

De la mesure environnementale à sa diffusion : mise en place d'une chaîne de traitement modulaire et générique pour les données de l'Observatoire de Recherche en Environnement AgrHyS

Geneviève Le Hénaff¹, Hervé Squidant², Ophélie Fovet¹, Mikaël Fauchoux¹,
Yannick Hamon¹, Nicolas Gillet¹, Arnaud Dubreuil³, Laurent Ruiz¹

Résumé. Dans un contexte où nous assistons à la généralisation du concept d'Open Science, l'Open Data et la réutilisation des données constituent un enjeu fort pour la recherche. Nous présentons dans cet article la démarche déployée dans le cadre de l'Observatoire de Recherche en Environnement AgrHyS pour établir une chaîne de traitement automatisée allant de la mesure à la diffusion. Nous détaillons le cadre générique de cette démarche initiée il y a une vingtaine d'années. Nous parcourons les différentes étapes de cette chaîne humaine et technique en retraçant le chemin de la donnée, de l'acquisition à la diffusion. Nous décrivons d'abord les différents développements qui ont permis l'automatisation du transfert, de l'expertise, et de la sauvegarde de la donnée dans une base de données. Nous décrivons ensuite la philosophie et la mise en place d'outils de publication et de réutilisation qui reposent sur une Infrastructure de Données Spatiales et sur le développement de modules complémentaires spécifiques à la dimension temporelle des données de l'observatoire. Enfin, nous expliquons comment le pari initial de 2009 qui consistait à suivre les préconisations de la directive INSPIRE (Trouver, Voir, Télécharger) permet en effet de répondre aux nouvelles exigences du FAIR (Findable, Accessible, Interoperable, Reusable) data.

Mots clés : données, expertise, FAIR, INSPIRE, OGC, interopérabilité



¹ UMR Sol Agro et hydrosystème spatialisé, Inra, 35000 Rennes, France ; genevieve.le-henaff@inra.fr

² Agrocampus Ouest, UMR Sol et hydrosystème spatialisé, Inra, 35000 Rennes, France ; herve.squidant@agrocampus-ouest.fr

³ UMR LISAH, 2 place Viala, 34060 Montpellier cedex 1, France

Introduction

Dans un contexte où nous assistons à la généralisation du concept d'Open Science, l'Open Data et la réutilisation des données constituent un enjeu fort pour la recherche, mobilisant à la fois les producteurs et les utilisateurs de données, ainsi que les informaticiens capables de développer les outils de sauvegarde, d'organisation et de diffusion de données. Cependant la mise en place de structures pour la mise en œuvre de ces concepts se heurte souvent à des difficultés majeures :

- Il est difficile de répondre aux multiples fonctionnalités attendues de ces outils (bancairisation, gestion des données par les producteurs, diffusion aux utilisateurs externes), et souvent les outils développés ne répondent qu'à un seul de ces objectifs, au détriment des autres ;
- l'adhésion des producteurs de données est difficile à mobiliser, en particulier parce qu'ils perçoivent l'injonction de configurer leurs jeux de données pour une activité de diffusion externe comme une charge de travail supplémentaire (et peu reconnue) ;
- la mise à disposition de données expertisées sous-entend une lourde responsabilité pour garantir la fiabilité/validité des données fournies aux utilisateurs, souvent à l'origine de long temps de latence entre production et diffusion de la donnée ;
- si les efforts informatiques et ceux des producteurs de données sont fournis, notre constat est que l'utilisation de ces données diffusées reste modérée donc décevante. Ceci est notamment lié au manque de capacité technique à mettre en place une infrastructure de données ouvertes facile à utiliser.

Cet article propose un cadre générique permettant de lever ces obstacles. Il s'appuie sur une expérience de 20 ans visant à développer le système d'information de l'Observatoire de Recherche en Environnement (ORE) AgrHyS (Fovet et al., 2018). La philosophie de la démarche repose sur les piliers suivants :

1. les outils développés doivent combiner l'ensemble des fonctionnalités attendues (i.e. que les outils utilisés pour bancairiser, traiter et diffuser les données soient intimement liés) et doivent constituer une aide, et non une contrainte, pour les producteurs de données (automatisation des procédures). Ceci n'est possible que si les choix des outils mis en œuvre s'appuient sur les outils préexistants et les contraintes spécifiques des producteurs. Ceci implique donc une capacité de l'outil à évoluer, et une concertation étroite entre métrologues, thématiciens (ex. : hydrologues, biogéochimistes, gestionnaires de l'eau...) et informaticiens ;
2. l'adhésion des producteurs de données doit être favorisée par d'une part le développement d'outils de traitement et de visualisation adossés à l'infrastructure, et d'autre part la transparence et la traçabilité des actions d'expertise de données (corrections, traitement...). La fourniture aux utilisateurs des données à la fois brutes et traitées (avec une qualification) permet de mieux assumer la responsabilité de diffuser des données expertisées ;
3. l'utilisation d'outils conformes aux standards d'interopérabilité doit permettre à la fois une visibilité maximale des données diffusées et un accès aux outils développés par la communauté de l'Open Data. Le "principe de subsidiarité" qui consiste à gérer les outils au plus près possible des producteurs de données, mais aussi à mobiliser les outils ou compétences voisines quand nécessaire, permet dans une large mesure de pallier les limites techniques locales.

Nous présentons ici en détail la chaîne de traitement allant de la mesure jusqu'au développement de services web interopérables pour la diffusion des données produites par l'observatoire AgrHyS. La **Figure 1** illustre l'ensemble de cette chaîne de traitement détaillée dans cet article.

