

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

Flora BRANGER^{1*}, Fabien THOLLET¹, Maxime CROCHEMORE², Mathieu POISBEAU², Nicolas RAIDELET²,
Pierre FARISSIER³, Mickaël LAGOUY¹, Guillaume DRAMAIS¹, Jérôme LE COZ¹, Alain GUÉRIN³,
Gaëlle TALLEC³, Julien PESCHARD³, Nicolle MATHYS⁴, Sébastien KLOTZ⁴, Mathieu TOLSA⁵

¹ Irstea, UR HHLY, Hydrologie-Hydraulique, 5 rue de la Doua, CS 70077, 69626 Villeurbanne Cedex

² Irstea, DSI, pôle Informatique Scientifique, 5 rue de la Doua, CS 70077, 69626 Villeurbanne Cedex

³ Irstea, UR HBAN, Hydrosystèmes et Bioprocédés, Parc de Tourvoie BP 44, 92163 Antony Cedex

⁴ Irstea, UR ETGR, Erosion Torrentielle, Neige et Avalanches, BP76, 38402 Saint Martin d'Hères

⁵ Irstea, GR RHAX, Groupe de Recherche en Hydrologie, 3275 route Cézanne CS 40061, 13182 Aix en Provence Cedex 5

RÉSUMÉ. – Les observatoires de long terme en environnement permettent d'acquérir une grande quantité de données. Se posent alors les questions de leur bancarisation, leur gestion par les équipes productrices, et leur mise à disposition sur Internet. Le projet BDOH (base de données pour les observatoires en hydrologie) vise à développer pour cela un outil performant, pour les données issues des observatoires hydrologiques d'Irstea. BDOH permet notamment : la visualisation graphique et le téléchargement en temps réel des chroniques, l'import et export de données sous forme de fichiers texte à différents formats avec possibilité d'interpolation du pas de temps, le calcul automatique de données indirectes via des lois de transformation de type courbes de tarage, la gestion de plusieurs niveaux d'utilisateurs avec des accès et droits différenciés, une interface cartographique conviviale.

Le projet implique 4 équipes / unités de recherche d'Irstea associées au pôle Informatique Scientifique de la Direction des systèmes d'information, qui assure le développement informatique. Une pré-version est actuellement en cours de test.

Mots-clés : observation de long terme, chroniques hydrologiques, gestion, mise à disposition, base de données

Database for hydrological observatories: a tool for storage, management and access of data produced by the long-term hydrological observatories of Irstea

ABSTRACT. – The long-term environmental observatories produce a great quantity of data. These data must be stored, managed properly and accessible to our research partners. The BDOH project aims at developing a database tool for the data produced by Irstea's hydrological observatories. It should allow real-time visualization and download of data, data upload and download as text files with various formats and options for time step interpolation, automatic calculation of derived data using transformation laws, easy browsing through a user-friendly cartographic interface.

This project involves 4 research teams/ units at Irstea along with the Pôle Informatique Scientifique which is in charge of application development. A pre-version of BDOH is currently being tested.

Key-words: long-term observation, hydrological time series, data management, data availability, data base

I. INTRODUCTION

Irstea (anciennement Cemagref) gère depuis de nombreuses années plusieurs observatoires hydrologiques de long terme. Il s'agit de sites expérimentaux de terrain sur lesquelles sont réalisées en continu ou lors de campagnes récurrentes des mesures de pluviométrie, hauteurs d'eau et débits dans les cours d'eau, niveaux de nappes, flux de matières en suspension, concentrations en diverses substances etc, le tout en plusieurs points dans l'espace. Parmi ces observatoires, certains fonctionnent ainsi depuis de nombreuses années (voir Figure 1) : depuis 1962 pour l'Orgeval (GIS ORACLE), 1966 pour le Réal-Collobrier, 1984 pour Draix (GIS Draix-Bléone), 1997 pour le bassin de l'Yzeron (dans le cadre de l'Observatoire de Terrain en Hydrologie Urbaine)...

