**Annexe 7 : CONTROLES DE COHERENCE/VALIDITE DES DONNEES**

Le coordinateur lance le contrôle de validité sur les données après que celles-ci aient été intégrées en base. L’AELB demande à ce qu’un maximum de contrôles de validité sur les données brutes soit effectué par le coordinateur à partir de l’application Lyxea® Desktop.

En fonction des tests de validité des données lancés par le coordinateur lors du contrôle de cohérence, l’application précise pour chacun les niveaux de criticité rencontrés par les pictogrammes suivants :



Ces différents niveaux de contrôles de validité gérés dans l’application Lyxea® Desktop identifient certains résultats de mesures ou métadonnées associées sur lesquels la qualité des données peut-être remise en question.

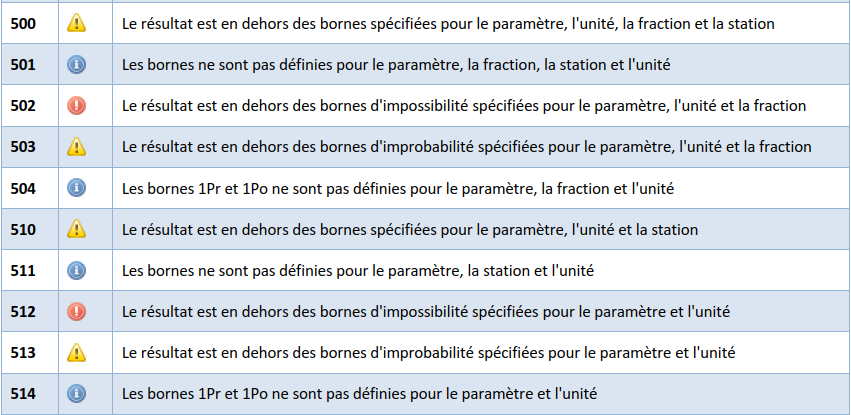
Les deux principaux contrôles de cohérence intégrés dans notre application de gestion sont :

* + - Comparaison des résultats à des bornes de vraisemblance par paramètre ou station/paramètre

Les principaux cas d’utilisation sont les suivants :

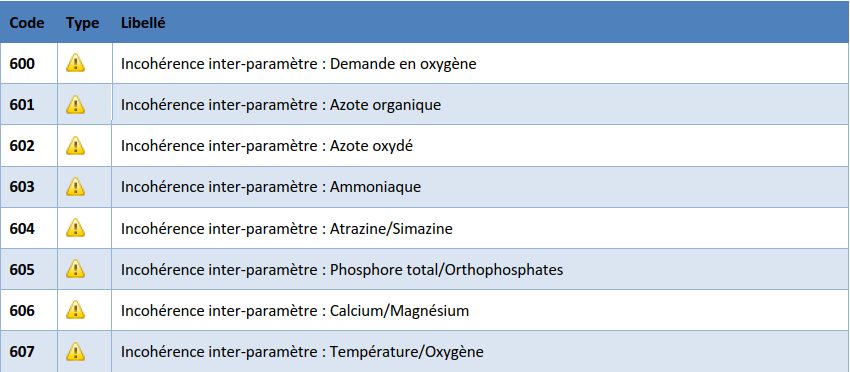
* Cas 1 : Si un résultat de mesure est supérieur à la borne impossible maximale par paramètre ou inférieur à la borne impossible minimale par paramètre, alors ce résultat ressort en erreur (code 502) ;
* Cas 2 : Si un résultat de mesure est compris entre la borne impossible par paramètre et la borne improbable par paramètre, alors ce résultat ressort en avertissement (code 503) ;

Les référentiels de bornes de vraisemblance par paramètre et par station/paramètre sont consultables dans Lyxea® Desktop et administrés par l’AELB.