Cette période de développement s'est composée de trois grandes étapes. La première étape débute en 2000, par un effort commun aux unités mixtes de recherche LISAH (Montpellier) et SAS (Rennes) de conception d'une base de données générique et de développement d'outils informatiques. Cette étape a donné naissance aux

premières bases de données des ORE OMERE (Observatoire Méditerranéen de l'Environnement Rural et de l'Eau, Molénat et al., 2018) et AgrHyS. Ces bases de données permettant la sauvegarde, l'organisation et l'exploitation des données, fonctionnaient sur une alimentation manuelle. En 2009, les deux unités se joignent à nouveau pour une deuxième étape visant à développer un outil de valorisation des données Agro-Environnementales existantes (VIDAE) par cartographie des sites d'étude des deux ORE. Le principe de cet outil étant de coupler les fonctionnalités d'un Système d'Information Géographique, l'accessibilité du web, et le géoréférencement. Le développement d'un outil commun devait privilégier des solutions transposables à d'autres contextes et donc à d'autres observatoires. Une philosophie clé de l'outil VIDAE a été le suivi des préconisations de la directive européenne INSPIRE (directive 2007/2/CE) en terme d'interopérabilité et de métadonnées. Ces préconisations étant initialement dédiées à la diffusion des données spatiales, il a été nécessaire de les adapter aux besoins et aux spécificités des données temporelles produites dans le cadre des deux ORE. D'un point de vue opérationnel, cela s'est d'abord traduit par le déploiement de l'Infrastructure de Données Spatiales (IDS) libre, modulaire et interopérable geOrchestra, alors à ses balbutiements ; puis par le développement d'outils spécifiques pensés comme des extensions (addons) à l'IDS. Il s'agissait d'une forme de pari à l'époque. Enfin depuis 2012, face à la montée en puissance des capteurs de terrain pour la mesure hydrométrique, et à la diversification des variables suivies, une troisième étape a donné lieu à plusieurs développements supplémentaires pour i) automatiser l'alimentation de la base de données, et ii) automatiser leur qualification et donc conserver l'expertise apportée sur les données. L'alimentation automatique a permis d'une part une simplification du travail de terrain, mais aussi d'améliorer la fiabilité du traçage de ces données et leur actualisation.

Après une description des données acquises sur l'observatoire et des protocoles d'acquisition dans une première section, sont détaillés dans une deuxième section les outils développés pour automatiser l'archivage des mesures vers la base de données, ainsi que l'expertise apportée. Ensuite sont décrites les différentes stratégies de diffusion des données permettant de répondre aux trois mots clé de la directive INSPIRE : "trouver, voir, télécharger". Enfin, la dernière partie démontre la généricité de la chaîne de traitement proposée à travers ses quelques transpositions à d'autres observatoires.

Description des données acquises et des protocoles de collecte

Présentation de l'Observatoire

AgrHyS (Fovet et al., 2018; Gascuel-Oudoux et al., 2018) est un des Observatoires de la Zone Critique de l'Infrastructure de Recherche Nationale Distribuée OZCAR (Gaillardet et al., 2018) dédié à l'étude des temps de réponse des flux hydrogéochimiques à l'évolution des agro-hydrosystèmes. Piloté par l'Inra depuis 1993, il s'agit d'un observatoire où les variables environnementales sont suivies sur le long-terme, c'est-à-dire plusieurs décennies, en cohérence avec les échelles temporelles des évolutions agricoles, paysagères, et celle de la variabilité climatique.

L'observatoire regroupe plusieurs sites, définis spatialement comme des bassins versants, au sein desquels sont mesurées des variables spatiales, éventuellement dynamiques dans le temps (ex: occupation des sols), et des variables temporelles sur des stations ponctuelles (ex: débit à l'exutoire). Le soutien récurrent de l'Inra permet à l'observatoire AgrHyS de produire un socle de variables suivies sur le long-terme. Certaines variables dont l'acquisition en continue serait trop onéreuse ou moins pertinente (variation lente), sont acquises dans le cadre des projets de recherche successifs qui s'appuient sur les sites de l'observatoire comme sites d'études.

Les données météorologiques sont acquises sur des stations du réseau Climatik de l'Inra. Cette présentation concerne les chroniques temporelles acquises sur les stations ponctuelles, en particulier relatives à l'hydrologie (niveau d'eau, débits, niveau de nappe, humidité des sols...) et la qualité chimique et physico-chimique de l'eau échantillonnée dans la pluie, les cours d'eau ou la nappe (paramètres physico-chimiques, concentrations).

Collecte des données

Deux types de chroniques acquises sur ces stations de mesures se distinguent selon la procédure d'acquisition par capteur ou par prélèvement (1a. **Figure 1**). Les données acquises *via* des capteurs initialement réservés aux variables hydrologiques, regroupent aujourd'hui toutes sortes de variables (également physico chimiques ou chimiques). Elles sont mesurées par des capteurs connectés à des centrales d'acquisition à une fréquence variant de 1 à 30 min. Elles peuvent être associées à tous les types de stations (météorologique, cours d'eau ou piézométrique). Les données associées à un prélèvement d'échantillon peuvent également concerner des variables hydrologiques, physico chimiques, ou chimiques. Elles sont mesurées à partir d'échantillons prélevés sur le site et analysés en laboratoire, ou mesurées par une lecture directe (niveau d'eau, paramètres physico-chimiques mesurés *in situ*) au moment du prélèvement de l'échantillon. Elles sont mesurées à plus basse fréquence : de 2 à 3 fois par mois jusqu'à une fréquence journalière. Elles peuvent être associées à tous les types de stations (météorologique, cours d'eau ou piézométrique).

Transfert des données brutes du terrain au serveur de l'unité

Le transfert des données brutes du terrain vers le serveur de l'unité est réalisé *via* le logiciel Hydras 3 ou des fichiers MS Excel (1b. **Figure 1**). Le logiciel Hydras est utilisé principalement pour la gestion des données mesurées automatiquement et en continu par des capteurs installés sur les sites (données capteur). Les données issues de mesures de laboratoire sur des échantillons prélevés sur le terrain ou de mesures ponctuelles *in situ* (données prélèvements) sont sauvegardées dans des fichiers Excel sur le serveur de l'unité.