Ces données sont utilisées à des fins scientifiques par les chercheurs d'Irstea et leurs partenaires, ainsi que par la communauté opérationnelle de l'environnement publique ou privée (services de l'Etat, collectivités territoriales, bureaux d'études). Les questions associées à la production de ces données en grande quantité sont celles de leur bancarisation, leur gestion par les équipes qui les produisent, et in fine leur mise à disposition sur Internet pour l'ensemble de la communauté d'utilisateurs. A l'heure actuelle, les différentes équipes scientifiques gestionnaires de ces observatoires ont des pratiques diverses dans ce domaine, dépendant notamment des outils dont elles disposent. Il peut s'agir soit d'autres systèmes de bases de données (base BICHE pour les données de HHLY¹,

1. <http://biche.lyon.cemagref.fr/>

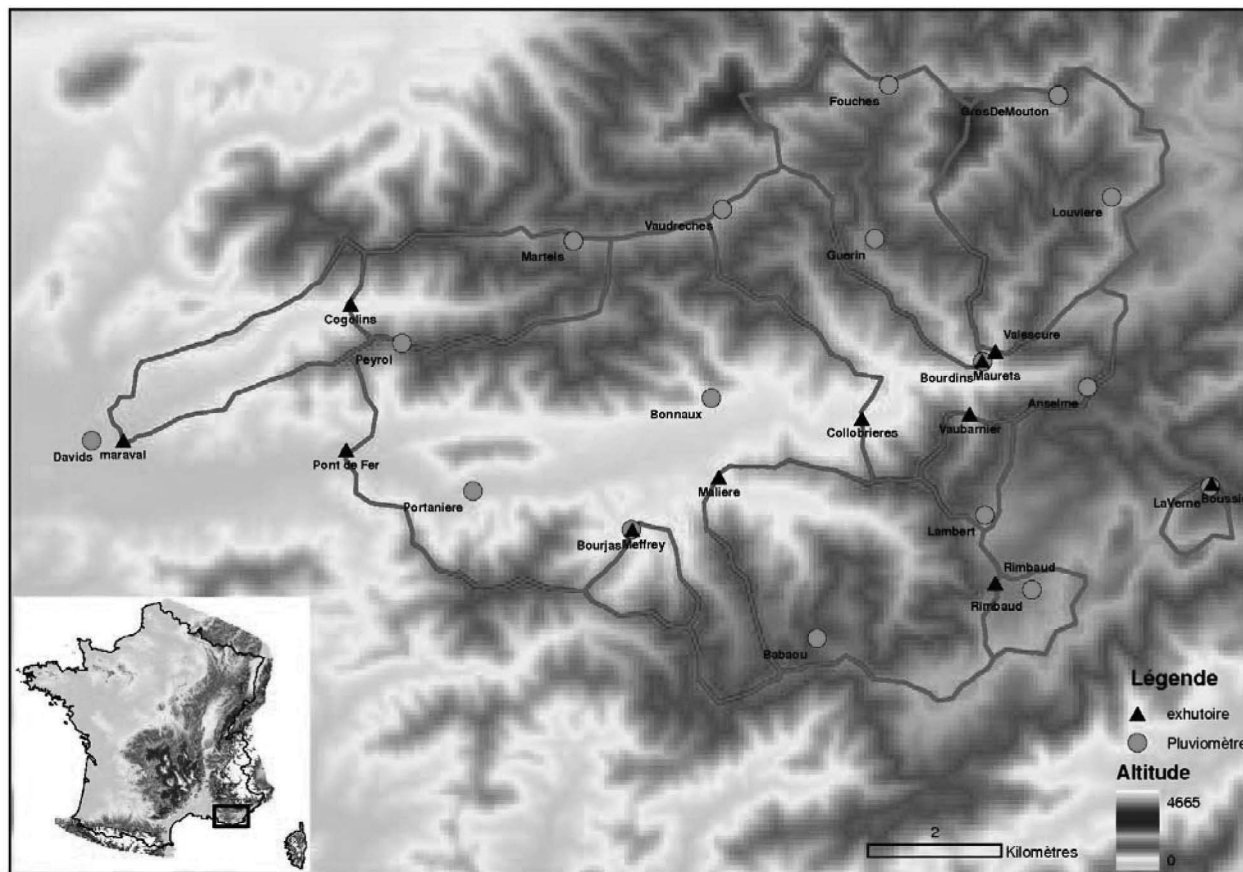


Figure 1 : Carte de situation d'un des observatoires d'Irstea : le Réal Collobrier.

base BD Oracle pour les données du GIS Oracle²), soit de systèmes de fichiers avec par exemple accès ftp pour la mise à disposition.