* + - Tests inter-paramètres :

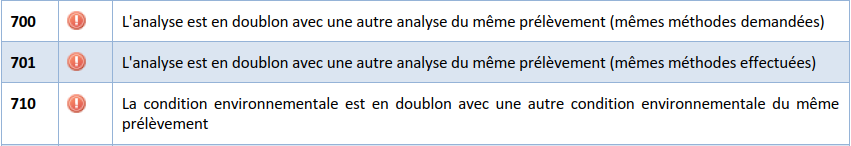
|  |  |  |
| --- | --- | --- |
| Libellé court erreur | Paramètres | Conditions de l’erreur |
| Demande en oxygène | 1314 : DCO  1313 : DBO5 | [1314] <= [1313] |
| Azote Organique | 1335 : NH4+  1319 : NK | [1335]\*0.776 >= [1319] |
| Azote oxydé | 1339 : NO2-  1340 : NO3- | [1339]\*0.304 >= [1340]\*0.226 |
| Ammoniaque | 1351 : Ammoniaque  1302 : pH | [1351] > 0 et [1302] < 8.3 |
| Phosphore Total / Orthophosphates | 1350 : P Total  1433 : Orthophosphates | [1350] < ([1433]/3.3)\*1 |
| Calcium/Magnésium | 1374 : Ca  1372 : Mg | [1374] <= [1372] |
| Température / Oxygène / Saturation | 1301 : T°C eau  1311 : O2  1312 : Taux de saturation en oxygène | 0,9 < ((([1311]/(14.6131-(0.4112\*[1301])+(0.008536\*[1301]\*[1301])-(0.00009\*[1301]\*[1301]\*[1301])))\*100)/[1312]) <1,1 |



Il est rappelé que c’est bien l’Agence de l’eau qui administre les référentiels de bornes de vraisemblance, des tests inter-paramètres mais que le coordinateur doit proposer des évolutions au regard de son expérience métier.

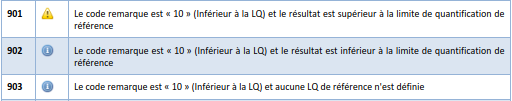
* + - Doublons :

Le contrôle de doublons concerne les analyses physico-chimiques et les conditions environnementales.



Est considéré comme doublon :

* Un prélèvement s’il a un même code prélèvement qu’un autre du « LABO\_DEST » ou un même code prélèvement commanditaire si l’application ayant générée les codes prélèvements est différente de celle qui intègre le « LABO\_DEST » ;
* Une analyse d’un même code prélèvement ayant un paramètre, une fraction, une méthode (inconnue ou non gérée) et une unité (inconnue) programmés ;
* Une condition environnementale d’un même prélèvement sur un même paramètre ;
* de LQ : le renseignement de la LQ dans la balise <LQAna> est rendu obligatoire par l’AELB et contrôlé par le module de validation de LQ. Celui-ci vérifie la cohérence de la LQ rendue pour chacun des résultats physico-chimiques avec son code remarque respectif égal à 10.



Les autres cas possibles de contrôles de respect de LQ sont prévus et détaillés dans l’autre prestation : Contrôles qualité complémentaires sur les données (cf. §IV-3);

* + - Autres types d’incohérence en lien avec les données métiers :

Le bon calage de tous ces référentiels métiers AELB est prépondérant pour que les contrôles de validité spécifiquement appliqués aux données soient cohérents et adaptés à :

