

OSR4 2015-2017

Outils numériques et
bancharisation de données



Observatoire
Sédiments
Rhône

Action V.4

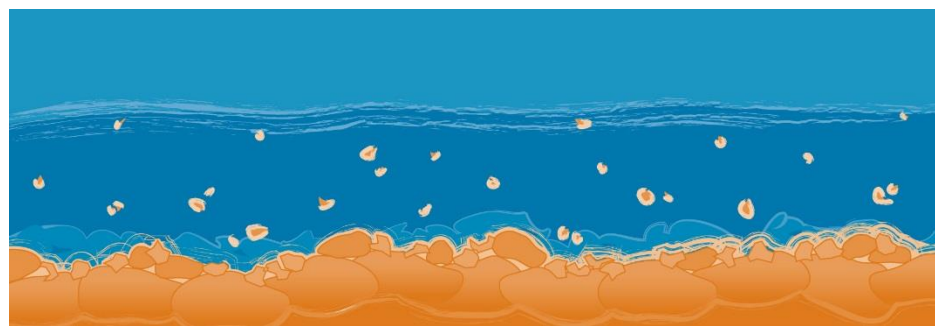
Rapport final, décembre 2017



2015 - 2020



Donnons un avenir à notre fleuve



Personnes impliquées	Equipe de recherche
C. Le Bescond, F. Branger, F. Thollet, J. Le Coz	Irstea – UR HHLY, Centre de Lyon-Villeurbanne
G. Pérreal, N. Raidelet	Irstea – Pôle Informatique Scientifique, Centre de Lyon-Villeurbanne

Outils de stockage et visualisation des données du réseau OSR d'observation des flux de matières en suspension et de contaminants particulaires (OSR 4)

La Base de Données pour les Observatoires en Hydrologie (BDOH)

Résumé

La Base de Données pour les Observatoires en Hydrologie (BDOH) a pour vocation de permettre la gestion, la bancarisation et la mise à disposition des données hydrologiques et biogéochimiques issues des observatoires de long terme. La base BDOH / OSR flux permet de capitaliser les données de flux particulaires et de contaminants pour l'ensemble du réseau d'observation de l'OSR. C'est à la fois un outil d'échange de données pour les scientifiques et aussi de mise en commun de ces données grâce, par exemple, à la gestion intégrée (et tracée) des calculs de flux de contaminants.

Développée à Irstea pour les besoins des chercheurs, BDOH est une base de données évolutive. Ce rapport fait état pour chaque année de l'avancement des développements effectués au cours du programme OSR4.

Mots-clés

OSR, BDOH, données, bancarisation, mise à disposition, calculs de chroniques dérivées

Table des matières

1	Présentation de BDOH	5
2	Vie de la base de données BDOH	7
2.1	Evolutions de BDOH en 2015	7
2.2	Evolutions de BDOH en 2016	7
2.3	Evolutions de BDOH en 2017	8
2.4	Perspectives et développements futurs	8
	Références	9

1 Présentation de BDOH

La Base de Données des Observatoires en Hydrologie (BDOH) a été développée par Irstea pour gérer, bancariser et mettre à disposition des séries temporelles de données hydrologiques et biogéochimiques issues des observatoires de long terme [Branger *et al*, 2014].

BDOH n'est pas une base de données « centralisée ». Les données sont organisées par Observatoires, qui sont le pendant des structures administratives de production et gestion de données (exemple : OSR). Dans un Observatoire, l'entité élémentaire de stockage de la donnée est la Chronique (voir figure 1), qui correspond à une série temporelle régulière ou irrégulière (pas de temps fixe ou variable) et monovariée (un seul paramètre et une seule valeur numérique par pas de temps). Une Chronique est rattachée à une Station, qui est un point de mesure localisé à la surface du sol et autour duquel sont réalisées les mesures. Des données de structure complexe telles que des champs 1D (humidités du sol à différentes profondeurs), 2D (pluies radar) ou 3D ne peuvent pas être bancarisées telles quelles et doivent être décomposées en multiples Chroniques ponctuelles. Outre les séries temporelles, BDOH permet aussi de charger un nombre restreint de données cartographiques : position des Stations, tracé des cours d'eau et contours des bassins versants.