Il existe plusieurs outils de gestion et mise à disposition de séries temporelles hydrologiques, utilisés par des organismes de recherche ou des gestionnaires, tels que les bases Hysae et Agrhys de l'INRA³, les bases AMMA et AMMA-CATCH⁴, la banque HYDRO⁵ en France. En Allemagne, le système RBIS⁶ (River Basin Information System) développé à l'université de Jena a également été étudié. Chacune de ces bases de données a été développée pour des objectifs spécifiques qui ne recoupent pas totalement le besoin des équipes d'Irstea. Par exemple, les bases Hysae, Agrhys, AMMA ou AMMA-CATCH ne permettent pas le téléchargement de données en temps réel, quels que soient le pas de temps et la longueur de la chronique. L'utilisateur récupère donc les données en différé (sous forme de fichiers sur un site ftp). Le système RBIS offre cette possibilité, mais ne peut pas gérer de données à pas de temps variable, et n'intègre pas de notion de validation de la donnée (code qualité attribué à chaque mesure), qui est un aspect important de la pratique de nos équipes. Hysae intègre une fonctionnalité de gestion de matériel (suivi des capteurs) qui pour Irstea fait double emploi avec le système de gestion de matériel déjà déployé dans l'établissement. Les bases AMMA et AMMA-CATCH ont également une approche

de la gestion des données très structurée par les instruments de mesure, qui ne correspond pas nécessairement à la vision d'Irstea privilégiant plutôt la répartition spatiale des points de mesure dans les bassins versants. Des fonctionnalités plus poussées en hydrométrie, telles que la gestion des courbes de tarage et le calcul automatique de données dérivées (débits à partir de hauteurs d'eau), ne sont disponibles que dans la banque HYDRO, qui par contre n'a pas vocation à bancariser d'autres types de données acquises sur les observatoires hydrologiques (pluviométrie, piézométrie, concentrations en diverses substances). Enfin, un aspect important du développement de ce type d'outil est son évolutivité et sa maintenance. Ayant vocation à être utilisées sur le long terme, ces bases de données doivent avoir un bon potentiel d'évolutivité. Cela est grandement facilité par la maîtrise des codes sources de l'application. Le partage et l'adaptation des codes source d'outils existants étant une question complexe (juridiquement et techniquement), il peut s'avérer plus simple de lancer un nouveau développement. Sur ce constat, le projet BDOH (Base de Données pour les Observatoires en Hydrologie) a donc été lancé de façon à développer un outil performant, répondant aux besoins exprimés par les équipes, pour la plupart des données produites par les observatoires hydrologiques de long terme d'Irstea.

II. ARCHITECTURE ET FONCTIONNALITES DE BDOH

Le développement de BDOH tire parti des points forts des outils existants mentionnés plus haut et des expériences passées

2. <http://bdoracle.irstea.fr>

3. <http://w3.rennes.inra.fr/hysae/> et <https://www.umr-lisah.fr/hysae/>

4. <http://www.amma-catch.org/> et <http://database.amma-international.org/>

5. <http://www.hydro.eaufrance.fr/>

6. <http://leutra.geogr.uni-jena.de/saaleRBIS/>

à Irstea pour la bancarisation de données hydrologiques (développement des bases de données BICHE et ORACLE).

II.1. Architecture

Au niveau informatique, BDOH repose sur une architecture en trois couches : base de données/application/client (navigateur web). Comme pour la plupart des outils de ce type développés en France, l'ensemble des technologies retenues est open-source, afin d'assurer une meilleure pérennité de l'outil dans le temps (maintenance, évolution, échange) et des possibilités de partage avec les communautés de développeurs. Le système de gestion de base de données relationnelles PostgreSQL et son extension PostGIS sont utilisés pour le stockage de données. Le langage de développement pour l'application est PHP (version 5), « augmenté » du framework Symfony2. Côté client, le langage Javascript est largement utilisé, ainsi que les bibliothèques suivantes : OpenLayers, jQuery, DataTables et jqPlot.