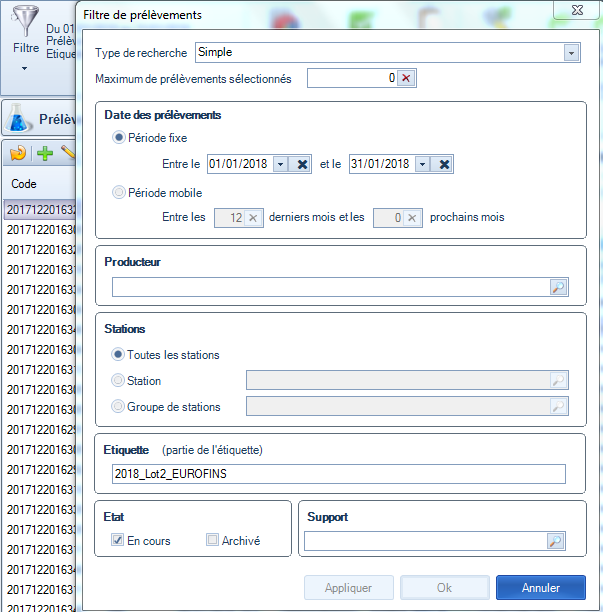
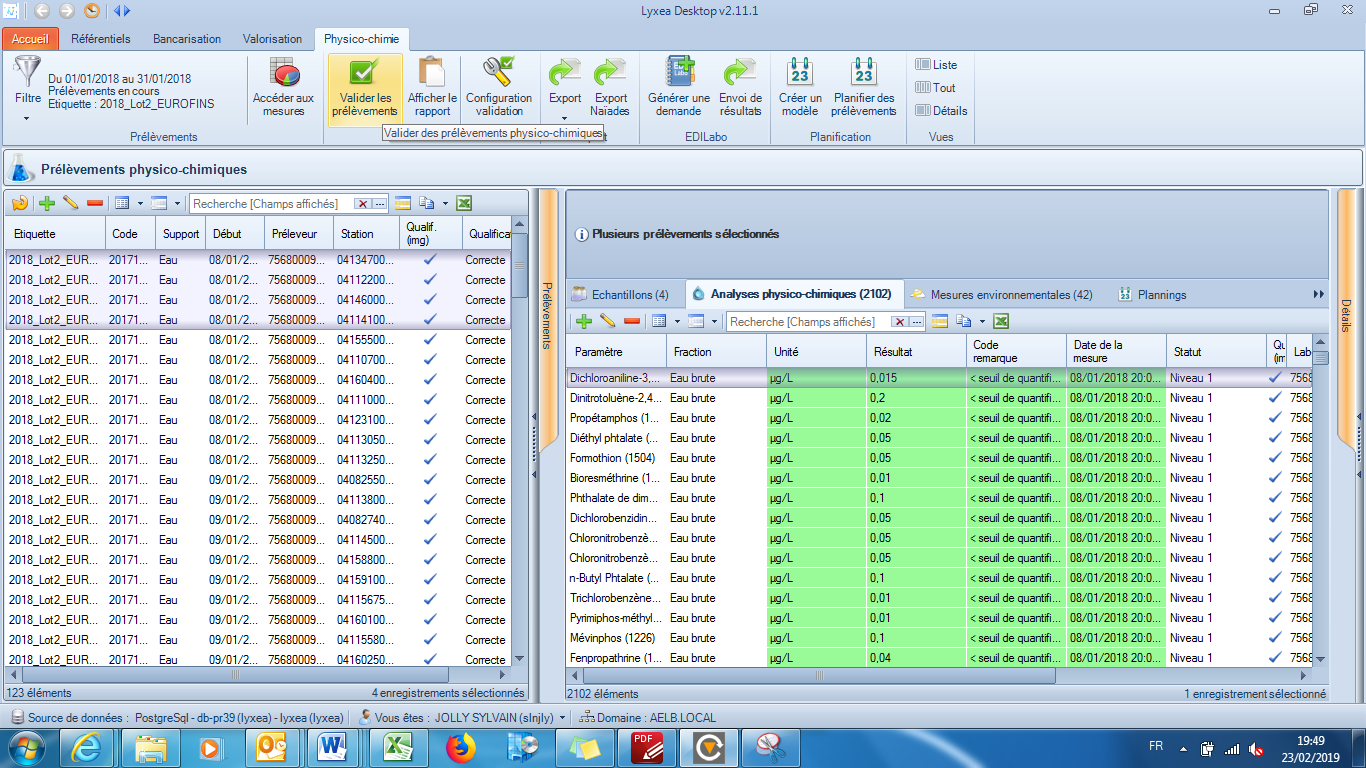
* la vérification du résultat d’analyse, du code remarque, des limites analytiques (LS, LD, LQ), de l’unité et de leur cohérence ;
* la cohérence de la date de l’analyse, de la date de la condition environnementale par rapport à la date de début et fin de prélèvement.