Le logiciel Hydras 3 de la société OTT est un outil multiposte de gestion des mesures hydrométriques. Il permet de paramétrer des centrales et des capteurs fabriqués ou commercialisés par la société OTT. Le téléchargement des données mesurées est réalisé soit sur le terrain par connexion avec un PC portable équipé du logiciel, soit à distance *via* des modems M2M. Les données téléchargées peuvent être visualisées et corrigées graphiquement. Un projet accessible uniquement avec le logiciel est créé pour organiser et gérer les données par capteur et par

point de mesure. Les données acquises par des centrales et des capteurs autres que OTT, peuvent être importées manuellement dans les projets du logiciel Hydras. Ceci permet une vision et une gestion globale de toutes les données mesurées automatiquement et en continu sur les sites de l'ORE.

Les données associées à un prélèvement sont soit lues directement *in situ* (sondes multi-paramètre de terrain) soit analysées dans différents laboratoires : principalement les laboratoires de l'UMR Géosciences Rennes, et de l'UMR SAS. Pour les mesures les plus anciennes, il existait également un laboratoire Inra à Quimper en plus de celui de Rennes où étaient analysés les échantillons du site finistérien notamment. Certains types d'analyse enfin sont sous-traités à des laboratoires extérieurs (ex. : isotopie). On associe donc pour chaque paramètre une série de données par laboratoire (**Tableau 1**). Les méthodes de détermination utilisées dans les laboratoires d'analyse sont standard et normalisées. Cependant certains paramètres peuvent être mesurés *via* différentes méthodes comme par exemple les ions calcium (Ca), identifiés par chromatographie ionique lorsqu'on cherche à déterminer les concentrations en cations majeurs, mais par spectrométrie de masse à plasma à couplage inductif (ICPMS) lorsqu'on cherche à déterminer la concentration d'une très large gamme d'éléments incluant les terres rares. On associe donc pour chaque paramètre une série de données par laboratoire et par méthode de détermination (**Tableau 1**). Les métadonnées détaillées concernant la détermination de la concentration pour un point donné sont archivées dans les référentiels et archives réalisés par les laboratoires dans le cadre de leur démarche qualité, et traçables par le nom de la série (paramètre, laboratoire et méthode) et la référence de l'échantillon dans le laboratoire. Les protocoles d'échantillonnages sont renseignés pour chaque donnée (prélèvement manuel, préleveur automatique, préleveur à bille, préleveur à pompe). Les protocoles de filtration ou non filtration de l'échantillon sont liés au paramètre mesuré. Après lecture ou analyses, ces données sont saisies manuellement dans des fichiers Excel d'accès partagés par les opérateurs des différents laboratoires sous forme de tableau. Un fichier correspondant à un site, avec éventuellement différents onglets correspondant aux types de stations (cours d'eau, nappe, pluie).

[Tableau 1 page suivante](#)

Tableau 1. Extrait de la table des types de données de BDDAgrhys illustrant la diversité des types de données (capteur/prélèvement), et la codification des données associées à un prélèvement (famille "chimie"). Ainsi une même grandeur peut être mesurée par un capteur ou associée à un prélèvement (ex : température de l'eau). Les variables dont la mesure est associée à un prélèvement sont identifiées par la combinaison de la grandeur mesurée (ex : concentration en nitrates), de la méthode de mesure (ex : chromatographie ionique), et du laboratoire de mesure (ex : SAS Rennes).

| Nom | Unité | Famille | Mesure | Description | Laboratoire |
|--------------------|-------|----------------|-------------|--|--------------------|
| h | m | hydrologie | capteur | Niveau d'eau dans le cours d'eau mesuré par un capteur à flotteur automatique | n.a. |
| Tw | °C | physico-chimie | capteur | température de l'eau en continu | n.a. |
| Tw in situ | °C | physico-chimie | prélèvement | température de l'eau mesurée <i>in situ</i> par sonde multiparamètre | n.a. |
| NO3-technicon-SASQ | mg/L | chimie | prélèvement | concentration en nitrates mesurée par analyseur Technicon | SAS Quimper |
| NO3-CI-SAS | mg/L | chimie | prélèvement | concentration en nitrates mesurée par chromatographie ionique | SAS Rennes |
| NO3-CI-GR | mg/L | chimie | prélèvement | concentration en nitrates mesurée par chromatographie ionique | Géosciences Rennes |
| Ca-CationBi-SASQ | mg/L | chimie | prélèvement | concentration en calcium mesurée par spectroscopie d'émission | SAS Quimper |
| Ca-ICPMS-GR | µg/L | chimie | prélèvement | concentration en calcium mesurée par spectrométrie de masse par plasma à couplage inductif | Géosciences Rennes |

Sauvegarde des données brutes, des données traitées et de l'expertise associée

La sauvegarde des données brutes est une bonne pratique élémentaire, et une obligation. Ces données brutes, en particulier si elles sont issues de capteurs, nécessitent toutefois un traitement pour être utilisables par la communauté disciplinaire au sens large. Le traitement des données d'observation implique de nombreux opérateurs, techniciens et chercheurs, certains des opérateurs étant non permanents en particulier les doctorants. Ce traitement utilise aussi des outils divers selon les opérateurs, les laboratoires, et évolutifs avec les technologies disponibles. Ces outils divers peuvent être :

- des tableaux associés à des graphiques dans un fichier Excel,
- des logiciels commerciaux développés par les fournisseurs de capteurs (ex. Hydras 3),
- des routines codées par l'opérateur (en VBA, R, Matlab, Python...).

Nous décrivons dans cette section comment l'alimentation automatique de la base de données s'appuie sur les protocoles et outils de traitement utilisés afin de valoriser l'expertise apportée lors des traitements.