Le pivot pour le stockage de données dans BDOH est le couple station / chronique (voir Figure 2), où une station désigne un point de mesure localisé dans l'espace et sur lequel sont mesurés un ou plusieurs paramètres physiques. Les résultats de ces mesures sont des chroniques, enregistrées à pas de temps fixe ou variable, qui peuvent être continues ou discontinues. Chaque station est rattachée à un ou plusieurs sites expérimentaux, dans un observatoire. La notion de site expérimental permet de gérer plus finement à la fois la présence éventuelle de plusieurs producteurs de

données sur un même observatoire, mais aussi l'acquisition de données dans des objectifs légèrement différents (par exemple suivi général de très long terme et mise en place de mesures plus intensives sur une zone plus restreinte pendant trois ans pour la durée d'un projet).

Dans une chronique, les données sont stockées sous la forme date – valeur mesurée – code qualité associé (un code par valeur). Une chronique est dite continue si on estime qu'il y a un lien de continuité entre une valeur et la suivante (cas typique des chroniques de pluie et de débit). Dans ce cas, une seule date est spécifiée. Elle est discontinue lorsqu'il n'y a pas nécessairement de lien entre les valeurs prises à deux dates successives (par exemple concentrations issues de prélèvements manuels). Le format de stockage est alors date début – date fin (qui peut être identique à la date début pour un prélèvement ponctuel) – valeur – code qualité. Les codes qualité sont obligatoires. Le jeu de codes qualité peut différer d'un observatoire à l'autre, mais il doit être unique sur tout l'observatoire. BDOH permet aussi, pour les chroniques continues et discontinues, le stockage des incertitudes de mesure (bornes min et max de l'intervalle de confiance à 95%) : une incertitude (un min et un max) peut-être ainsi rattachée à chaque valeur mesurée.

II.2. Principales fonctionnalités

Le cahier des charges de BDOH comprend plusieurs fonctionnalités adaptées aux différents profils d'utilisateurs de l'application.

