans les fiches de saisies et présents dans la base de données Pandore sont utilisés. Pour chaque donnée, les informations suivantes sont renseignées : son format, une description de sa nature et une liste des anomalies mises en évidence par les contrôles pré-bancarisation d'Irstea.

4.2.1 Données d'identification et de référencement de l'opération de contrôle

Organisme

- Format : Texte.
- Description : identification de la structure, généralement établissement public de type DREAL ou bureau d'étude qui réalise le relevé. Il peut éventuellement s'agir de sigles. Cette donnée est à renseigner conformément à l'Article 3, volet II/3d de l'Arrêté Ministériel « Agrément » du 27 octobre 2011 portant sur les modalités d'agrément des laboratoires effectuant des analyses dans le domaine de l'eau et des milieux aquatiques au titre du code de l'environnement.
- Anomalies : donnée absente ; non concordance entre les fiches stationnelle et floristique.

Opérateurs

- Format : Texte.
- Description : nom en clair (pas uniquement les initiales) et éventuellement prénoms des personnes ayant réalisé le relevé. Cette donnée est à renseigner conformément à l'Article 3, volet II/3d de l'Arrêté Ministériel « Agrément » du 27 octobre 2011 portant sur les modalités d'agrément des laboratoires effectuant des analyses dans le domaine de l'eau et des milieux aquatiques au titre du code de l'environnement. Il est intéressant de préciser dans le champ « observations » qui a réalisé le relevé de terrain et qui s'est occupé de la phase laboratoire s'il s'agit de personnes différentes.
- Anomalies : donnée absente ; non concordance entre les fiches stationnelle et floristique.

Nom du cours d'eau

- Format : Texte.
- Description : nom du cours d'eau communément admis (cartes IGN 1/25 000ème ou BD Carthage). C'est le nom présent dans le référentiel « Cours d'eau » du Sandre.
- Anomalies : donnée absente ; donnée différente de celle du référentiel « Cours d'Eau » du Sandre.

Nom de la station

- Format : Texte.
- Description : nom de la station de mesure présente dans le référentiel « Station de mesure de la qualité des eaux superficielles continentales » du Sandre. Ce nom comprend généralement le nom du cours d'eau suivi du nom de la commune sur laquelle se localise la station, ex : « Elorn à Sizun », auxquels peuvent s'ajouter des précisions d'ordre géographique, ex : « Bras de l'Isle à Saint-Seurin-sur-l'Isle (amont prise d'eau) » ou encore « Le Basarun Erreka à Saint Jean de Luz (Lieu-dit Xibau Berri) ».
- Anomalies : donnée absente ; donnée non conforme aux dénominations du référentiel « Station de mesure de la qualité des eaux superficielles continentales » ; non concordance entre fiches stationnelle et floristique ; plus rarement, une anomalie qui n'est pas une erreur apparaît quand une station fait l'objet d'un prélèvement alors qu'elle n'a pas encore été intégrée dans le référentiel du Sandre.

Code Sandre de la station

- Format : Numérique (8 chiffres).
- Description : les codes station correspondent au référentiel de codification nationale du Sandre. Ils sont produits par les Agences de l'Eau. Il s'agit d'une référence essentielle qui identifie de façon unique le couple « cours d'eau / station » dans une banque de données. Il s'agit d'un numéro d'ordre national à 8 chiffres dont les deux premiers (commençant par 0) désignent un bassin hydrographique.
- Remarque : lors du passage des données par un format Excel, les données au format « nombre » ou « standard » qui commencent par « 0 » sont automatiquement amputées de ce premier chiffre. Cela peut induire des anomalies de format du code station.
- Anomalies : donnée absente ; faute de frappe engendrant un chiffre manquant ou erroné ; non concordance entre le nom de station et le code station relativement au référentiel Sandre « Stations de mesure de la qualité des eaux superficielles continentales » ; non concordance entre fiches stationnelle et floristique. Régulièrement les codes stations saisis ne sont pas au format « Sandre » avec la suppression des 2 premiers chiffres qui identifient le bassin hydrographique (01, 02, etc.) ; plus rarement, une anomalie qui n'est pas une erreur apparaît quand une station fait l'objet d'un prélèvement alors qu'elle n'a pas encore été intégrée dans le référentiel du Sandre.

Date

- Format : Date (jour – mois – année).
- Description : date à laquelle le relevé a été effectué. Lorsque les relevés ne peuvent pas être réalisés sur une seule journée, ils doivent l'être, au maximum, sur deux jours consécutifs. Dans le cas où un relevé est réparti sur deux journées (par exemple relevé stationnel le jour 1 et relevé floristique le jour 2), la date notée sera celle du relevé floristique.
- Anomalies : donnée absente ; donnée incohérente (ex : 2025 ; 1985) ; décalage de plus d'une journée entre la date du relevé des données stationnelles et celle des données floristiques sans précision dans le champ « observations ».

Coordonnées géographiques

- Format : Numérique, exprimé dans le système Lambert 93.
- Description : coordonnées relevées en limite amont du point de prélèvement. Indispensables à l'identification de chaque opération de contrôle pour le positionnement géoréférencé du relevé dans les SIG, l'affectation des paramètres environnementaux à un point de prélèvement, le contrôle de validité des positionnements, le rattachement des opérations de prélèvements de différents éléments biologiques à une station, etc.
- Anomalies : non concordance entre les coordonnées du point de prélèvement et les coordonnées de la station dans le référentiel « Station de mesure de la qualité des eaux superficielles continentales » du Sandre. Les écarts observés peuvent être importants (5 km à 10 000 km). Il s'agit dans les cas extrêmes d'erreur de saisie, de saisie dans un système différent du Lambert 93. Il s'agit très souvent d'un décalage du point de prélèvement par rapport à la position géographique de la station telle que référencée dans le Sandre. Ceci est fréquent. Les opérateurs de terrain ayant eu la charge de fixer les points de prélèvements pour chaque EQB ont en effet eu la possibilité de retenir l'endroit le plus favorable selon eux.

Rive

- Format : Liste (menu déroulant : droite – gauche – centre).
- Description : indique si les coordonnées du point de prélèvement ont été prises en rive gauche, rive droite ou au centre du cours d'eau. Ceci n'a pas de conséquence dans le cas des petits cours d'eau mais s'avère important dans le cas des grands cours d'eau (largeur de plus de 100 m) pour un positionnement correct de la limite du point de prélèvement.
- Anomalie : donnée absente.

Référence du dossier

- Format : Texte.
- Description : champ réservé à l'opérateur réalisant l'opération de contrôle. La donnée est destinée à identifier le dossier auquel elle se rapporte. Le mode d'identification et l'utilisation sont libres.
- Remarque : donnée non obligatoire pour la réception des données ni requise pour la qualification et l'exploitation des données.
- Anomalies : NULL⁵.

⁵ Dans la base de données PADORE, sous moteur MySQL, une donnée absente renvoie le code « NULL »

4.2.2 Données de description de végétation

Nombre d'unité de relevé

- Format : Numérique (Chiffre « 1 » ou « 2 »).
- Description : L'unité de relevé (UR) est un ensemble de faciès regroupés selon leur caractère « lent » et « rapide ». Cette distinction est relative aux conditions locales : l'unité de relevé « la plus rapide » regroupera les faciès les plus rapides du point de prélèvement, l'unité « la plus lente » regroupera les faciès les plus lents. Pour les grands cours d'eau, il est possible d'utiliser les deux UR pour distinguer les couples rive droite / rive gauche ou bien chenal / berges. Il s'agit donc d'indiquer le nombre d'unité de relevé identifié lors de la réalisation du point de prélèvement. Ce nombre est de « 2 » par défaut. Il peut être réduit à « 1 » si la morphologie du tronçon de cours d'eau analysé est très homogène et ne permet pas de distinguer deux unités de relevés.
- Anomalies : donnée absente ; données aberrantes car non concordance entre le nombre d'UR indiqué et le nombre d'UR effectivement décrits ; données aberrantes car non concordance entre fiches stationnelle et floristique.

Pourcentage de recouvrement de chacune des unités de relevé

- Format : Numérique (en %).
- Description : Pourcentage de recouvrement de chacune des unités de relevé identifiée (1 ou 2). C'est la contribution de chaque unité de relevé sur l'ensemble de la surface du point de prélèvement. Ce renseignement est indispensable au calcul de la note IBMR (pondération des recouvrements des taxons par UR). Dans le cas où une seule unité de relevé est observée, la valeur est obligatoirement de 100%. Dans le cas où 2 unités de relevés sont distinguées, la somme des valeurs de recouvrement doit être de 100%.
- Anomalies : donnée absente ; données aberrantes car non concordance entre fiches stationnelle et floristique ; donnée aberrante quand le relevé porte sur une seule unité de relevé et que le pourcentage de recouvrement est différent de 100% ; donnée aberrante quand le relevé porte sur deux unités de relevé et que la somme du recouvrement des deux unités de relevés est différente de 100% ; données aberrantes quand le recouvrement d'une unité de relevé est 0% alors que des taxons sont notés sur cet UR.

Liste floristique

- Format : Texte et Numérique.
- Description : La liste floristique est la liste résultant de l'identification et de l'estimation des recouvrements (estimation de la proportion d'occupation surfacique) de chaque appellation de taxon sur la zone prélevée. C'est cette liste assortie des abondances de chaque taxon (les recouvrements dans le cas des macrophytes) qui constitue le « résultat de la mesure hydrobiologique », au sens de la DCE. Il est possible que dans certaines stations, aucune végétation ne soit détectée et que la liste floristique soit vide. Il convient alors de l'expliquer dans le champ « Observations ». Dans cette liste floristique, les Phanérogames, Bryophytes et Ptéridophytes sont à renseigner à l'espèce, les Algues et les Hétérotrophes au genre (sauf si il s'agit d'un taxon contributif présent dans la norme). Les ligneux ne doivent pas figurer dans la liste. Les taxons non contributifs, bien que n'ayant pas d'influence sur la valeur de la note IBMR, doivent impérativement figurer sur la liste floristique. Ils sont en effet une indication importante pour l'interprétation des résultats, et leur compilation doit permettre à terme de préciser le caractère bioindicateur des taxons contributifs et l'évolution de la liste de référence normalisée. Ils sont aussi le support du calcul de certaines métriques indicatrices (ratio contributifs/totaux, par exemple). De plus, les protocoles actuels d'acquisition des données biologiques dans les réseaux de surveillance nécessitent d'établir un inventaire complet sur lesquels seront calculer plusieurs métriques et indices pouvant par exemple rendre compte de la biodiversité. Bien que la norme « IBMR » actuelle focalise sur le calcul de l'indicateur, ce référentiel est principalement utilisé dans les applications DCE en tant que protocole d'acquisition de données.