Protocoles de traitement des données acquises (2a Figure 1)

Gestion des données capteur

Les données issues de capteurs sont donc traitées à l'aide du logiciel Hydras 3 avec pour chaque site (projet), une région, dans laquelle trois domaines ont été définis en concertation avec l'ensemble des opérateurs impliqués dans la procédure de traitement (environ cinq personnes). Le domaine I correspond aux acquisitions

par les capteurs. Dans le domaine II, les données aberrantes ou erronées sont supprimées, et les dérives de capteurs, corrigées, à l'aide d'un 2^e capteur par exemple. Il s'agit donc de corrections qualifiées de fiables. Le domaine III propose des corrections expertes. Par exemple, pour des fenêtres identifiées erronées à l'étape précédente, une interpolation peut être proposée. Également, certains événements correspondant à des interventions de terrain, comme la purge des piézomètres avant prélèvement d'un échantillon d'eau, y sont supprimées dans le but de fournir des séries nettoyées utilisables pour le calcul de données journalières. La fiabilité de ces corrections est donc plus hétérogène. L'information de l'intervention est par ailleurs traçable par comparaison entre les trois domaines. Le choix de diffuser à la fois l'information brute (domaine I) et l'information traitée (domaine III) a été fait dans un souci de traçabilité et de réutilisabilité : de nouveaux traitements pouvant être proposés *a posteriori* et générer de nouvelles séries de données traitées. Le **Tableau 2** donne quelques exemples des différents traitements pouvant être apportés selon les domaines et de la qualification correspondante associée.

Gestion des données de prélèvements

Les données associées aux prélèvements d'échantillons sont comparées aux gammes habituelles sur les 25 années de suivi et rassemblées dans les fichiers Excel. Si la visualisation de la chronique met en évidence des valeurs en dehors des gammes habituelles auquel cas l'échantillon est ré-analysé. Notons qu'au bout de 25 années de suivi ces gammes sont relativement robustes. Le tableau Excel consigne la valeur de concentration corrigée dans la version traitée de la série, tandis que la valeur initiale est conservée dans la version brute, permettant ainsi de tracer le chemin de l'échantillon. Les données traitées sont qualifiées de fiables, même dans le cas où il y a eu correction.

Transfert des données vers le serveur virtuel, sauvegarde des données et des qualifications (2b Figure 1)

Le nom de domaine Agrhys.fr a été déclaré auprès des services de l'Inra et pointé sur la machine virtuelle (ubuntu 18.04) sur laquelle est hébergée la base de données (BdD) BDDAgrhys sur un serveur Postgresql/Postgis, et les applications développées en PHP 5.5.

La **Figure 2** propose un schéma simplifié de la base de données. L'automatisation de la sauvegarde consiste à synchroniser les différentes étapes de gestion des données du serveur de l'unité vers la base de données (**Figure 3**).

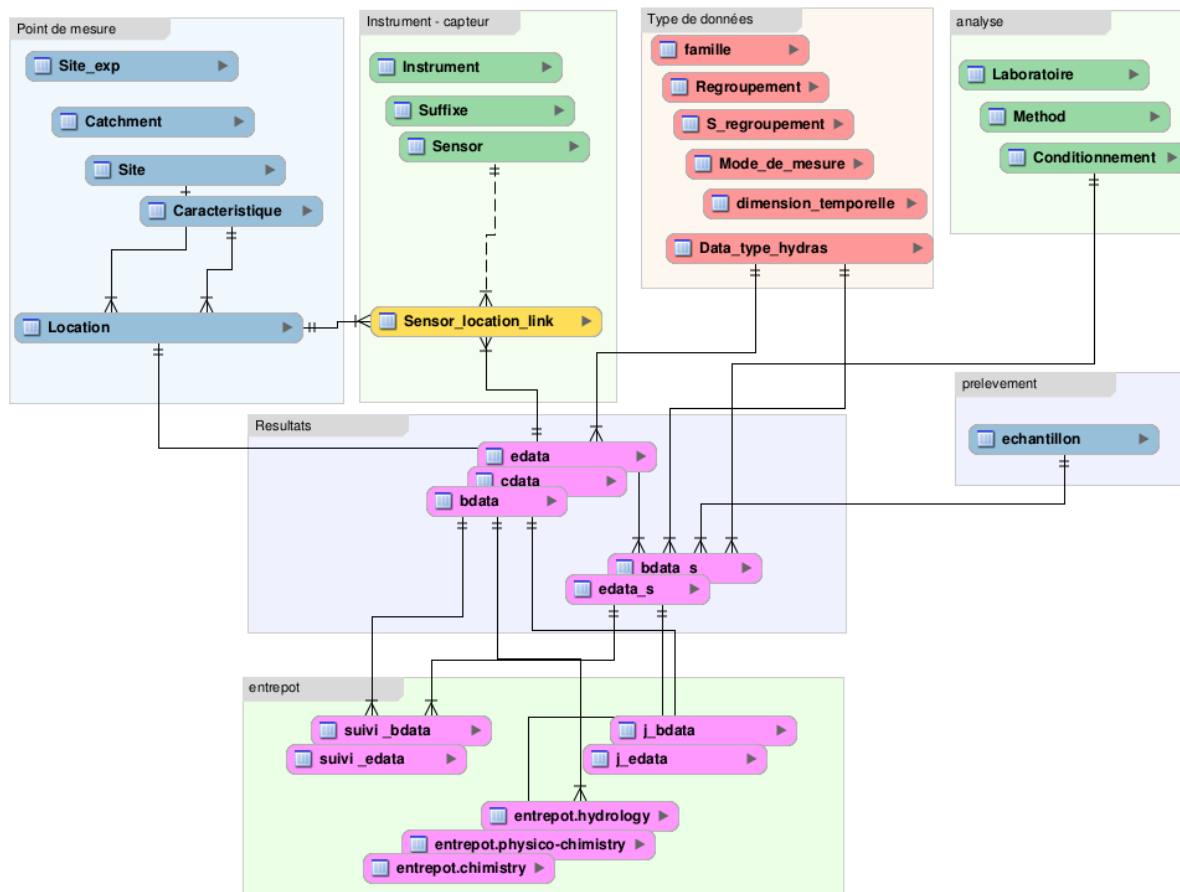


Figure 2. Schéma simplifié de la base de données BDDAgrhys.