Les tests de validation unitaires avec leur niveau de criticité associé (cf. annexe 7a) sont à appliquer par le coordinateur sur chaque fichier de données. Ces tests de validation seront précisés à la réunion de démarrage et révisables chaque année.

Ils sont tous disponibles dans l’application de l’AELB qui compte au total 83 contrôles de validation. Le coordinateur s’assurera que tous ces tests demandés par l’AELB (évolutifs sur la période d’exécution du marché) sont bien activés dans l’application.

Le coordinateur sera d’autant plus vigilant sur la configuration exhaustive de ces tests de validation pour un fichier de données PC à MO AELB que beaucoup de ces tests sont débrayés pour le traitement d’un fichier de données PC d’autres maitres d’ouvrage

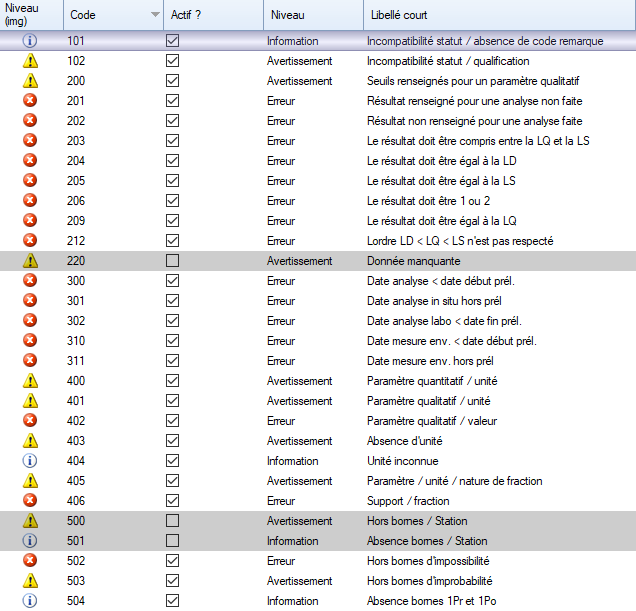
Afin de simplifier le contrôle de validation des données en masse, le coordinateur filtre puis sélectionne un lot de prélèvements et actionne la validation automatique des analyses sous-jacentes (cf. Figures ci-dessous).



Dans l’exemple ci-dessus, le gestionnaire de données a filtré les prélèvements de Janvier 2018 réalisés par EUROFINS sur le lot 2.

L’AELB demande au coordinateur d’étiqueter les prélèvements traités dans l’application de gestion afin d’apporter une souplesse de gestion et une indexation explicite pour des sélections rapides et d’éventuelles mises à jour à posteriori. Après sélection des prélèvements avant import, le coordinateur leurs attribue une étiquette du type : « annee\_LotX\_nomlabo\_support\_mois ».

Pour les paramètres hydrobiologiques, le nombre de contrôle est seulement de 25 tests de validation pour un fichier de données HBIO MO AELB et partenaires, et de 24 pour un fichier de données HBIO d’autres suivis locaux (cf. capture ci-dessous)



Le contrôle de validation induit automatiquement un changement du statut de la donnée qui passe de « donnée brute » à « donnée de niveau 1 » et de la qualification de ces mêmes analyses et conditions environnementales. La qualification devient « correcte », « incorrecte », « incertaine » ou « non définissable » à la place de la qualification « non qualifié ».