La liste floristique est saisie en utilisant le menu déroulant qui permet de sélectionner le nom latin des taxons dans le référentiel mis à disposition par Irstea dans la fiche de saisie des données floristiques. Une autre possibilité pour l'opérateur est de saisir manuellement le nom du taxon ou son code 6L. Si le taxon ainsi saisi contient une faute de frappe ou est inconnu, un message en informe l'opérateur. Si le taxon a déjà été saisi dans la liste, l'opérateur est également informé par un message d'erreur.

- Anomalies : donnée absente sans explication dans le champ « Observations » ; données aberrantes : seuls les taxons contributifs sont assortis de leur recouvrement ; présence de ligneux, de taxons identifiés au niveau taxinomique de la famille ; présence dans la liste d'hybrides non officiellement reconnus par les référentiels nationaux (la notion d'hybride est parfois utilisée à tort pour exprimer qu'un taxon a été observé avec des traits rappelant morphologiquement deux taxons différents sans que l'opérateur soit parvenu à trancher).

Confer (cf.)

- Format : Texte (cf.)
- Description : permet d'indiquer que la détermination est associée à un faible degré de confiance. Cette information est conservée dans l'optique d'une éventuelle confirmation ultérieure ou dans celle d'une évolution des connaissances taxinomiques.
- Remarque : la mention Confer n'est pas bancarisée. Le taxon associé de la mention cf. est pris en compte dans la liste floristique et bancarisé. S'il s'agit d'un taxon contributif, il est pris en compte dans le calcul de l'indice malgré la mention cf.
- Anomalie : mention Confer utilisée avec le commentaire « en attente d'une validation taxinomique » par un expert sans que le résultat de l'expertise soit donné.

Code alternatif des taxons

- Format : Texte (6 lettres).
- Description : Le code alternatif désigne un code différent du code Sandre qui est utilisé dans certains référentiels taxinomiques majeurs pour décrire les appellations de taxon. Aussi parfois appelé « code métier », il s'agit le plus souvent d'une codification qui est entrée dans les mœurs par les opérateurs de terrain, gestionnaires de données et personnes en charge de l'exploitation des données. Ce sont des codes plus explicites que les codes numériques du Sandre. Pour les macrophytes en cours d'eau, le code alternatif est un code en 6 lettres (code 6L), qui suit des règles de construction précises qui garantissent sa cohérence et son exclusivité.

Le codage 6L pour les macrophytes : De façon générale, les 3 premières lettres correspondent au nom de genre et les 3 dernières au nom de l'espèce soit « GENESP » (ex : *Apium nodiflorum*, APINOD). Pour les taxons renseignés au genre (Algues et hétérotrophes), le code 6L est sur le modèle GENSPX (ex : *Vaucheria* sp., VAUSPX). Pour les sous-espèces, variétés, formes, le code 6L se construit en prenant les 3 premières lettres du genre suivi des 2 premières lettres du nom d'espèce et de la première lettre du nom de sous-espèce, variété ou forme (ex : *Nuphar lutea* var. *pumila*, NUPLUP). Pour les hybrides reconnus dans les référentiels taxinomiques nationaux et internationaux (et non un taxon pour lequel l'opérateur hésite entre deux appellations), le code 6L se construit en prenant les 3 premières lettres du genre, un X puis les 2 premières lettres du nom d'espèce de l'hybride (ex : *Potamogeton x nitens*, POTXNI) ou bien lorsqu'il s'agit d'un hybride qui n'a pas de nom propre mais dont le nom est composé du nom des deux taxons qui s'hybrident, le code 6L se construit en prenant les 3 premières lettres du genre suivis de la première lettre du nom de l'espèce 1, un X, et la première lettre de l'espèce 2 (ex : *Potamogeton lucens x perfoliatus*, POTLXP).

L'attribution d'un code 6L pour les macrophytes en cours d'eau ne doit pas générer de redondance. Deux genres différents ne peuvent pas présenter les mêmes 3 premières lettres. Si les 3 premières lettres sont déjà prises, il faut utiliser les 2 premières et la première 3^{ème} lettre disponible (ex : *Hydrocharis* sp., HYDSPX ; *Hydrocotyle* sp., HYRSPX ; *Hydrurus* sp., HYUSPX ; etc.).

A chaque code Sandre doit correspondre un seul et unique code alternatif. Afin de rapprocher le référentiel national des taxons (Sandre) à celui des banques de données (code alternatif), il est nécessaire de pouvoir établir des correspondances dans les référentiels respectifs (code alternatif mentionné dans le référentiel « Appellations de taxon » du Sandre et code Sandre mentionné dans les banques de données).

Dans le cadre de l'appui à la gestion des référentiels « métiers » pour les méthodes hydrobiologiques et des référentiels Taxons du SANDRE, Irstea est en charge de la déclinaison des codes alternatifs au fur et à mesure de l'observation de nouveau taxon.

- Anomalies : donnée aberrante quand le code 6L saisi est inconnu avec ou sans précision du nom du taxon que l'opérateur à chercher à saisir.

Nom latin des taxons

- Format : Texte.
- Description : nom latin valide des taxons figurant dans le menu déroulant du référentiel fourni par Irstea dans la fiche floristique ou bien nom latin d'un nouveau taxon ne figurant pas dans le référentiel mais reconnu par les référentiels taxinomiques nationaux et internationaux.
- Anomalies : donnée aberrante quand erreur de saisie dans le nom d'un taxon du référentiel Irstea ou d'un nouveau taxon. En principe, l'utilisation du modèle de saisie Excel Irstea garantit la conformité des appellations de taxons, par automatisation de ce renseignement d'après le code 6L saisi.

Code Sandre des taxons

- Format : Numérique.
- Description : Le Sandre établit le référentiel des données sur l'Eau du Système d'Information sur l'Eau. Ce référentiel comprend notamment les listes de codes qui permettent l'échanges de données sur l'eau à l'échelle de la France. Ces référentiels visent à harmoniser et standardiser la dénomination et l'utilisation des données par le codage de toutes les méthodes, paramètres, unités, opérateurs, stations, etc., dans tous les domaines de la surveillance des milieux aquatiques. S'agissant des taxons, le code d'appellation de taxon est un numéro, non signifiant (simple incrémentation automatique au fil des demandes) donné par le Sandre à chacun des noms de taxons qui émergent de la production des données et susceptibles d'être échangés. Ce système de code est associé à une description du taxon en question et permet à différents acteurs de l'eau de communiquer en étant certains de parler de la même entité. (ex : *Ctenophora* sp. peut à la fois désigner une Diatomée ou bien un Invertébré mais ils disposent chacun de leur propre code Sandre, 4169 pour *Ctenophra* sp.– Diatomées et 9525 pour *Ctenophora* sp.- Invertébrés).
- Anomalies : le code Sandre des taxons est incrémenté automatiquement lors de la saisie d'un code 6L ou d'un nom de taxon figurant dans le menu déroulant. Il n'y a donc pas d'erreur possible s'agissant du code Sandre des taxons, sauf si l'opérateur saisit un nouveau taxon et renseigne manuellement un code Sandre erroné. Des anomalies peuvent toutefois apparaître si des codes non mis à jour ou gelés sont utilisés (problèmes de version des référentiels).

Pourcentage de recouvrement des taxons

- Format : Numérique (en %).
- Description : Le recouvrement doit être noté en pourcentage sur le modèle de relevé floristique Irstea. Le recouvrement des taxons doit être renseigné pour chaque unité de relevé où il a été observé.
- Remarque : La somme des recouvrements des taxons par unité de relevé peut être supérieure à 100% (forts recouvrements avec stratification de la végétation, par exemple).
- Anomalies : donnée absente ; données aberrantes : égale à 0% ou supérieur au recouvrement total par la végétation (information « végétalisation » de la description stationnelle) pour un taxon dans une unité de relevé.

Nom d'un nouveau taxon

- Format : Texte.
- Description : si un taxon n'est pas présent dans le menu déroulant à disposition dans la fiche floristique, le modèle de saisie Excel prévoit la possibilité d'ajouter le nom de ce taxon par une saisie manuelle. La saisie du code 6L « NEWCOD » ouvre un espace de saisie du nom scientifique dans la colonne « nouveaux taxa hors liste de référence ». Un code Sandre peut également être saisi manuellement.
- Remarque : cette possibilité d'ajout ne modifie pas le référentiel taxinomique intégré au modèle de saisie. Il permet simplement de noter des taxons observés qui ne sont pas listés dans le référentiel. Ces ajouts sont traités ensuite, après le traitement de pré-bancarisation, et intégrés aux référentiels (codes 6L et Sandre) après validation.
- Anomalies : données aberrantes quand le taxon existe dans le menu déroulant mais qu'il est saisi comme un nouveau taxon ; données aberrantes quand le nouveau nom saisi est le synonyme d'un taxon figurant dans le menu déroulant ou déjà saisi dans la liste floristique ; données aberrantes quand l'orthographe du nom de taxon est erroné ; données aberrantes quand le nom de taxon n'existe pas dans les référentiels taxinomiques nationaux et internationaux ; données aberrantes quand le taxon ajouté est un ligneux ou que le taxon est renseigné à la famille (le niveau taxinomique doit être conforme à celui préconisé par la norme).