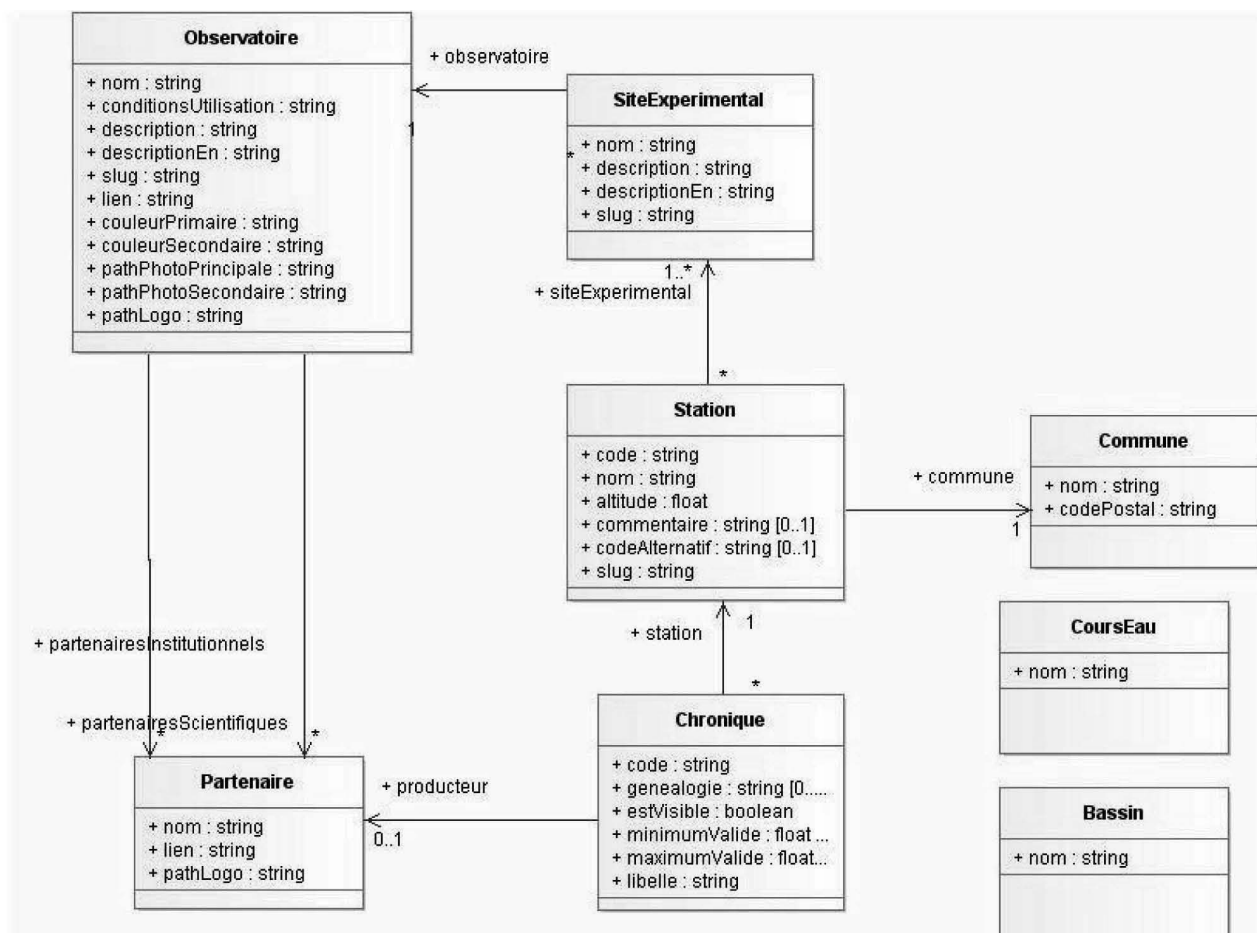


Figure 2 : Extrait du modèle conceptuel de données de BDOH : signalétique d'une station.

Pour la consultation et le téléchargement de données (accès externe), chaque observatoire dispose également de sa propre page d'accueil et d'un accès spécifique aux données. BDOH permet l'affichage des données signalétiques d'une station et d'une chronique de façon à ce que l'utilisateur puisse voir d'un seul coup d'œil quelles sont les données disponibles, sur quelles périodes temporelles et avec quelle continuité pour le cas des chroniques continues (affichage des taux de remplissage des chroniques) (voir Figure 3). La visualisation graphique des chroniques, entre les dates spécifiées par les utilisateurs, est également prévue. Les principales étapes métrologiques de l'acquisition de la donnée sont rappelées dans un cadre généalogie pour aider l'utilisateur à mieux comprendre d'éventuels changements dans la chronique. Le téléchargement des données doit être rapide (temps réel), même pour les très longues chroniques. L'export de données se fait sous forme de fichiers texte (un fichier par chronique). Un format commun dit « format standard BDOH » a été défini. Il reprend le format de stockage mentionné plus haut. L'export dans des formats alternatifs (notamment QTVAR⁷ et QJO⁸ de la banque HYDRO) est également possible. Le cahier des charges prévoit enfin l'export de données interpolées au pas de temps choisi par l'utilisateur (en fonction des possibilités offertes par les gestionnaires des données). Une interface conviviale comprenant un moteur de recherche cartographique est prévue.

En termes de gestion de données (accès interne), BDOH comprend plusieurs fonctionnalités spécifiques : calcul automatique de données indirectes (par exemple débits) via des lois de transformation (par exemple courbes de tarage) dont

l'évolution dans le temps est gérée avec archivage des lois de transformation historiques et traçage des changements. Une journalisation des actions est prévue de façon à pouvoir tracer les différentes opérations réalisées sur un observatoire donné. Une interface spécifique est prévue pour les imports de données dans la base (Figure 4). Plusieurs formats d'import sont possibles, dont les anciens formats utilisés par les observatoires (pour l'initialisation de la base), et pour le fonctionnement sur le long terme le format standard BDOH et les formats QTVAR et QJO de la banque HYDRO. Au niveau administration, BDOH permet de créer plusieurs niveaux d'utilisateurs avec des accès et droits différenciés, aussi bien pour les accès en consultation (accès externes) qu'en administration (accès internes).