Les données issues du réseau d'observation des flux particuliers et de contaminants associés obtenues dans le cadre du programme de l'OSR sont stockées dans la base de données BDOH et accessibles sur une page spécifique à l'adresse suivante :

<https://bdoh.irstea.fr/OBSERVATOIRE-DES-SEDIMENTS-DU-RHONE/>

BDOH permet de calculer des chroniques de flux particuliers et de contaminants associés en combinant les mesures de débit, de matières en suspension (MES) et de contaminants particuliers issues d'une même station, ou d'une station proche par propagation du débit par exemple. Toutes les actions suivantes sont effectuées directement dans BDOH :

- Chaque chronique de turbidité est convertie en chronique de concentrations en MES par l'application d'un barème de conversion turbidité/MES établi pour chaque station et chaque capteur de turbidité. Les barèmes de conversion utilisés sont également stockés dans BDOH pour une période d'application et une plage de validité définies et la liste de barèmes peut être complétée lorsque la courbe de calibration turbidité/MES du capteur de turbidité est modifiée ;
- La chronique de concentration en MES, calculée à partir de la chronique de turbidité, est multipliée par la chronique de débit à la station correspondante pour obtenir une chronique de flux de MES. Les deux chroniques mères sont mises au même pas de temps par interpolation linéaire à la seconde près entre deux points ;
- Une chronique continue des concentrations en contaminant associé aux MES doit être établie à partir de mesures et d'hypothèses. En général, il est préférable de la construire à partir des résultats du piège à particules, avec vérification sur des prélèvements ponctuels (centrifugeuse), ou de considérer des teneurs moyennes par année ou par type de régime hydrologique. Puis, une chronique de flux de contaminant particulaire est obtenue en multipliant cette chronique continue de concentration en contaminant particulaire avec la chronique de flux de MES.



Métadonnées, généalogie de la chronique

Taux de remplissage des données

Visualisation des données

Figure 1 : Visualisation de la fiche chronique « concentration en MES sur l’Arve à Genève ».

Les chroniques de flux calculées sont consultables directement à travers l’interface Web. L’utilisateur peut définir la période sur laquelle il souhaite calculer un flux instantané ou un cumul de MES et/ou de contaminants associés.

Des conventions bilatérales entre Irstea et chaque producteur de données ont été établies pour définir les modalités de mise à disposition de ces données.

L’avancement de la bancarisation dans la base BDOH/OSR des données issues du suivi particulière et de contaminants est présenté dans le livrable sur l’action III.1 (C. Le Bescond et Al, 2017).

2 Vie de la base de données BDOH

BDOH est un produit Irstea conçu par ses équipes scientifiques et techniques et développé par son Pôle Informatique Scientifique. L'outil reste donc évolutif et de nouvelles fonctionnalités pour les besoins des scientifiques ont été implémentées au cours du programme OSR4. Un groupe de travail reste actif pour assurer le lien entre les développeurs et les utilisateurs de la base. Il est composé de membres de l'équipe informatique, des utilisateurs les plus actifs en termes d'alimentation de données et des coordinateurs scientifique et technique qui gèrent le projet depuis son lancement.

2.1 Evolutions de BDOH en 2015

En 2015, les évolutions suivantes ont été apportées :

- Mise en place d'une gestion des DOI (Digital Object Identifier) sur le jeu de données BDOH / OSR. Un DOI est une identification unique et pérenne qui permet la citation des jeux de données dans les bases de données bibliographiques et les citations d'article. C'est un enjeu important d'une manière générale pour tous les producteurs de données expérimentales, car cela ouvre la possibilité de faire reconnaître la production de données en tant que telle comme activité "publiante".
- Formalisation des règles de gestion de codes qualités supplémentaires, nécessaires à l'établissement des flux OSR : codes LQ/LD/E pour une concentration inférieure à la limite de quantification ou de détection, et des données estimées (ou reconstituées), avec les règles de propagation aux chroniques-filles de flux (implémentation prévue en 2016).
- Formalisation de règles communes OSR pour les calculs de chroniques de flux instantané et de cumulés avec notamment un code R développée pour la conversion d'une chronique discontinue en chronique continue ; l'implémentation est envisagée dans BDOH.
- Correction de bugs et maintenance de l'application.