Code Sandre d'un nouveau taxon

- Format : Numérique.
- Description : le code Sandre de nouveau taxon doit être demandé auprès du Master Data Management (MDM) du Sandre. Le MDM⁶ est l'outil qui permet de faire des demandes d'ajout ou de modification dans les jeux de données de référence du Sandre.
- Anomalies : donnée absente ; donnée aberrante en cas de non-concordance entre le nom du taxon et le code Sandre du référentiel « Appellations de taxons ».

Code alternatif d'un nouveau taxon

- Format : Texte (6 lettres).
- Description : le code alternatif d'un nouveau taxon est généré par les experts chargés des référentiels (actuellement Irstea) pour permettre la non-redondance des codes et le respect de la nomenclature. Le code alternatif 6L par défaut à utiliser par les opérateurs lors de la saisie d'un nouveau taxon (taxon absent du menu déroulant de la fiche floristique) est par convention « NEWCOD »
- Anomalies : donnée aberrante quand le code 6L a été créé par l'opérateur.

4.2.3 Données de description stationnelle

Altitude

- Format : Numérique.
- Description : altitude exprimée en mètre, elle est généralement issue de l'indication d'élévation donnée par le GPS servant à relever les coordonnées géographiques.
- Remarque : ce champ est fréquemment vide, sans doute du fait de sa redondance avec les caractéristiques données pour la station par les référentiels « Stations », ou de son manque de lien apparent avec la végétation. La donnée, lorsqu'elle est renseignée, n'a jusqu'alors jamais fait l'objet de vérification pré-bancarisation par Irstea. La donnée est pour autant bien bancarisée.

⁶ <http://mdm.sandre.eaufrance.fr/>

Pour le traitement des données impliquant la prise en compte de l'altitude, Irstea utilise les altitudes renseignées dans le référentiel « Station de mesure » du Sandre.

- Anomalies : donnée absente ; donnée aberrante de par sa valeur qui doit être comprise entre 0 et 4000m ou de par son manque de cohérence par rapport au référentiel Sandre « Stations de mesure de la qualité des eaux superficielles continentales).

Pourcentage de végétalisation de chacune des unités de relevé

- Format : Numérique (en %).
- Description : correspond à la projection verticale de l'ensemble de la végétation par rapport à la surface en eau du point de prélèvement. Le pourcentage de végétalisation de chacune des unités de relevé ne peut pas dépasser 100%.
- Anomalies : donnée absente ; données aberrantes car non-concordance entre fiches stationnelles et floristique ; données aberrantes lorsqu'une unité de relevé présente un pourcentage de végétalisation supérieure à 100%. Données aberrantes lorsque le pourcentage de végétalisation total est inférieur à celui d'un des taxons.

Pourcentage de recouvrement des taxons exprimés par groupes taxinomiques

- Format : Numérique (%).
- Description : estimation, en pourcentage, des recouvrements des principaux groupes taxinomiques (Hétérotrophes, Algues, Bryophytes, Lichens, Ptéridophytes et Phanérogames). Cette donnée complète la liste floristique qui donne le recouvrement de chacun des taxons et permet de décrire le peuplement végétal sous plusieurs angles. Cela permet d'apprécier les erreurs d'estimation de recouvrements notés dans la liste floristique. Le recouvrement des groupes taxinomiques est exprimé en pourcentage. Le total peut dépasser 100% (ex : Algues recouvrant des Phanérogames ou bien encore végétation flottante au-dessus des populations immergées, etc.).
- Remarque : cette donnée n'a pas vocation à être bancarisée, elle ne constitue pas un résultat de la mesure « macrophytes ». Elle est utilisée par les macro-commandes de la feuille de saisie pour un premier contrôle de cohérence.
- Anomalies : données aberrantes quand les pourcentages de recouvrements par groupes taxinomiques sont non-concordants avec les recouvrements des taxons saisis dans la liste floristique.

Pourcentage de recouvrement des taxons exprimés par groupes fonctionnels

- Format : Numérique (%).
- Description : estimation, en pourcentage, des recouvrements des principaux groupes fonctionnels. Les Hydrophytes flottants ou à feuilles flottantes regroupe les espèces et les formes dont au moins une partie de l'appareil foliaire s'étend à la surface de l'eau (nénuphars, lentilles, feuilles flottantes de rubaniers et de grands potamots, etc.). Les Hydrophytes immergés sont toutes les espèces ou les formes qui se développent dans la masse d'eau sans atteindre la surface. Les Hélophytes sont toutes les espèces et les formes qui présentent au moins une partie de leur appareil foliaire dressé au-dessus de la surface de l'eau (ex : acore, baldingère, rubanier, massettes, etc.). Cette donnée permet de décrire la morphologie du peuplement en relation avec le morphodynamisme du cours d'eau au niveau du point de prélèvement. Le total peut dépasser 100%.
- Remarque : idem recouvrements par groupes floristiques.
- Anomalies : données aberrantes quand les pourcentages de recouvrements par groupes fonctionnels sont non-concordants avec les recouvrements des taxons saisis dans la liste floristique.

Recouvrement des types de faciès par unité de relevé

- Format : Numérique (classes de 0 à 5 : 0 : x = absent ; 1 : $x < 1\%$; 2 : $1\% \leq x \leq 10\%$; 3 : $10\% \leq x \leq 25\%$; 4 : $25\% \leq x \leq 75\%$; 5 : $x \geq 75\%$).
- Description : classes de recouvrement pour chaque type de faciès et pour chacune des unités de relevé. Noter la part de chaque faciès morphologique dans chaque unité de relevé permet de caractériser ces unités de relevé, de contribuer à l'analyse de la composition floristique et donc à la qualification des données.
- Anomalies : donnée absente ; données aberrantes : la saisie par un menu déroulant offrant la possibilité de saisir les classes de 0 à 5 réduit les erreurs. La somme totale des recouvrements minimaux renseignés par les classes doit toutefois être inférieure à 100 % et la somme des recouvrements maximaux doit être supérieure à 100 % (ex d'anomalies : 2 classes de recouvrement « 5 » conduisant à un recouvrement total au moins supérieur à 150% ; 5 classes de recouvrement « 1 » conduisant à un recouvrement total maximal inférieur à 5%).

Recouvrement des types de substrat par unité de relevé

- Format : Numérique (classes de 0 à 5). Idem précédent.
- Description : Classes de recouvrement pour chaque type de substrat et pour chacune des unités de relevés. Noter la part de chaque substrat dans chaque unité de relevé permet de caractériser ces unités de relevé, de contribuer à l'analyse de la composition floristique et donc à la qualification des données.
- Anomalies : donnée absente ; données aberrantes : la saisie par un menu déroulant offrant la possibilité de saisir les classes de 0 à 5 réduit les erreurs. La somme totale des recouvrements minimaux renseignés par les classes doit toutefois être inférieure à 100 % et la somme des recouvrements maximaux doit être supérieure à 100 %.

Recouvrement des classes de profondeur, vitesse et éclaircissement par unité de relevé

- Format : Numérique (classes de 0 à 5). Idem précédents.
- Description : Il s'agit d'estimer la part prise par chaque classe de profondeur, vitesse et éclaircissement sur chacune des unités de relevés. Noter la part de chaque classe de profondeur, vitesse et éclaircissement dans chaque unité de relevé permet de caractériser ces unités de relevé et contribue à l'analyse de la composition floristique et à la qualification des données.
- Anomalies : donnée absente ; données aberrantes : la saisie par un menu déroulant offrant la possibilité de saisir les classes de 0 à 5 réduit les risques erreurs. La somme totale des recouvrements minimaux renseignés par les classes doit toutefois être inférieure à 100 % et la somme des recouvrements maximaux doit être supérieure à 100 %.

Périphyton

- Format : Liste (menu déroulant : absent – peu abondant – abondant - très abondant).
- Description : il s'agit de considérer les biofilms bien visibles, feutrages algaux, couvertures de diatomées autres que les types macroscopiques pris en compte comme « macrophytes » présents sur le substrat ou sur les végétaux. Cette estimation n'est pas prise en compte dans la liste floristique ni dans les calculs de métriques ou d'indice. Elle constitue uniquement une aide à l'interprétation des résultats, en particulier lorsque l'état de la végétation est rapproché des paramètres physico-chimiques (très forte amplitude en oxygène dissous ou en pH alors que le recouvrement en macrophytes est très faible, par exemple).
- Remarque : cette donnée complémentaire, qui ne fait pas partie de la description du peuplement macrophytique, n'est pas bancarisée.
- Anomalies : donnée absente ; données aberrantes (la saisie par un menu déroulant empêche en principe les erreurs de libellé).

Hydrologie

- Format : Liste (menu déroulant : étiage sévère – basses eaux, étiage normal – moyennes eaux – hautes eaux).
- Description : décrit les conditions hydrologiques au moment de la réalisation du point de prélèvement. Elle est estimée par observation des indicateurs de niveau d'eau relatif sur les berges : végétation ripicole se trouvant immergée, lignes de bryophytes au ras de la ligne d'eau, algues sèches sur les pierres émergées, etc. La norme mentionne les conditions dans lesquelles un relevé IBMR doit être réalisé, qui doivent correspondre à des conditions des basses eaux – étiage normal (débit de l'ordre du QMNA₅).
- Anomalies : donnée absente ; données aberrantes (la saisie par un menu déroulant empêche en principe les erreurs de libellé).

Météo

- Format : Liste (ensoleillé – faiblement nuageux – fortement nuageux – pluie fine – orage, pluie forte – conditions crépusculaires).
- Description : conditions météorologiques au moment de la réalisation du point de prélèvement. C'est un facteur conditionnant l'observation de la végétation subaquatique car influençant la pénétration de la lumière dans l'eau (incidence, lumière diffuse) et l'état à la surface.
- Anomalies : donnée absente ; données aberrantes (la saisie par un menu déroulant empêche en principe les erreurs de libellé).