III. ORGANISATION DU PROJET ET ETAT D'AVANCEMENT

Le projet implique 4 équipes / unités de recherche d'Irstea associées au pôle Informatique Scientifique de la Direction des systèmes d'information, qui assure le développement de l'application. Le projet a démarré à l'été 2011. Un comité de pilotage se réunit régulièrement pour examiner l'avancement du projet et les moyens à mobiliser.

La première étape du développement a consisté à définir un lexique commun de façon à proposer une architecture de base de données qui convienne à tous. Un catalogue des données à bancariser dans BDOH a également été réalisé, de façon à prévoir l'ensemble des cas de figure nécessaires

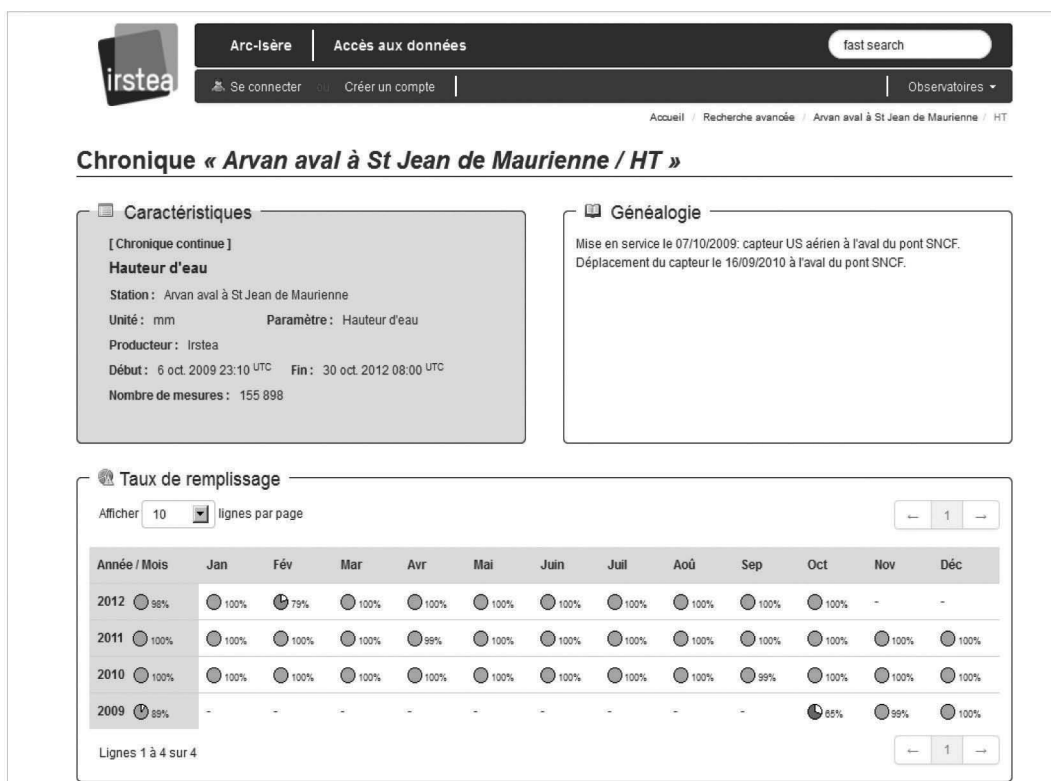


Figure 3 : Exemple de fiche signalétique d'une chronique dans BDOH.

7. Débit à pas de temps variable

8. Débit moyen journalier

Figure 4 : Exemple d'interface pour l'import de données dans une chronique : visualisation des chevauchements entre les données présentes et les nouvelles données importées.

pour le développement de la base de données. Cela a été pour les équipes scientifiques l'occasion d'échanger sur leurs pratiques météorologiques, de critique et de validation de données, et de réfléchir à l'harmonisation des descripteurs de données et des formats d'échange et de stockage (formats de fichiers de données validées, métadonnées à renseigner sur la généalogie des données...).