Export automatique des données capteur du logiciel commercial Hydras 3

Le logiciel Hydras 3 permet d'exporter les données sous forme de fichiers csv. Ces fichiers sont créés dans un répertoire temporaire sur un serveur de l'unité. Chaque fichier exporté a un nom précis de la forme `IdPoint_IdDonnee_IdDomaine.csv` où figure l'identifiant du point de mesure (IdPoint, ex: 55 pour exutoire E30), du type de données (IdDonnee, ex:102 hauteur d'eau) et l'identifiant du domaine HYDRAS à l'origine des données (IdDomaine, à savoir I, II ou III). Pour les données brutes (domaine I) les exports sont programmés automatiquement tous les soirs. Il a été convenu d'exporter uniquement les données des 5 derniers mois. Pour les domaines II et III, il est nécessaire d'intégrer le temps de l'expertise par un ou plusieurs opérateurs. À l'issue de cette étape, les exports sont réalisés ponctuellement (Figure 3), et portent sur la totalité de la série de données pour intégrer d'éventuelles nouvelles interventions (correction, suppression, reconstitution) sur des données récentes ou anciennes.

Import automatique des données capteur dans BDDAgrhys, calcul des qualifications et du débit

À la suite de ces exports, le serveur Agrhys.fr lance l'outil Crontab qui est un planificateur de tâches. Le Crontab active d'abord le transfert sécurisé des fichiers csv entre deux machines distantes (SCP) : le serveur de l'unité (où se trouvent les fichiers exports csv) et le serveur virtuel (dans un fichier temporaire) où est installée la base de données BDDAgrhys. Le planificateur lance ensuite l'import des données des fichiers csv vers la base archive BDDAgrhys, dans trois tables correspondant aux trois domaines d'hydras décrits précédemment, et définies comme suit :

- la table des données issues du domaine I : `agrhy.bdata`, où chaque enregistrement est de la forme `PointDeMesure*TypeDeDonnee*Date*Valeur` ;
- la table des données issues du domaine II: `agrhy.cdata`, où chaque enregistrement est de la forme

*PointDeMesure*TypeDeDonnee*Date*Valeur*Code1 ;*

- la table des données issues du domaine III : agrhys.edata, où chaque enregistrement est de la forme *PointDeMesure*TypeDeDonnee*Date*Valeur*Code1*Code2*.

Le programme d'import décrypte le nom des fichiers et en déduit le point de mesure, le type et l'origine des données (domaine I, II, III). Si la destination des données est le domaine I, il complète la table de données brutes existantes après contrôle d'unicité et de cohérence. Si la destination des données est le domaine II ou le domaine III, le programme d'import annule et remplace la totalité des données de la table concernée, afin de prendre en compte les corrections réalisées sur des données récentes ou anciennes. Cette étape est rigoureusement effectuée sur le domaine II avant le domaine III. L'unicité et la cohérence des données sont contrôlées.

Les codes sont calculés automatiquement par comparaison des tables I et II pour le code 1, et comparaison des tables II et III pour le code 2 selon leur différence ou identité (**Tableau 2**). La combinaison des deux codes permet l'attribution d'un code de traitement d'une part et d'une fiabilité aux données de la table III. Ces codes de qualification sont fournis à l'extraction des données.

Le programme annule, insère et calcule aussi certaines variables, en particulier pour les stations cours d'eau, le débit à l'exutoire calculé à partir de la hauteur d'eau et d'une courbe de tarage établie par les opérateurs. Le débit n'est calculé que sur les valeurs traitées et fiables de hauteur d'eau.

Tableau 2. Extrait de la table de calcul automatisé des codes corrections pour les données capteur. Cette codification est affectée comme suit, 1 : pas de correction, 2 : donnée corrigée, 3 : donnée supprimée, 4 donnée reconstituée, 5 donnée non acquise. L'extrait illustre ici les valeurs de code affectées selon différents types de traitement classiquement apporté aux données. Le code 1 est obtenu par comparaison entre I et II, le code 2 par comparaison entre II et III. Le code de traitement et la fiabilité sont associés à la donnée traitée (table III) pour chaque enregistrement.

| I (brut) | | II | | III (traité) | Illustration | Code 1 | Code 2 | Code traitement | Fiabilité |
|----------|---|--------|---|--------------|--|--------|--------|-----------------|-----------|
| existe | = | existe | = | existe | Aucune modification : cas le plus courant | 1 | 1 | 1 | Oui |
| existe | ≠ | existe | = | existe | Correction du niveau du cours d'eau à l'aide d'un 2 ^e capteur | 2 | 1 | 2 | Oui |
| existe | = | existe | ≠ | existe | Purges d'un piézomètre | 1 | 2 | 2 | - |
| existe | = | existe | | vide | Niveau piézométrique inférieur à la profondeur du capteur | 1 | 3 | 3 | - |
| vide | | vide | | existe | Interpolation de données sur une période longue | 5 | 4 | 4 | - |
| vide | | existe | = | existe | Interpolation de données piézométrique sur une durée courte (<1h) | 4 | 1 | 4 | Oui |

Enfin ce serveur lance l'insertion des données dans les tables correspondantes de BDDAgrhys. À l'issue de cette étape un compte rendu l'insertion est rédigé et transféré dans une boîte mail applicative et partagée si-agrhys@inra.fr.