Turbidité

- Format : liste (menu déroulant : nulle – faible – moyenne – forte).
- Description : conditions de turbidité lors de la réalisation du relevé. La turbidité est un facteur limitant l'observation de la végétation subaquatique et l'estimation des recouvrements.
- Anomalies : donnée absente ; données aberrantes (la saisie par un menu déroulant empêche en principe les erreurs de libellé).

Largeur du cours d'eau au niveau du point de prélèvement

- Format : Numérique.
- Description : L'estimation sur site de la largeur « moyenne » du cours d'eau vise principalement à caractériser le cours d'eau par rapport à une typologie française. L'estimation de la largeur de la section mouillée est une mesure simple, qui est utilisée comme « proxy » de l'ordre de drainage, validé par les travaux sur la typologie nationale. Elle peut être mesurée selon l'équipement disponible et les conditions locales, au décimètre, au télémètre ou autres techniques équivalentes. Les mesures devront intégrer la variabilité de la largeur sur le tronçon délimité pour la réalisation du point de prélèvement (indication sur la naturalité du lit).
- Remarque : la feuille de terrain prévoit que puisse être renseignée la largeur sous le format : « largeur plus ou moins » pour prendre en compte l'amplitude des largeurs sur le point de prélèvement. Les fiches de saisies ne prévoient pas de champ pour cette amplitude mais elle doit être prise en compte dans la valeur de largeur saisie.
- Anomalies : donnée absente ; données aberrantes par rapport au type national de la station indiqué dans le référentiel « Stations de mesure » du Sandre.

Longueur du point de prélèvement

- Format : Numérique.
- Description : la norme AFNOR prévoit une longueur du point de prélèvement de 50 à 100 m selon la densité de la végétation macrophytique. Les documents d'encadrement des applications DCE imposent une longueur systématique de 100 m (retour d'expérience et préconisations de la norme européenne). Il est toutefois parfois inévitable, selon les conditions locales, de tronquer le point de prélèvement.

Lors de la mise en œuvre du protocole « grand cours d'eau » (utilisation du protocole par points-contacts ou largeur supérieure à 100 m), la longueur du tronçon de relevé (point de prélèvement) est portée à 200 m (préconisations techniques reprises du projet de guide d'application de la norme, incluses dans les cahiers des charges).

- Anomalies : données absentes ; données aberrantes

Longueur respective de chacune des unités de relevé

- Format : Numérique.
- Description : longueur de chacune des unités de relevés identifiées. Elle est exprimée en mètres. Cette donnée permet de décrire la répartition de la longueur du point de prélèvement entre les deux unités de relevé. Ils peuvent être successifs (par exemple 20 m et 80 m sur les 100 m du point de prélèvement) ou juxtaposés au moins en partie (100 m de l'unité de relevé la plus rapide et 60 m de l'unité de relevé la plus lente sur un point de prélèvement de 100m également). Ces valeurs peuvent être estimées *a posteriori* sur le schéma du point de prélèvement si celui-ci a été levé de façon juste et proportionnée.
- Anomalies : donnée absente ; données aberrantes par rapport aux pourcentages de recouvrement des unités de relevés ou à la longueur du point de prélèvement.

Largeur respective de chacune des unités de relevé

- Format : Numérique.
- Description : largeur de chacune des unités de relevé identifiées. Elle est souvent différente entre unité de relevé la plus rapide et la plus lente. Elle permet de préciser la valeur générale de largeur du point de prélèvement.
- Anomalies : donnée absente ; donnée aberrante par rapport aux pourcentages de recouvrement des unités de relevés ou à la largeur du point de prélèvement.

Protocole de relevé

- Format : Liste (menu déroulant : IBMR standard ; points contacts; mixte).
- Description : type d'approche suivi : protocole standard pour les petits et grands cours d'eau prospectables à pied, protocole par points-contacts pour les grands cours d'eau ou en cas de très faible transparence, protocole mixte pour les relevés comprenant des zones prospectables à pied et d'autres par points-contacts ou utilisation d'un râteau.
- Anomalies : donnée absente.

4.2.4 Observations et schéma

Observations

- Format : Texte.
- Description : champ réservé à la notation en clair de tous les éléments susceptibles d'expliquer ou d'impacter le peuplement macrophytique observé : présence d'un ouvrage hydraulique en amont ou en aval, traces d'entretien des berges, fréquentation par le bétail, particularités morphologiques, influence visible d'un rejet ou d'un affluent, etc. Les précisions à reporter concernent également une éventuelle délimitation particulière du point de prélèvement (contraintes locales, faciès particuliers) ou son dépassement par rapport aux campagnes antérieures (modifications morphologiques, impossibilité d'accès, etc.) ou toutes adaptations du protocole ayant été nécessaires. L'absence de végétation macrophytique au sein du point de prélèvement doit être inscrite dans le champ « observation » pour expliquer que la liste floristique est vide.
- Anomalies : donnée absente quand la liste floristique est elle aussi absente.

Schéma de la station

- Format : Image.
- Description : Le schéma cartographique doit répondre à deux objectifs principaux : décrire clairement les bornes amont et aval du point de prélèvement où le relevé a été précisément effectué. Il doit également permettre de décrire les unités de relevés c'est-à-dire d'indiquer la localisation et la superficie des unités de relevés. Tous les éléments susceptibles de fournir des repères fixes sont reportés sur le schéma : arbres caractéristiques (possibilité d'indiquer l'espèce sur le schéma mais pas dans la liste floristique), occupation particulière de la berge, clôtures, affluence de fossés ou canaux, bâtiments, etc. Il doit également fournir une vue synthétique des conditions stationnelles (écoulements, zones émergées, substrats remarquables, embâcles ; aménagements de berges, etc.) de la végétation et sa répartition afin d'aider à l'exploitation des résultats floristiques à une analyse diachronique éventuelle. Il pourra le cas échéant également être utilisé pour estimer les changements de conditions et de répartition des unités de relevé selon le débit. La graduation du schéma doit faciliter un levé aux proportions. Le sens d'écoulement doit être figuré.
- Anomalies : donnée absente.

4.3 Impacts potentiels des anomalies

Le degré d'impact des anomalies constatées dans les données dépend de leur nature obligatoire, descriptive de la station ou des conditions d'observation ou complémentaires. Ainsi, on peut distinguer 3 niveaux d'anomalies, de fortement à potentiellement impactantes. Dans un système de pré-traitement, ces niveaux pourront être codés pour chaque anomalie identifiée, pour évaluer le degré de non-conformité global résultant.

4.3.1 Anomalies fortement impactantes

Les anomalies considérées comme fortement impactantes portent sur les données obligatoires d'après la norme NF T90-395 et les documents d'accompagnement (prescription de l'agrément, en particulier), c'est à dire les données qui permettent l'identification et la localisation précise de l'opération de contrôle et celles qui sont utilisées pour le calcul de l'indicateur IBMR. Elles peuvent en effet conduire à des erreurs dans l'évaluation de la qualité écologique des masses d'eau par un calcul erroné de la note ou par un rattachement erroné à la station.

Les anomalies à identifier seront donc celles portant sur :

- le code station,
- le nom de la station (même si son libellé exact peut ne pas être strictement identique à celui du référentiel Sandre),
- le nom du cours d'eau,
- la date du relevé,
- les coordonnées géographiques du point de prélèvement,
- le nom de l'opérateur (organisme et agent)
- le pourcentage de recouvrement de chacune des unités de relevés,
- la liste floristique comprenant les codes 6L ou noms des taxons,
- le recouvrement en % de chaque taxon.

4.3.2 Anomalies impactantes

Les anomalies portant sur les données non obligatoires au sens des référentiels méthodologiques, utiles pour la description de l'environnement de la station et la caractérisation de l'opération de contrôle, sont impactantes pour l'évaluation de la qualité écologique des masses d'eau. Ce sont en effet les données permettant à l'opérateur de décrire les conditions dans lesquelles ont été effectués

les relevés (donc participant à l'évaluation de leur fiabilité), conditions pouvant influencer sur la visibilité dans l'eau : la météo, la turbidité. Ces informations sont utiles pour qualifier la liste.

Il s'agit également des anomalies portant sur les données relatives aux conditions hydrologiques au moment du relevé. En effet un relevé effectué dans des conditions hydrologiques ne correspondant pas aux « basses eaux » ou « étiage normal », mais en « moyennes eaux » pourra induire une liste floristique plus longue incluant davantage d'espèces végétales de berges (ce qui impacte des métriques utilisant la richesse taxinomique). De plus, ces conditions sont exclues du domaine d'application de la norme.

Des erreurs sur la longueur, la largeur du point de prélèvement et de chacune des unités de relevés peuvent être impactantes car elles permettent de vérifier la cohérence du pourcentage de recouvrement de chacune des deux unités de relevés, pourcentage qui est pris en compte dans le calcul de la note IBMR et donc participe à l'évaluation.

Les erreurs concernant le pourcentage de végétalisation des unités de relevé sont également impactantes car cette donnée permet de qualifier les informations concernant le recouvrement des taxons dans chaque unité de relevé, pourcentages qui sont pris en compte dans le calcul de la note IBMR.

Le non-renseignement du type de protocole de relevé est également impactant pour la qualification des données car suivant le protocole suivi, la liste des taxons observés peut être différente. Savoir qu'une station habituellement prospectée à pied a dû, au moins en partie, être prospectée par points-contacts (au râteau voire avec une embarcation), peut renseigner sur les conditions hydrologiques dans lesquelles le relevé a été effectué. Cette donnée est importante pour qualifier la liste floristique sur laquelle repose la note IBMR.

Enfin, l'absence de renseignement ou les anomalies concernant les pourcentages de recouvrements des différents types de faciès, de classes de profondeur, vitesse, éclaircissement et types de substrat par unité de relevé peuvent également poser des problèmes pour qualifier les listes floristiques. En effet, ces paramètres sont à mettre en relation avec l'autoécologie des taxons observés ce qui doit permettre de vérifier qu'il n'existe pas d'aberration (ex. : un taxon rhéophile dans un contexte très lentique). De plus, ce sont des données essentielles pour l'utilisation des données à des fins de développement de méthode d'évaluation et de diagnostic. Elles constituent en effet un jeu de données hydromorphologiques à l'échelle de la station qui peut être mis directement en lien avec les communautés végétales en présence ce qui apparait plus précis que des données à l'échelle du tronçon voire du bassin versant.