2.2 Evolutions de BDOH en 2016

L'outil BDOH permet désormais de bancariser des données reconstituées et inférieures aux limites de quantification ou de détection. Depuis juin 2016, BDOH dispose de 3 nouveaux codes qualité permettant la gestion de données estimées et de données inférieures aux limites de quantification ou de détection :

- E pour Estimé, la donnée manquante (lacunaire ou invalidée) a été reconstituée à l'aide d'une équation mathématique qui relie le paramètre manquant à un autre paramètre (relation entre le débit et la concentration en MES par exemple) ou d'un modèle numérique (modèle hydraulique Rhône 1D par exemple) ;
- LQ pour Limite de Quantification, la valeur est inférieure à la limite de quantification de la méthode d'analyse utilisée ;
- LD pour Limite de Détection, la valeur est inférieure à la limite de détection de la méthode d'analyse utilisée.

Pour prendre en compte les évolutions récentes de la base de données (gestion de DOI, nouveaux codes qualités...), la documentation utilisateur de l'interface a été mise en jour. Il est prévu de la rendre disponible en ligne pour l'ensemble des utilisateurs.

2.3 Evolutions de BDOH en 2017

En 2017, le pôle Informatique Scientifique d'Irstea a travaillé sur les montées de version des briques logicielles utilisées pour l'interface Web (logiciels libres). Ces mises à jour ont bousculé le code, ce qui a parfois nécessité une réécriture complète de certaines fonctionnalités.

Ce travail a été ou sera mis à profit pour apporter des améliorations ou répondre à des évolutions demandées comme par exemple :

- la prise en compte des heures dans la gestion des barèmes de conversion et une nouvelle présentation de la page de gestion des barèmes;
- la visualisation des valeurs inférieures aux limites de détection et de quantification par un symbole ou une couleur différente dans le visualisateur simple ;
- une meilleure gestion des imports en cas d'erreur ;
- la réécriture des briques de code pour l'export de données et l'interpolation ;
- l'amélioration du recalcul des taux de lacunes ;
- la correction d'erreurs dans la gestion des droits.

Les montées de version des logiciels ont également bouleversé la mise en forme des pages d'administration ; leur code a dû être réécrit pour être plus robuste et résistant aux mises à jour des logiciels. Suite au départ des agents contractuels de l'équipe de développement, la gestion de BDOH a été confiée au personnel permanent ce qui, à terme, devrait contribuer à améliorer la pérennité du code lors de futures évolutions.

Toutes ces évolutions ont dans un premier temps été apportées au code de la base de test, qui est une image de la base de production de BDOH. Les utilisateurs testent régulièrement ces nouvelles fonctionnalités pour vérifier leur adéquation avec celles proposées dans la base de production actuelle et faire remonter d'éventuelles erreurs à corriger. Lorsque le niveau de la base de test sera jugé suffisant, les évolutions seront intégrées à la base de production.

En 2017, le service de Forge Logiciel pour la capitalisation et le traitement des évolutions techniques a été remplacé par l'interface Gitlab (<https://gitlab.irstea.fr/pole-is/bdoh>).

2.4 Perspectives et développements futurs

L'utilisation importante de l'outil BDOH au sein des observatoires de long terme et la diversification des données qui y sont stockées nous conduit à y apporter régulièrement des corrections de bugs, des améliorations modestes et parfois de nouvelles fonctionnalités pour mieux répondre au besoin des utilisateurs (ex : gestion des LQ / LD). Aussi leurs retours d'expérience auprès des gestionnaires de la base sont-ils précieux pour faire progresser cet outil (contact : bdoh.support@lists.irstea.fr).