Import automatique des données de prélèvement dans la base BDDAgrhys

Il existe donc deux tables correspondant aux onglets Excel contenant les données brutes et traitées, définies comme suit :

- la table "agrhys.bdata_s" contient les données brutes où chaque enregistrement est de la forme *N°echantillon*PointDeMesure,TypeDeDonnee,Date,valeur*
- la table "agrhys.edata_s" contient les données traitées où chaque enregistrement est de la forme *N°echantillon*PointDeMesure,TypeDeDonnee,Date,valeur, code1, code2.*

Les résultats des données de prélèvements sont insérés manuellement (**Figure 3**), soit par l'opérateur de terrain (données physico-chimie) soit par les analystes de laboratoires qui réalisent l'analyse des échantillons, dans un fichier partagé et sécurisé sur le serveur de l'unité. Toutes les semaines, le serveur Agrhys.fr transfère de manière sécurisée (SCP) ces fichiers sur un répertoire temporaire. La procédure retenue annule et remplace la totalité de la série concernée. Le programme annule, insère et calcule aussi certaines variables, en particulier pour les stations de cours d'eau, le débit à l'exutoire calculé à partir de la hauteur d'eau relevée manuellement au moment du prélèvement. À l'issue de cette étape un compte rendu d'insertion est rédigé et transféré dans une boîte mail applicative et partagée.



Figure 3. Chronologie des chemins selon les données (4 time lines) illustrant la synchronisation des différentes étapes de l'insertion vers la base de données BDDAgrhys.

Services web pour la diffusion des données

Principes

La diffusion des données de l'observatoire repose sur la mise en œuvre opérationnelle de l'interopérabilité en mobilisant les différents protocoles de communication standards internationaux de l'Open Geospatial Consortium (OGC) qui rendent possibles le partage et la mise à disposition des données spatiales ou à entrée spatiale (chronique sur un point de mesure géoréférencé) *via* le web. On parle ici d'interopérabilité effective entre des services de diffusion et des outils ou logiciels "clients", par opposition à une simple compatibilité deux à deux comme on l'entend parfois.

Cela est rendu possible par le déploiement et l'utilisation de l'Infrastructure de Données Spatiales (IDS) de l'UMR SAS nommée GéoSAS (Bera, 2015). Cette IDS est une implémentation de l'architecture libre geOrchestra composée de modules indépendants et tous interopérables, parmi lesquels :

- un catalogue de métadonnées, basé sur l'application de catalogue [GeoNetwork](#) Open-source,
- un serveur cartographique : [GeoServer](#),
- un serveur de tuiles : [GeoWebCache](#), pour la mise en cache de tuiles de cartes (stockage transitoire de données pour accélérer leur réutilisation),
- un visualiseur et éditeur,
- un extracteur,
- un système d'authentification unique (Central Authentication System),
- et plusieurs autres modules permettant d'administrer l'architecture geOrchestra.

Nous avons repris l'approche modulaire et interopérable de l'IDS pour y intégrer des services qui répondent aux spécificités des données produites dans le cadre de l'observatoire, en particulier par leur dimension temporelle. Cela s'est traduit par le déploiement d'un serveur ISTSOS dédié à la diffusion des données d'observation et par le développement du grapheur web dynamique VIDAE proposant la visualisation et le téléchargement des chroniques de mesures. Nous avons également réalisé des développements qui reposent sur des modules existants de l'IDS comme le portail thématique de l'observatoire qui intègre un visualiseur et les fiches de métadonnées.

La finalité de ces services web que nous avons déployés, développés ou alimentés dans ce projet consiste à proposer différents services répondant aux trois principales fonctions attendues et préconisées par la directive INSPIRE pour *Trouver, Voir, Télécharger* les données de l'observatoire. Pour répondre à cette exigence, l'administration des données et des services s'inscrit dans un processus à trois temps : *la publication, la réutilisation, le monitoring/reporting* des consultations.

Publication

L'étape de publication des données de l'observatoire AgrHyS consiste à mettre à disposition sur internet *via* l'IDS GéoSAS les différents jeux de données produits sur l'observatoire (3b **Figure 1**). Parmi les données publiées, on peut distinguer deux grandes familles : les données spatiales "référentiel" et les données temporelles géoréférencées correspondant aux différentes variables mesurées. Chaque jeu de données est nécessairement accompagné de la publication d'une métadonnée.

Données spatiales "référentiel"

La publication des données spatiales "référentiel" résulte d'une utilisation classique de l'IDS et plus particulièrement de son serveur cartographique GeoServer. Ces données spatiales "référentiel" sont soit de type

“vecteur” soit de type “raster” (ou matriciel). Elles ont été produites au moyen de logiciels SIG (QGIS, ArcGIS...) et sont constituées de :

- cartes pédologiques au 1/25000,
- cartes parcellaires et d'occupation du sol,
- cartographies du réseau hydrographique,
- cartographies de la délimitation des bassins versants,
- cartographies des réseaux de fossés ou de haies,
- des modèles numériques de terrain représentant l'altitude.

Les couches disponibles sur l'IDS GéoSAS relatives à l'Observatoire AgrHyS sont récapitulées dans le **Tableau 3**. En pratique, la publication des données spatiales sur l'IDS consiste à déposer les fichiers au format SHAPE pour les données vecteur et au format GEOTIFF pour les données raster sur le serveur de l'IDS puis de créer la couche correspondant sur le serveur cartographique GeoServer. Une fois créée, la couche est immédiatement accessible sur internet *via* les différents protocoles de communication standards de l'OGC (**Figure 4**) permettant :

- la visualisation : WMS (Web Mapping Service),
- la sémiologie : SLD (Style Layer Descriptor),
- l'accès aux données vecteur : WFS (Web Feature Service),
- l'accès aux données raster : WCS (Web Coverage Service).