4.3.3 Anomalies potentiellement impactantes

L'absence de renseignement du champ « observations » peut être impactante lorsque des contraintes ont été rencontrées sur le terrain pour la réalisation du relevé et que celles-ci n'ont pas été rapportées. On peut citer par exemple le remaniement d'une berge impactant la végétation riparienne, la présence d'un gué avec perturbation morphologique et trophique par le bétail, le constat d'une pollution, un accès difficile n'ayant pas permis la réalisation de tout ou partie du relevé.

L'absence de transmission du schéma de station peut elle aussi s'avérer impactante. Le schéma peut en effet permettre de valider la part prise par chaque unité de relevé au sein du point de prélèvement ainsi que le recouvrement par les formations végétales représentées. Le schéma peut aussi permettre de qualifier les données sur les paramètres soutenant la biologie s'il comporte des annotations concernant les profondeurs, les vitesses, etc. Enfin, lors de campagnes spécifiques du

type de celles organisées pour évaluer l'incertitude de la mesure hydrobiologique inter-opérateur ou saisonnière, l'absence de schéma peut compliquer la location précise du point de prélèvement, le schéma permettant en effet de matérialiser les limites amont et aval du point de prélèvement ainsi que des particularités hydromorphologiques ou végétales permettant de se repérer le long du linéaire de cours d'eau.

5 ELÉMENTS DE CONTRÔLE ET DE VALIDATION

5.1 Éléments de contrôle automatisé

La transmission des données sous format standardisé, en l'occurrence le modèle MS.Excel de 2006 à 2014, permet de mettre en place des contrôles informatiques automatiques. Après analyse de la nature et du format des données attendues dans les fiches de saisie, Irstea a développé un script en langage Python permettant de mettre en évidence les anomalies de la plupart des champs. Cette automatisation des contrôles permet de traiter simultanément un grand volume de données. De plus, elle est plus rigoureuse qu'un contrôle visuel car si l'erreur est possible à la saisie, elle l'est également à la vérification. Les résultats sont obtenus sous la forme de tableau Excel indiquant pour chaque opération de contrôle, les données validées et les données à vérifier. C'est sous cette forme de tableau synthétique établi pour chaque Bassin que sont adressés, aux Agences, les rapports de contrôle de pré-bancarisation.

Les contrôles automatiques mis en place détectent tout d'abord la non-concordance des champs redondants entre fiches floristique et stationnelle c'est-à-dire :

- les identifiants de la station, du point de prélèvement et de l'opérateur ;
- la date. Un décalage de plus d'un jour est signalé comme anomalie potentielle. La justification est éventuellement appréciée en fonction des observations reportées dans le champ « Observations » ;
- le nombre d'unité de relevé ;
- le pourcentage de recouvrement des unités de relevé ;
- le pourcentage de végétalisation des unités de relevés.

Les contrôles automatiques permettent de dresser la liste des champs qui n'ont pas été remplis.

Des contrôles automatiques sont également effectués sur les champs numériques s'agissant des données ayant un format prédéfini et / ou une amplitude de valeur déterminée :

- le code station doit être de 8 chiffres. Il doit correspondre à celui qui est indiqué dans le référentiel « Station de mesure » du Sandre. Ce code est souvent amputé du premier chiffre 0 par le passage par des applications telles qu'Excel ;
- la date doit être du type JJ-MM-AAAA. Les jours doivent être compris entre 1 et 31. Pour les mois, il est vérifié que la valeur saisie est comprise entre mai et octobre qui correspond au domaine d'application de la méthode (période de végétation, préconisé par la norme) ; La non-redondance avec un relevé déjà bancarisé à cette station / date est également contrôlé (pas de doublon possible, car le couple station/date définit l'opération de contrôle) ;

- les coordonnées géographiques du point de prélèvement doivent être renseignées en Lambert 93 avec une coordonnée x comprise entre 100000 et 125000 m et une coordonnée y entre 6000000 et 7200000 m correspondant à l'emprise du territoire français métropolitain. Elles doivent par ailleurs correspondre à celles données par le référentiel « Station de mesure » du Sandre.
 - un décalage de quelques centaines de mètres (moins d'1 km) peut être considéré comme plausible,
 - un décalage de quelques kilomètres (moins de 5 km) doit être confirmé s'il n'y a pas d'historique actant la localisation,
 - un décalage de plusieurs kilomètres (supérieur à 5 km, sans justification dans le champ « observations ») entre la position du point de prélèvement donnée par les coordonnées saisies et les coordonnées de la station donnée par le référentiel du Sandre caractérise une anomalie ;

De même, des coordonnées strictement identiques à celle de la « station » au sens du Sandre nécessitent une confirmation (dans ce cas, ce sont souvent les coordonnées de la « station » qui ont été recopiées, et non celles du point de prélèvement indiquées). Lorsque la station a fait l'objet de plusieurs années de relevé, l'historique peut être utilisé pour déceler des changements de localisation. De tels changements doivent être motivés.

- les pourcentages de recouvrement des types de faciès, profondeur, vitesse, éclaircissement, substrats renseignés par classe. Les contrôles automatiques ne peuvent être que partiels, visant à détecter des aberrations formelles. Il est vérifié que la somme des recouvrements minimaux correspondant aux classes renseignées est inférieure ou égale à 100% et la somme des recouvrements maximaux est supérieure ou égale à 100%. Pour le recouvrement minimum, les valeurs de 0%, 0%1%, 10%, 25%, 75% sont respectivement attribués aux classes 0 (0%), 1 (0-1%), 2 (1-10%), 3 (10-25%), 4 (25-75%), 5 (75-100%) et 0%, 1%, 10%, 25%, 75% et 100% pour le recouvrement maximal.
- le pourcentage de recouvrement d'un taxon au sein d'une unité de relevé ne peut pas dépasser 100% ni être nul dans les deux unités de relevé.

Pour les données chiffrées non formatées et dont l'amplitude des valeurs possibles est large, des contrôles ont été mis en place en fixant des valeurs au-delà desquelles la donnée est vraisemblablement aberrante :

- la longueur d'un point de prélèvement ne peut pas dépasser 200 m (valeur haute prenant en compte les relevés en grand cours d'eau). Il serait utile de tenir compte de l'information « type de relevé » pour estimer la conformité de cette longueur.
- la longueur d'une unité de relevé ne peut pas être plus grande que la longueur du point de prélèvement.
- la largeur d'un point de prélèvement ne peut dépasser 2000 m (valeur haute prenant en compte les relevés en grand cours d'eau ;
- la somme des pourcentages de recouvrement des taxons au sein d'une unité de relevé doit être égale au pourcentage de végétalisation saisi pour l'unité de relevé à plus ou moins 20%.

Pour les données textuelles qui sont à renseigner via un menu déroulant, les contrôles automatiques dans le formulaire de saisie permettent d'assurer la conformité du texte à l'un des intitulés de ces menus. Ces vérifications portent sur les champs : protocole utilisé, rive, hydrologie, météo, turbidité, périphyton et codes 6L des taxons ou nom des taxons.

5.2 Éléments de contrôle par expertise

Si un grand nombre d'informations peuvent faire l'objet de contrôles informatiques automatisables, il en demeure certaines dont la validation ne peut être effectuées que par un avis d'expert hydrobiologiste.

Ainsi la validation de la liste taxinomique elle-même est un type de contrôle essentiel. Toutefois, elle nécessite une vérification de la cohérence biogéographique et écologique entre les taxons saisis d'une part et la localisation et les caractéristiques hydromorphologiques et physico-chimiques de la station d'autre part. Cette validation, en l'état actuel des données informatisées ou informatisables, n'est pas automatisable. Elle ne peut donc être effectuée que par un expert disposant des connaissances requises en floristique, écologie et hydrobiologie et s'appuyant sur les données stationnelles soutenant la biologie relevées pendant l'opération de contrôle.

Vu le nombre très élevé de données collectées et bancarisées annuellement (potentiellement plus de 600 en 2011 à 1200 en 2012-13), un examen de chaque relevé par un expert n'est pas envisageable. Seules quelques aberrations sont parfois constatées, lors de la visualisation fortuite de liste ou de leur utilisation en développement méthodologique (relevés qui ressortent comme outliers dans les analyses statistiques).

Actuellement, ce contrôle doit donc être réalisé au niveau de la vérification de la bonne réalisation des marchés par les prestataires. Il revient donc à la charge des responsables de données (les Agences de l'eau très principalement) ou de prestataires spécialisés dument mandatés. Toutefois, pour les mêmes raisons de quantité de données à contrôler, ces vérifications ne peuvent porter que sur des échantillons de données, ou par examen en amont des compétences des opérateurs et des mesures qu'ils prennent pour assurer la fiabilité des résultats. Ces points sont en théorie examinés et validés lors des processus d'accréditation.

Certains points particuliers devraient faire l'objet d'un suivi. Par exemple la prise en compte par les opérateurs de la validation par des experts extérieurs de la détermination de certains échantillons, processus qui est prévu dans les cahiers des charges. Plusieurs cas ont été constatés, portant sur des résultats transmis pour bancarisation alors que la validation taxinomique n'était pas encore revenue des experts consultés par les opérateurs. Dans ces cas, aucun correctif n'a été transmis *a posteriori* par les opérateurs.

S'agissant de la liste floristique, lorsqu'un nouveau nom de taxon est rencontré, c'est-à-dire l'appellation d'un taxon qui ne figure pas dans le référentiel (menu déroulant) intégré au modèle de saisie Irstea, une expertise est là aussi nécessaire pour vérifier que le nom saisi est valide taxinomiquement. Pour se faire, l'expertise est étayée par une consultation des référentiels reconnus : principalement le référentiel national TAXREFF (Inventaire National du Patrimoine Naturel ; MNHN) pour les Phanérogames, Bryophytes et Ptéridophytes et AlgaeBase (algeabase.org) pour les Algues et les Hétérotrophes. Le « nouveau » taxon doit lui aussi être cohérent par rapport à la biogéographie de la station et à ses caractéristiques hydromorphologiques et physico-chimiques.