Le statut d’une donnée détermine son état dans son cycle de traitement. Le coordinateur vérifie les statuts particuliers attribués aux données selon les possibilités du SANDRE :

* « Donnée brute » : Données issues du processus d'acquisition n’ayant pas subi d’examen. Ce sont celles transmises directement par le laboratoire et qui ne doivent plus avoir ce statut après le contrôle validation ;
* « Donnée contrôlée niveau 1 » : le coordinateur expertise les résultats par rapport à des contrôles automatisés dans l’application de gestion (bornes min-max classiquement rencontrées sur le point, tests inter-paramètres, …) et apporte une analyse critique des données par rapport à la connaissance métier sur cette station. **Pour garantir la qualité des données bancarisées, le coordinateur expose dans son offre les plus-values techniques qu’il peut apporter.**
* « Donnée contrôlée niveau 2 » : le coordinateur vérifie l'ensemble de la chaîne d'acquisition et la cohérence des données (par exemple : depuis le prélèvement, conditionnement, flaconnage, transport, mesure en laboratoire). L’AELB apporte aussi en complément son expertise technique. Au terme des réunions de qualification annuelles, le coordinateur doit attribuer le statut « Donnée contrôlée niveau 2 » à toutes les données.
* « Donnée interprétée » : la valeur a été utilisée dans un rapport ou valorisée. Cette mise en perspective de l'information permet de consolider son niveau de validité et détecter les dernières erreurs. Par exemple : diagrammes binaires, comparaison faciès, etc.

L’AELB ou un expert peut attribuer une qualification manuelle à la donnée dont le statut devient alors «Donnée contrôlée niveau 2 » ou « Donnée interprétée ».

* **Qualification des données**

La qualification décrit la qualité de la donnée une fois contrôlée :

* Non qualifié : État initial de la mesure qui n'a pas encore subi d’audit ou d’interprétation de la part de l’AELB ou du coordinateur en vue de sa validation. Correspond à un statut « Donnée brute » ;
* Incertaine : une valeur sera déclarée « Incertaine » si la validité de la donnée reste « douteuse » au stade de validation indiquée dans l'information « statut de la donnée ». D’après le SANDRE, la qualification « Douteuse » doit être une étape transitoire de la validation de la donnée et doit être réservée à des avancements intermédiaires de la validation. Dans les faits, l’AELB n’applique pas forcément cette règle et conserve la qualification incertaine.
* Incorrecte : Une valeur est déclarée « Incorrecte » lorsqu’elle est estimée erronée au stade de validation indiqué dans l'information « statut de la donnée » et vis-à-vis de la finalité recherchée. Par exemple, une donnée hors limites de ces bornes d’impossibilité ;
* Correcte : Une valeur est déclarée « Correcte » lorsqu’elle est estimée valide au stade de validation indiqué dans l'information « statut de la donnée » et vis-à-vis de la finalité recherchée ;
* Qualification non définissable : Une valeur sera non définissable lorsque l’AELB et le coordinateur sont dans l'impossibilité d'obtenir les informations nécessaires pour évaluer la conformité de la donnée. Il s'agit par exemple de données historiques récupérées des archives dont toutes informations sur le processus de production ont été perdues.

L’outil de validation des données permet de qualifier automatiquement chaque donnée, puis chaque prélèvement présent en base en effectuant une batterie de tests. Le coordinateur n’aura qu’à contrôler en masse les qualifications attribuées et se contentera d’interventions minimes à cette phase, sauf indications contraires de l’AELB.

À l’issue de ce contrôle de validation, la qualification des prélèvements et des mesures est identifiée à partir des icônes suivantes :



La qualification des données diffère selon les cas :

Cas 1 : au moins un avertissement confère une qualification incertaine à la donnée ;

Cas 2 : au moins une erreur confère une qualification incorrecte à la donnée ;

Cas 3 : les informations ne dégradent pas la qualification de la donnée.

Dans l’application, le prélèvement hérite de la qualification la plus déclassante de ses mesures.