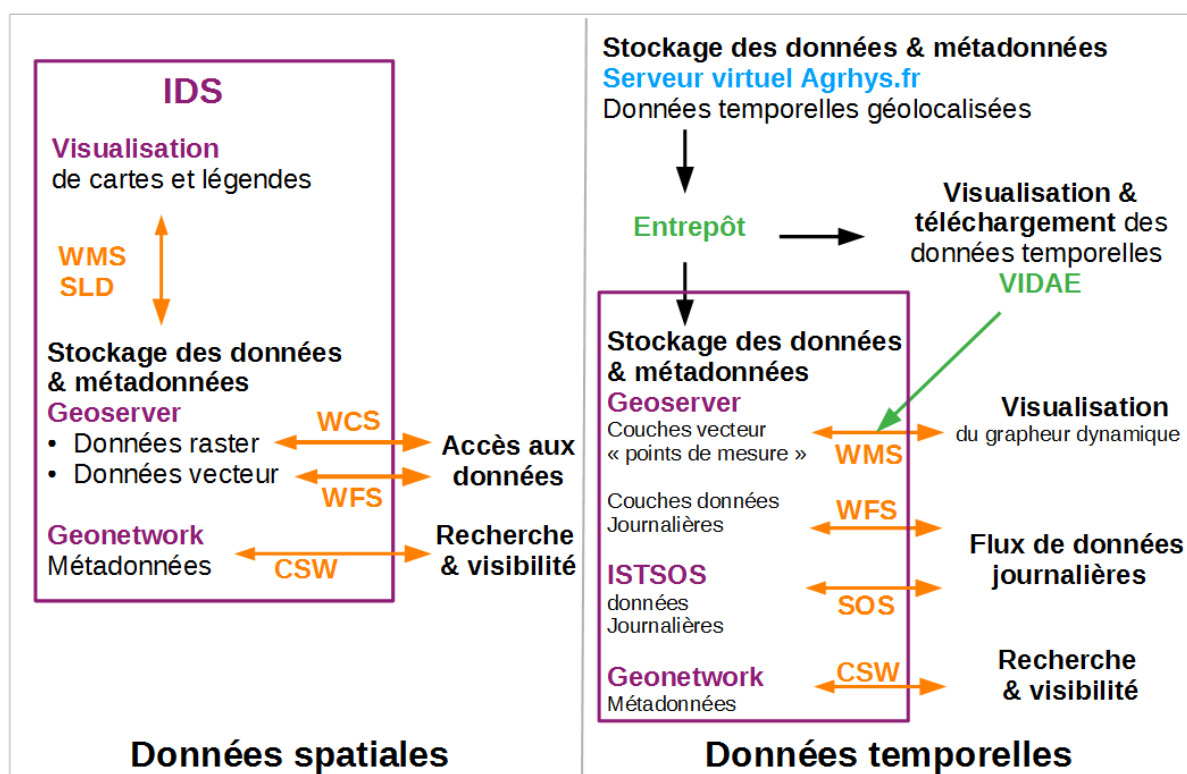


Figure 4. Synthèse des outils et protocoles de diffusion des données spatiales et temporelles. L'infrastructure de données spatiales et ses modules sont représentés en violet, les protocoles de communication standards en orange, les autres développements en vert.

Données temporelles géoréférencées

La publication des données temporelles représentant les mesures hydrologiques, physico-chimiques et chimiques du SOERE nécessite une utilisation spécifique et originale de l'IDS. La solution proposée est de

permettre deux modes de réutilisation distincts dans l'accès à ces données temporelles en publiant deux types de couches spatiales sur GeoServer. Le premier mode permet l'accès à l'application de représentation des chroniques (grapheur dynamique VIDAE) par simple interrogation *via* la fonction GetFeatureInfo du standard OGC WMS. Le second propose un accès direct et donc automatisable aux données journalières *via* des flux WFS (**Figure 4**). Le premier mode est destiné à des utilisateurs humains, le second est adapté à un usage automatique, par exemple par des modèles. Dans les deux cas et pour des raisons de sécurité et d'optimisation dans les temps d'accès aux données, une structure intermédiaire entre la BdD archive (BDDAgrhys) et les couches publiées sur le GeoServer est automatiquement créée sous la forme d'une BdD entrepôt (3a **Figure 1**). Cet entrepôt, non existant dans l'architecture geOrchestra, a été développé sous l'implémentation ISTSOS (Cannata, 2010).

Publication des graphes dynamiques *via* le mode WMS GetFeatureInfo

L'accès aux données temporelles fait appel à une application pour visualiser et télécharger ces données présentée plus loin (grapheur VIDAE). Le lien entre la représentation spatiale de la position des points de mesure et cette application est effectuée par la publication sur GeoServer d'une couche vecteur "Points de mesure" configurée spécifiquement pour établir ce lien. La liaison est alors concrétisée par la mise en place d'une interrogation standard "WMS GetFeatureInfo" sur cette couche.

Cela se fait en trois étapes, la première étant la création d'une table "Point de mesure" dans l'entrepôt du serveur de BdD de l'IDS, par export automatique à partir de la BdD archive (BDDAgrhys). Cette table au format Postgres/PostGIS, contient les informations de base de chaque point de mesure (Identifiant, type de station, coordonnées X, Y...). Cette table est actualisée à chaque modification du dispositif d'observation de l'ORE dans la BdD archive. La seconde étape est la création de la couche vecteur "Points de mesure" sur le serveur cartographique GeoServer pointant sur la table "Points de mesure" de l'entrepôt. Enfin, la dernière étape est la configuration sur le GeoServer de la réponse à l'interrogation spatiale WMS GetFeatureInfo. GetFeatureInfo est un appel standard WMS qui permet de récupérer des informations sur les couches spatiales affichées sur une carte. Par défaut cette requête renvoie la table attributaire des objets sélectionnées. La démarche consiste ici à personnaliser ce renvoi à l'aide d'un modèle développé en FreeMarker dans GeoServer (Doc GeoServer GetFeatureInfo) pour qu'il renvoie pour chaque point de mesure :

- une photo du capteur,
- un tableau synthétique généré à la volée et décrivant pour chaque type de données, le nombre de mesures ainsi que la date du premier et du dernier enregistrement,
- un accès au grapheur dynamique de ce capteur et sa représentation d'une chronique par défaut de données journalières.

Pour l'utilisateur, cela se matérialise par l'apparition d'une fenêtre lorsqu'il clique sur un point de mesure de la couche spatiale localisant ces points (**Figure 5**).