6 COMMENTAIRES, PERSPECTIVES

Malgré l'obligation faite aux opérateurs d'utiliser des formulaires standardisés de saisie des données « macrophytes en cours d'eau » et hydrobiologiques en général (jusqu'à présent sous modèles Excel pour la plupart, plus ou moins évolués⁷), il existe de multiples anomalies dans les données transmises : données oubliées, erreurs de saisie, données aberrantes, etc., voire non-utilisation des formulaires prescrits. Certaines anomalies interdisent même la bancarisation des données concernées sous le format de la base de données.

Ces anomalies peuvent avoir des niveaux d'impacts variables sur la qualité des données suivant la nature des données concernées et leur utilisation, de fortement impactantes lorsqu'il s'agit de données directement liées au calcul de la note IBMR à faiblement impactantes s'il ne s'agit que de données descriptives complémentaires.

Une autre vision pourrait également être appliquée, en distinguant les anomalies formelles (non-respect factuel de préconisations liées à la norme ou aux procédures qualité de l'accréditation et de l'agrément, par exemple) de celles susceptibles d'altérer la fiabilité ou l'exploitabilité des données. Par exemple, un taux de recouvrement végétal total non cohérent avec les taux cumulés par taxon est un indicateur d'imprécision importante ou d'erreur potentielle, alors que ce type de vérification n'est pas faite au titre du respect de la méthode. De même, l'absence des coordonnées géographiques est une non-conformité formelle (cette information est obligatoire), mais sa cohérence avec la position de la station, non encadrée par les textes, est un défaut qui peut rendre la donnée inutilisable. Ou encore l'absence du nom de l'opérateur est une non-conformité par rapport à l'agrément, mais n'impacte pas forcément la qualité du relevé.

L'annexe 1 regroupe une synthèse des anomalies, impacts et type de vérification associés à chaque type de données « macrophytes en cours d'eau ».

Ces données ayant vocation à alimenter les banques de données sur lesquelles s'appuient les protocoles d'évaluation, de rapportage et de diagnostic, mais également le développement méthodologique par les équipes scientifiques d'Irstea, la mise en œuvre de protocoles de qualification des données avant bancarisation est apparue indispensable. Afin de garantir la transparence sur la « correction » des données, ces protocoles de qualification sont mis en œuvre en collaboration avec les acteurs impliqués : les Agences de l'Eau, et les opérateurs de terrain, laboratoires de DREAL et bureaux d'étude prestataires. A l'exception de quelques cas évident (erreur de frappe dans le libellé d'une station clairement identifiée, par exemple), aucune modification des données originales n'est appliquée sans accord du producteur.

Pour réduire le temps consacré aux contrôles et aux corrections des données saisies, il est possible d'envisager des outils permettant une saisie davantage assistée et cadrée. Rappelons à ce sujet que les systèmes qui ont été mis en place dès le début de la surveillance pour la chimie, comme l'application Edilabo encadrant la production et la transmission des résultats sous des formats standardisés (scénarios SANDRE), n'existent pas pour les paramètres hydrobiologiques.

Cependant, il n'est pas pertinent d'aller plus loin dans la qualification des données avec les applications de saisie actuellement développées sous MS-Excel, pour des raisons de lourdeur de développement et d'utilisation de cet outil pour des applications complexes, mais aussi et surtout pour les problèmes de compatibilité qu'engendre l'évolution des versions de ces logiciels Microsoft ou les dysfonctionnement selon les plateformes informatiques utilisées (non compatibilité des

⁷ Ces modèles sont diffusés sur le site <http://hydrobio-dce.irstea.fr/>

fonctions VBA développées sous environnement Windows avec les logiciels libres ou les systèmes Mac, par exemple). A cela s'ajoute le fait que les modèles connaissent de fréquentes évolutions liées aux mises à jour de la liste taxinomique, aux décisions prises au niveau national pour l'application des mesures dans les réseaux et aux demandes émanant directement des utilisateurs qui se révèlent parfois difficiles à mettre en œuvre dans la structure de feuilles de calculs de Microsoft Excel. Les fiches Excel, bien qu'offrant un certain nombre d'avantages depuis maintenant plus de 10 ans, atteignent leurs limites.

Au regard de l'évolution de la structuration des circuits de collecte et de bancarisation des données de surveillance au niveau national, qui se réoriente vers une gestion complète autonome par bassins, il est probable que les outils indépendants, qui apparaissent sur le marché, se développent. Ils devraient permettre la saisie et la transmission harmonisée des données produites. Toutefois, ces outils devront répondre à un corpus de spécifications homogènes, en veillant à leur compatibilité au regard des protocoles d'acquisition de données, à la nature, la pertinence des types de contrôles intégrés, ainsi qu'à la complétude et aux procédures de mise à jour utilisés pour les contrôles à la saisie que ces outils comprendront.

Par exemple, l'application ALISMA *Application pour le traitement des LISTes floristiques des relevés MACrophytes* a été élaborée et diffusée en v1.0 en août 2016 par Irstea sur la base de l'expérience acquise durant le développement des versions successives de l'application MS-Excel. Cette application multiplateformes, développées sous licence libre (Java et PostgreSQL), intègrent un certain nombre de fonctions d'aide à la saisie et de contrôles durant la saisie et lors du calcul d'indicateurs. Cet outil comprenant une base de données, ces contrôles peuvent s'appuyer sur des référentiels intégrés à l'application et mis à jour (référentiels du Sandre, en particulier). Ce type d'application a vocation à évoluer en fonction des besoins des utilisateurs et de l'avancée des connaissances qui permettront des contrôles plus précis, en particulier sur les listes floristiques (utilisation des traits écologiques, par exemple).

Quelle que soit la technicité des outils de saisie utilisés, la validation de certaines données telles que les listes floristiques semble difficile à automatiser sans avis d'un expert hydrobiologiste. Il est envisageable cependant que le travail d'expertise puisse être en partie relayé par la consultation de base de données sur la biogéographie et l'autoécologie des taxons. Des catalogues tels que celui développé par Irstea à partir des données « macrophytes en cours d'eau » bancarisées depuis 2002, définissant les profils écologiques des macrophytes en cours d'eau (Loriot *et al.*, 2014), pourront servir de base à ce type d'approche. Dans cet atlas, par exemple, le profil écologique de 332 taxons a été établi et pour chacun d'entre eux, il comprend : une carte de répartition géographique du taxon suivant les occurrences de la base de données Pandore (relevé IBMR) avec la précision de son abondance exprimée en pourcentage de recouvrement et un tableau présentant les probabilités de présence (%) du taxon pour chaque paramètre et pour chaque modalité. Les occurrences relatives sont présentées ainsi que le nombre total d'occurrence dans la base Pandore ; Les probabilités de présence en fonction des modalités pour les paramètres hydromorphologiques et physico-chimiques. Ces éléments pourront servir à la mise en place d'une base de données directement interrogée par des applicatifs de saisie, et ainsi contribuer à la mise en évidence d'anomalies dans les listes taxinomiques.

La validité des listes floristiques elles-mêmes, en composition comme en complétude, reste le point essentiel de qualification des résultats « macrophytes en cours d'eau ». Ce contrôle n'est pas encore opérationnel, ni réalisé de manière systématique par experts. Un tel examen systématique n'est d'ailleurs pas envisageable, car il reviendrait à demander aux experts de faire une deuxième fois le travail réalisé par les opérateurs, ce qui serait une aberration opérationnelle. Cependant, l'examen ponctuel des résultats recueillis montre que ce sujet est une source d'altération majeure des résultats et d'incertitude dans les calculs d'indicateur d'évaluation.

Cet état de fait interroge sur la pertinence de l'obligation faite aux opérateurs dans les marchés de conserver les échantillons récoltés pendant une durée fixée (souvent plus de 2 ans), pour d'éventuels contrôles de détermination taxinomique. Dans la pratique, ces échantillons conservés ont un coût réel mais ne sont que très peu souvent utilisés, par manque de contrôle des résultats fournis qui pourraient déclencher une demande de vérification.

Pour les prochaines années, il semble souhaitable que le développement de systèmes de contrôle de la composition floristique des listes fasse l'objet de travaux pour fournir des bases utilisables par les applicatifs de saisie ou par les protocoles de validation lors de la bancarisation.

En tout état de cause, comme l'analyse des cycles de bancarisation le montre, le temps nécessaire aux échanges avec les opérateurs pour compléter ou corriger les données transmises constitue une source importante d'allongement des délais de mise à disposition des résultats pour l'exploitation en diagnostic ou en évaluation.

Au vu de la qualité actuelle des données brutes transmises par les opérateurs et des délais constatés pour les échanges visant à obtenir des données conformes ou confirmées, il paraît difficile que des données utilisées moins de 6 mois après leur fourniture puissent garantir une fiabilité satisfaisante, à moins que la réorganisation des circuits de collectes et de bancarisation des données ne permettent d'améliorer de façon importante l'efficacité des processus de contrôle, correction et validation des résultats.

Plusieurs points apparaissent comme devant faire l'objet d'amélioration dans la chaîne de production-contrôle-bancarisation :

- le contrôle de la conformité aux prescriptions des marchés de la livraison des résultats (complétudes des résultats, items renseignés dans les formulaires) ;
- le contrôle de la validation effective des listes floristiques par les prestataires, avant leur transmission aux donneurs d'ordre (identification des échantillons soumis à validation par experts, par exemple) ;
- l'efficacité des contrôles « métier » de pré-traitement dès réception des données brutes (identification des informations manquantes ou douteuses) ;
- la prise en compte des différentes phases de contrôle et de validation dans les protocoles de bancarisation, pour une qualification effective de chaque donnée. Dans cet optique, les critères de « qualification » des données hydrobiologiques devraient faire l'objet d'une définition objective et complète (notion de « statut » de la donnée, telle qu'elle se définit pour les résultats en chimie) ;
- l'intégration de protocoles de détection d'anomalie des listes floristiques (par traitement de l'historique ou analyse des profils écologiques). Ces protocoles restent encore à élaborer.