Le développement proprement dit a commencé en février 2012. Le mode d'organisation est le suivant : la liste des tâches à réaliser et leur ordre de priorité est fixé entre le représentant de la maîtrise d'œuvre et celui de la maîtrise d'ouvrage, puis validé en comité de pilotage. Une rencontre hebdomadaire est fixée entre les développeurs et le représentant des équipes scientifiques pour un suivi technique précis de l'avancement du projet. A chaque étape du développement, l'ensemble des équipes scientifiques est associé lors de séances de travail sur des cas d'utilisation précis : comment créer une nouvelle station ; quelles sont les étapes nécessaires pour l'import de données ; quelles sont les étapes pour le calcul automatique de débit via une courbe de tarage... Cela permet de préciser à la fois les fonctionnalités nécessaires et l'ergonomie des différentes pages web (voir Figures 3 et 4). L'ensemble des informations, compte-rendus de réunions, spécifications techniques, documentation utilisateur etc sont consignés sur le site web du projet, hébergé par la forge logicielle d'Irstea⁹, qui abrite aussi une sauvegarde de l'évolution du code source de l'application. Cela permet d'avoir à disposition en permanence les informations nécessaires au suivi du projet.

Début 2013, une version de test est disponible en interne pour les équipes. Cette version comprend les affichages principaux des signalétiques observatoire, site expérimental, station et chronique, un moteur de recherche avancée pour accéder aux données, un visualisateur simple (pour une

chronique) avec des fonctions de zoom, les imports de données dans l'ensemble des formats prescrits et l'export de chroniques (sans modifications ni interpolations). La gestion des droits d'accès pour les utilisateurs et gestionnaires des observatoires est la dernière fonctionnalité implémentée. Sont prochainement prévus une première intégration des informations géographiques (localisation des stations) dans le moteur de recherche avancée (sans interface cartographique pour l'instant) et l'historisation des actions. Avec cette fonction, une première version de la base de données pourra être ouverte aux partenaires des observatoires. L'ensemble de ces fonctionnalités devrait être opérationnel d'ici mars 2013 avant de passer aux prochaines étapes du cahier des charges. L'ouverture de la base sur internet est prévue pour l'automne 2013 à l'adresse <http://bdoh.irstea.fr/>.

IV. CONCLUSIONS ET PERSPECTIVES

Le projet BDOH est un projet ambitieux qui vise à doter les observatoires hydrologiques de long terme gérés par l'établissement d'un outil adapté à la gestion et à la mise à disposition des longues chroniques de données. Le développement d'un nouvel outil a été jugé nécessaire car à l'heure actuelle les outils existants sont soit limités dans leurs fonctionnalités que ce soit en accès externe (pas d'export de chroniques en temps réel, pas de visualisation graphique des chroniques) ou interne (pas de gestion des transformations) soit restreints à un petit nombre de paramètres (pluie / débit uniquement par exemple). La mobilisation d'informaticiens et de thématiciens (scientifiques gestionnaires des observatoires) permet un développement tenant compte « en temps réel » des demandes et besoins des uns et des autres. Entre équipes gestionnaires des observatoires, cela permet une réflexion commune fructueuse qui va même au-delà du seul projet BDOH, appuyée entre autres par le réseau Mesures

9. <https://forge.cemagref.fr/projects/bdoh>

de l'établissement. Le projet suit son cours et la première version opérationnelle est attendue avec impatience par l'ensemble des équipes.

L'ouverture de l'outil à l'utilisation par d'autres organismes est envisageable sous deux formes. La première est celle de l'hébergement par Irstea des données dans un observatoire dédié. Elle est possible dans la mesure où Irstea est partenaire de l'observatoire et y acquiert aussi des mesures. L'application permet ensuite d'identifier plusieurs producteurs de données.

La seconde forme est la diffusion de l'application elle-même. Techniquement, BDOH est développée en utilisant des logiciels libres ; il n'y a donc pas d'obstacle juridique externe à la diffusion. Les conditions précises de diffusion ne sont pas encore définies à l'heure actuelle et dépendront de l'intérêt des autres organismes et de leurs capacités de développement : cela peut aller de la mise à disposition gratuite de l'application selon une licence logicielle adaptée au partage des codes source, dans un objectif de co-développement.