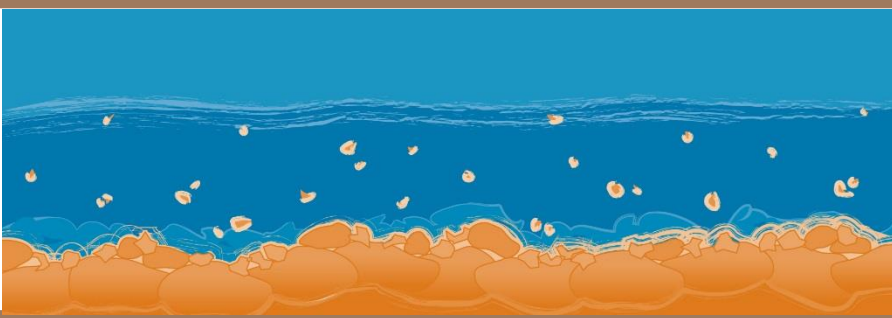
Ainsi, la poursuite des développements de fonctionnalités est prévue notamment sur les calculs et la visualisation de flux mais aussi sur l'ergonomie de l'interface web :

- possibilité de hiérarchiser l'affichage des chroniques en liste pour chaque station (par ex : par famille de contaminant), à utiliser aussi pour sélectionner les chroniques mères pour le calcul des chroniques filles,
- affichage différent des mesures inférieures aux LQ / LD et Estimées,
- calcul du ratio pourcentage de mesures « vraies » / mesures estimées dans les flux cumulés,
- proposer une statistique du nombre de mesures par mois pour les chroniques discontinues (sur l'exemple des camemberts pour les chroniques continues),
- calcul automatique des dates début/fin de la chronique-fille pour prendre en compte les périodes de recouvrement des chroniques-mères,
- transformation de chroniques discontinues en chronique continue, applicable sur des lots de chroniques. Un programme R extérieur à la base est actuellement utilisé ; son implémentation dans BDOH permettrait de le déployer pour l'ensemble des utilisateurs,
- une amélioration importante est attendue pour le séquençement des calculs effectués par la base, sous forme de « Jobs » dont l'état sera visible et archivé. Une réflexion entre équipes de développement et gestionnaires de la base a été menée en 2017 : un cahier des charges précis a été rédigé et l'implémentation sera conduite au cours du programme OSR5. Cela permettra d'organiser les calculs (interpolations de pas de temps à l'export, calculs de chroniques dérivées) qui peuvent être d'autant plus gourmands en puissance que la quantité de données stockées dans BDOH a fortement augmenté au cours des deux dernières années. De plus, l'absence de message indiquant que le calcul est en cours peut faire croire à l'utilisateur que le lancement n'a pas fonctionné et le conduire à relancer plusieurs fois le même calcul, saturant complètement la base. En 2016 et 2017 nous avons eu plusieurs alertes de ce genre. Le séquençement des jobs permettra de gérer cela de façon efficace, et aussi de lancer automatiquement le calcul de chroniques filles lors de la mise à jour d'une chronique mère (cas typique des flux OSR), ce qui facilite grandement le travail des opérateurs. Par exemple, si on ajoute les derniers enregistrements de turbidité à une chronique existante, alors toutes les chroniques dérivées de cette chronique pourront être étendues automatiquement : chronique de concentration en MES, de flux de MES, et de flux en tous les contaminants particuliers dont la teneur est bancarisée. Ces mises à jour se font fastidieusement à la main actuellement.

Références

Branger et al (2014). Le projet Base de Données pour les Observatoires en Hydrologie : un outil pour la bancarisation, la gestion et la mise à disposition des données issues des observatoires hydrologiques de long terme à Irstea, Houille Blanche-Revue Internationale de l'Eau, vol. 1, p. 33-38, doi: 10.1051/lhb/2014005

Le Bescond et al (2017). Rapport sur le fonctionnement du réseau OSR d'observation des flux de matières en suspension et de contaminants particuliers et sur la bancarisation des données pour l'année 2017 (OSR 4)



Observatoire des Sédiments du Rhône



Provence-Alpes-Côte d'Azur



Observatoire des Sédiments du Rhône

GRAIE – OHM Vallée du Rhône

66 bd Niels Bohr – CS 52132

69603 Villeurbanne Cedex

www.graie.org/osr/

Contact : dad.roux@graie.org