Avec l'évolution et la réorientation récente des systèmes de collecte et de bancarisation des données de surveillance, qui refocalise sur une gestion autonome par bassin après avoir privilégié une harmonisation et une centralisation au niveau national, il est d'autant plus important de s'assurer que l'ensemble de ces conditions sont remplies pour tous les systèmes qui contribuent à l'évaluation nationale, et qu'un tronc commun harmonisé est mis en place.

Enfin, au regard de l'option prise en France de baser la surveillance sur une multitude d'opérateurs soumis à des processus d'accréditation et d'agrément, il semble nécessaire d'utiliser ce retour d'expérience sur la bancarisation des données pour identifier les points sur lesquels ces processus ne garantissent pas la qualité effective des résultats. Le rapprochement entre encadrement amont et résultats effectifs serait à mener et analyser, pour permettre une évolution des processus d'encadrement de la production des données de surveillance vers une meilleure fiabilité qui améliorerait la qualité intrinsèque des résultats produits sans nécessiter des opérations de contrôle a posteriori trop lourdes et coûteuses.

7 DOCUMENTS CITES

AFNOR, 2003. Qualité des eaux - Indice Biologique Macrophytique Rivière. Norme NF T90-395 octobre 2003.

Arrêté du 7 août 2015 modifiant l'arrêté du 25 janvier 2010 établissant le programme de surveillance de l'état des eaux en application de l'article R. 212-22 du code de l'environnement. JORF, 28 août 2015

Arrêté du 27 juillet 2015 modifiant l'arrêté du 25 janvier 2010 relatif aux méthodes et critères d'évaluation de l'état écologique, de l'état chimique et du potentiel écologique des eaux de surface pris en application de l'article R. 212-22 du code de l'environnement. JORF, 28 août 2015

Arrêté du 27 octobre 2011 portant modalités d'agrément des laboratoires effectuant des analyses dans le domaine de l'eau et des milieux aquatiques au titre du code de l'environnement. JORF, 9 novembre 2011.

Chauvin C., Peltre MC, Haury J., 2009. L'indice Biologique Macrophytique en Rivière (NF T90-395). PrGA Document de travail de la Commission de normalisation AFNOR T95F. 99 p. + annexes.

Loriot S., Feret T., Chauvin C., 2014. Traits morphologiques et écologiques des macrophytes en cours d'eau. Recherche de métriques de diagnostic des perturbations hydromorphologiques et physico-chimiques. 61 p. + annexes.

Loriot S., Laplace-Treytore C., Boutry S., Bertrin V., 2014. Des applications informatiques pour faciliter l'acquisition des données « macrophytes » dans les réseaux de surveillance des masses d'eau continentales. Sciences, Eaux et Territoires, la revue d'Irstea, n°15 : 64 - 68.

SANDRE – Référentiels et dictionnaires. Site web <http://www.sandre.eaufrance.fr>

Modèles et formulaires de saisie et transmission des données de surveillance DCE "macrophytes en cours d'eau". Site web <http://hydrobio-dce.irstea.fr/>

8 ANNEXES

Annexe 1	Tableau des anomalies associées à chaque item des données « macrophytes cours d'eau »
Annexe 2	Modèle de fiche-terrain de relevé des données stationnelles et floristiques
Annexe 3	Modèle de formulaire de saisie des listes floristiques et de calcul de l'IBMR (extrait)
Annexe 4	Modèle de saisie des données stationnelles

Annexe 1 : Anomalies associées à chaque item des données « macrophytes cours d'eau », selon les formulaires de saisie Irstea « Formulaire de saisie de la description des points de prélèvement IBMR v_3.7 » et « Formulaire de saisie informatique « listes floristiques et calcul de l'IBMR v_4.3.1 »

	Donnée absente	Donnée aberrante	Non concordance fiches station et floristique	Non concordance avec les référentiels	Type de vérification
opérateurs	x		x		-
nom du cours d'eau	x	x	x	x	automatique
nom de la station	x	x	x	x	automatique
code Sandre station	x	x	x	x	automatique
référence du dossier	x		x		-
date	x	x	x		automatique
protocole de relevé	x				automatique
coordonnées	x	x		x	automatique
rive	x				automatique
organisme	x		x		-
altitude	x	x		x	automatique
hydrologie	x				automatique
météo	x				automatique
turbidité	x				automatique
nombre d'UR	x	x	x		automatique
% rec UR1 et UR2	x	x	x		automatique
longueur	x	x			automatique
largeur	x	x			automatique
longueur des UR	x	x			automatique
largeur des UR	x	x			automatique
% végétalisation des UR	x	x	x		automatique
périphyton	x				automatique
rec types faciès par UR	x	x			automatique
rec profondeur par UR	x	x			automatique
rec éclaircissement par UR	x	x			automatique
rec vitesse par UR	x	x			automatique
rec substrat par UR	x	x			automatique
liste floristique	x	x			expertise
code 6L ou nom du taxon	x	x		x	expertise
code Sandre taxons				x	automatique
% recouvrement taxon	x	x			automatique + expertise
cf.					expertise
nom nouveau taxon	x	x		x	expertise
code sandre newcod	x			x	automatique
% rec gr. taxinomiques	x	x			automatique
% rec types fonctionnels	x	x			automatique
observations	x				expertise
schéma de la station	x	x			expertise

- informations obligatoires ; anomalies fortement impactantes
- informations de description ; anomalies pour la qualification et l'interprétation
- informations complémentaires ; anomalies mineures

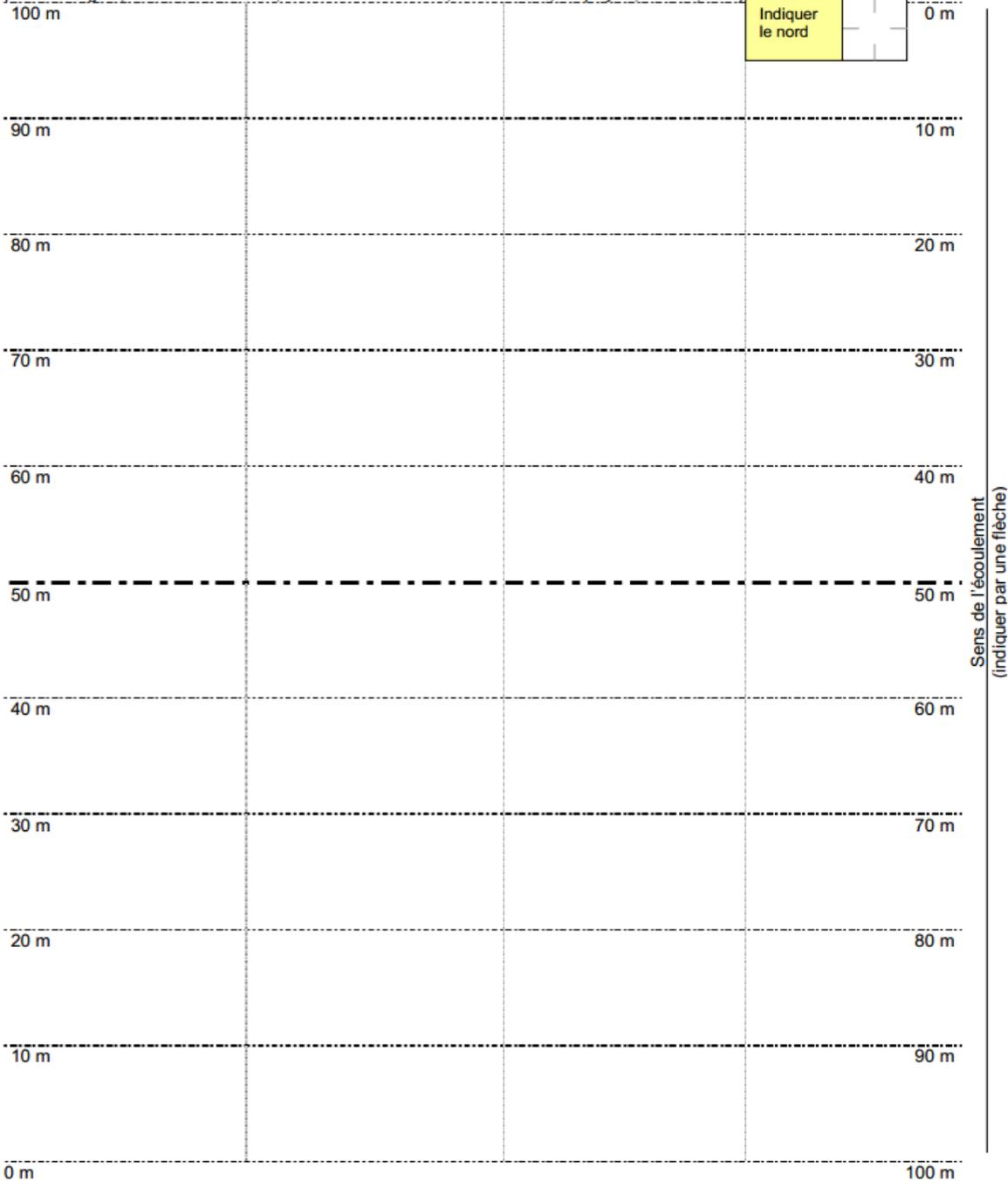
Annexe 2 : Modèle de fiche-terrain de relevé des données stationnelles et floristiques dans le cadre de la réalisation de relevé IBMR. Modèle Irstea v 3.46.

<http://hydrobio-dce.irstea.fr/telecharger/macrophytes-rivieres-2/>

Indice Biologique Macrophytique en Rivière – I.B.M.R.		Fiche Terrain IBMR		
STATION ET POINT DE PRELEVEMENT		conforme norme NFT90-395 oct. 2003		
<i>Consulter la notice explicative de ce formulaire de relevé « terrain » avant utilisation</i>				
Cours d'eau :		Date :		
Nom station :		Code station :		
Point de prélèvement :				
Organisme / Opérateur :		Réf. dossier :		
POINT DE PRELEVEMENT				
Coordonnées de la limite	prises en rive : <input type="checkbox"/> droite <input type="checkbox"/> gauche <input type="checkbox"/> centre en <input type="checkbox"/> amont <input type="checkbox"/> aval			
Lambert 93 (système français)	relevées sur : <input type="checkbox"/> GPS <input type="checkbox"/> carte IGN ou Géoportail			
WGS 84 (système GPS)	en mètres	X :	Y :	
	<input type="checkbox"/> Deg.min. sec. <input type="checkbox"/> Deg. décimaux	E/W :	N :	
Conditions d'observation	hydrologie : <input type="checkbox"/> étiage sévère <input type="checkbox"/> basses eaux, étiage normal <input type="checkbox"/> moyennes eaux <input type="checkbox"/> hautes eaux			
	météo : <input type="checkbox"/> ensoleillé <input type="checkbox"/> faiblement nuageux <input type="checkbox"/> fortement nuageux <input type="checkbox"/> pluie fine <input type="checkbox"/> orage, pluie forte <input type="checkbox"/> conditions crépusculaires			
	turbidité : <input type="checkbox"/> nulle <input type="checkbox"/> faible <input type="checkbox"/> moyenne <input type="checkbox"/> forte			
Dimensions	fond visible : <input type="checkbox"/> oui <input type="checkbox"/> non <input type="checkbox"/> partiellement			
longueur = m largeur = m ± m				
UNITES DE RELEVÉ				
Nombre d'unité de relevés retenues	UR 1		UR 2	
	<input type="checkbox"/> unité la plus courante		<input type="checkbox"/> unité la moins courante	
	<input type="checkbox"/> unité de relevé unique			
	<input type="checkbox"/> chenal (en grands cours d'eau)		<input type="checkbox"/> berges (en grands cours d'eau)	
<input type="checkbox"/> rive droite (en grands cours d'eau)		<input type="checkbox"/> rive gauche (en grands cours d'eau)		
Caractéristiques				
6 classes : 0 : x = absent ; 1 : x < 1% ; 2 : 1 ≤ x < 10% ; 3 : 10 ≤ x < 25% ; 4 : 25 ≤ x < 75% ; 5 : x ≥ 75%				
Faciès constitutifs (en classes de recouvrement) chenal lotique plat lentique chenal lotique plat lentique
 radier mouille radier mouille
 cascade chenal lentique cascade chenal lentique
 plat courant fosse dissipation plat courant fosse dissipation
 rapide zones de bordure rapide zones de bordure
 autre : autre : autre : autre :
% de l'unité de relevé dans la station				
Longueur de l'unité de relevé (en m)				
Largeur de l'unité de relevé (en m) ± m				
Morphologie				
Profondeur (m) (en classes de recouvrement) P < 0,1 1 ≤ P < 2 P < 0,1 1 ≤ P < 2
 0,1 ≤ P < 0,5 P ≥ 2 0,1 ≤ P < 0,5 P ≥ 2
Vitesse de courant (m/s) (en classes de recouvrement) 0,5 ≤ P < 1	 0,5 ≤ P < 1	
 V < 0,05 0,5 ≤ V < 1 V < 0,05 0,5 ≤ V < 1
Eclairement (en classes de recouvrement) 0,05 ≤ V < 0,2 V ≥ 1 0,05 ≤ V < 0,2 V ≥ 1
 0,2 ≤ V < 0,5	 0,2 ≤ V < 0,5	
 très ombragé éclairé très ombragé éclairé
 ombragé très éclairé ombragé très éclairé
 peu ombragé	 peu ombragé	
Substrats				
Type de substrat (en classes de recouvrement) Va Te De	
 Ca Bl Ar	
 Sa Ra		
 Va Te De	
 Ca Bl Ar	
 Sa Ra		
Va : Vases, limons (substrat meuble) ; Ca : Cailloux, pierres, galets ; Sa : Sables, graviers (substrat mobile) ; Te : Terre, argile, marne, tourbe (subs. solide) ; Bl : Blocs, dalles ; Ra : Racines et branchages De : Débris organiques ; Ar : Artificiel (béton) ;				
Fiche terrain IBMR v_3.46.doc		© Irstea/ GIS Macrophytes des eaux continentales		
		1/4		

Cours d'eau :		Date :	
Nom station :		Code station :	
Organisme / Opérateur :		Réf. dossier :	

(Noter la largeur, les unités de relevés, les herbiers dominants, les embâcles, la ripisylve, un bras, etc...)



Annexe 3 : Modèle de formulaire de saisie des listes floristiques et de calcul de l'IBMR. Application MS.Excel. Modèle Irstea v. 4.31 (extrait). <http://hydrobio-dce.irstea.fr/telecharger/macrophytes-rivieres-2/>

Relevés floristiques aquatiques - IBMR										modèle Irstea-GIS version 4.3.1 - février 2016		
(Organisme)		(Opérateurs)										
(Cours d'eau)		(Nom de la station)				(Code station)		(dossier, type réseau)				
(Date)	01/01/1901				Résultats							
Unité(s) de relevé	UR1	UR2	station		IBMR							
Faciès dominant			0		niveau trophique							
% UR/pt. prêt			0		cote sp. coef stén.							
VEGETALISATION				tot. pondéré		moyenne						
% surf. tot. veg./UR			0.00		écart-type							
périphyton					nb taxons							
% hétérotrophes			0.00		hét. -					mini -		
% algues			0.00		alg. -					maxi -		
% bryophytes			0.00		bryo. -					total -		
% ptérido. & lichens			0.00		pté./lich. -					contrib. -		
% phanérogames			0.00		phan. -					sténo. 1 -		
% vég. flottante			0.00		ratio contrib/total -					sténo. 2 -		
% vég. immergée			0.00							sténo. 3 -		
% héliophytes			0.00									
rec par UR			0	0.00								
rec. pondéré			0.00	0.00								
CODES	%	%	% sta.	grp	Csi	Ei	NOMS	(Cf)	SANDRE	Nouveaux taxons hors référentiel (si nécessaire)		
										Noms	cd_sandre	

Nouvelle feuille

IBMR Tri noms

Robustesse Tri groupes

Tri CS

Tri E

Récapitulatif

Export (feuille Excel)

Lignes suppl. Suppr. lignes

Référentiel Taxons

Notice

Annexe 4 : Modèle de saisie des données stationnelles dans le cadre de la réalisation d'un relevé IBMR. Feuille MS-Excel. Modèle Irstea v3.7. <http://hydrobio-dce.irstea.fr/telecharger/macrophytes-rivieres-2/>

ica Biologique Macrophytique en Rivière_I.B.1		v_3.7 /46cembre 2014	
DONNEES GENERALES SUR LA STATION ET LE POINT DE PRELEVEMENT			
Station			
Organisme			
Opérateur			
Code station			
Nom de cours d'eau			
Nom de la station			
Date (jj/mm/aaaa)			
Point de prélèvement			
Protocole de relevé			
Coordonnées prises en riverie:			
Coordonnées Lambert 93	X		
AMONT point de contrôle	Y		
Altitude (en m)			
Hydrologie			
Météo			
Turbidité			
Fond visible			
Longueur (en m)			
Largueur (en m)			
Unité de relevé			
Sélections possibles de recouvrement			
B: <0,1X	Q: <0,1X	Z: 1,5<=0,1X	M: 1,5<=0,25X
4: 2,5<=0,75X	5: >0,75X		
Nombre d'unités de relevé			
UNITE DE RELEVÉ 1 la plus rapide ou unique (au chenal au rive droite)		UNITE DE RELEVÉ 2 la plus lente (au bord au rive gauche)	
% de recouvrement de l'UR1		% de recouvrement de l'UR2	
longueur de l'UR1 (en m)		longueur de l'UR2 (en m)	
largeur de l'UR1 (en m)		largeur de l'UR2 (en m)	
% surface végétalisée de l'UR1		% surface végétalisée de l'UR2	
periphyton		periphyton	
Type de facies		Type de facies	
chenal lentique		chenal lentique	
plat lentique		plat lentique	
maillage		maillage	
faux-dérapant		faux-dérapant	
chenal latique		chenal latique	
radier		radier	
carcade		carcade	
plat courant		plat courant	
rapido		rapido	
autre type:		autre type:	
recouvrement de "autre type"		recouvrement de "autre type"	
Profondeur (m)		Profondeur (m)	
P < 0,1		P < 0,1	
0,1 ≤ P < 0,5		0,1 ≤ P < 0,5	
0,5 ≤ P < 1		0,5 ≤ P < 1	
1 ≤ P < 2		1 ≤ P < 2	
P ≥ 2		P ≥ 2	
Vitesse de courant (m/s)		Vitesse de courant (m/s)	
V < 0,05		V < 0,05	
0,05 ≤ V < 0,2		0,05 ≤ V < 0,2	
0,2 ≤ V < 0,5		0,2 ≤ V < 0,5	
0,5 ≤ V < 1		0,5 ≤ V < 1	
V ≥ 1		V ≥ 1	
Eclaircissement		Eclaircissement	
très ombragé		très ombragé	
ombragé		ombragé	
peu ombragé		peu ombragé	
éclairé		éclairé	
très éclairé		très éclairé	
Type de substrat		Type de substrat	
Vare, limon		Vare, limon	
Terre, argile, marne, tourbe		Terre, argile, marne, tourbe	
Cailloux, pierres, galets		Cailloux, pierres, galets	
Éboulis, dalles		Éboulis, dalles	
Sables, graviers		Sables, graviers	
Racines, branchages		Racines, branchages	
Détritus organiques		Détritus organiques	
Artificiels		Artificiels	
OBSERVATIONS			