[Accéder au grapheur pour l'Exutoire E30](#) - Site de Kerbernez, Finistere



| Type de données | Date de début | Date de fin | Nombre |
|-----------------------------------|---------------|-------------|----------|
| Hydrology | 11/08/97 | 25/02/19 | 25053920 |
| Physico-chemistry | 11/08/97 | 28/02/19 | 9501088 |
| Chemistry | 11/08/97 | 04/12/18 | 7744 |

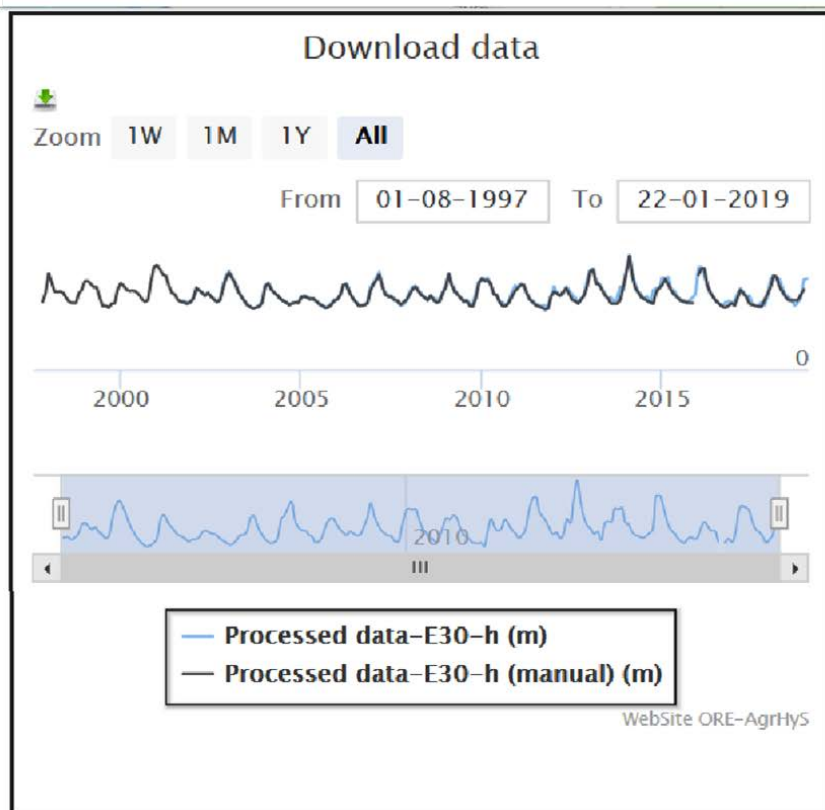


Figure 5. Personnalisation de la requête WMS GetFeatureInfo de la couche "Points de Mesure" pour renvoyer la photo du point de mesure, une information synthétique sur le nombre de mesures disponibles sur ce point, et un lien vers le grapheur dynamique avec affichage d'une chronique par défaut.

Publication de flux de données temporelles en mode flux WFS temporel

Le principe du mode "flux WFS temporel" consiste à détourner l'usage classique du standard OGC WFS habituellement utilisé pour diffuser des données spatiales pour dans notre cas, proposer un service d'accès interopérable et automatique à des chroniques de données mesurées, tout en conservant l'entrée géographique. La mise en place du mode "WFS temporel" est effectuée par la publication sur le serveur cartographique GeoServer de trois couches correspondant aux trois familles de mesures réalisées sur l'observatoire AgrHYS: hydrologiques, physico-chimiques et chimiques. En présupposant les types de réutilisation, nous avons opté pour la diffusion de données validées et moyennées au pas de temps journalier. Cette publication se fait en deux étapes, la première consiste à créer trois tables "Données hydrologiques", "Données physico-chimiques", "Données chimiques" dans l'entrepôt du serveur de BdD de l'IDS. Ces tables sont alimentées quotidiennement par export automatique à partir de la BdD archive BDDAgrhys. La deuxième étape est la création des trois couches associées "Données hydrologiques journalières", "Données physico-chimiques journalières", et

“Données chimiques journalières” sur le serveur cartographique GeoServer pointant sur la table de l'entrepôt correspondant.

Publication de flux de données temporelles en mode flux SOS

Défini par le groupe de travail Sensor Web Enablement (SWE) de l'OGC, le standard Sensor Observation Service (SOS) est un protocole de découverte, d'accès et d'échange de données temporelles adapté aux observatoires de recherche. La mise en place du mode flux SOS est effectuée par la publication sur le serveur ISTSOS de données validées et moyennées au pas de temps journalier. Le service SOS est alimenté quotidiennement par exports automatiques à partir de la BDD archive BDDArchive. À ce jour le serveur de test est opérationnel et nous envisageons une mise en production rapide.

Métadonnées

Une fiche de métadonnées est associée à chaque couche spatiale “standard” publiée sur GeoServer. La création et la saisie des fiches de métadonnées s'effectuent sur le catalogue GeoNetwork, module de l'IDS. Conformément à la directive européenne INSPIRE, les fiches de métadonnées produites respectent la norme ISO19115/19139. Le catalogue est également conforme au standard OGC CSW (Catalog Service for the Web), ce qui permet le moissonnage automatique des fiches de métadonnées par d'autres catalogues.

La métadonnée décrivant la couche “Points de mesure” est spécifique puisqu'elle fait référence à plusieurs couches spatiales et temporelles décrites précédemment. Elle pointe également vers des métadonnées décrivant les dictionnaires d'attribut des trois flux WFS vers les données hydrologiques, physico-chimiques et chimiques selon la norme ISO 19110.

Enfin, la métadonnée de service INSPIRE constitue une porte d'entrée vers l'ensemble des données et outils accessibles sur le web. Elle est conforme à la norme ISO 19119. (MIG 2013).

Par ailleurs, Il est possible d'affecter un identifiant numérique (DOI) à l'ensemble des jeux de données une fois publiés (UMR SAS, 2010).

Réutilisation (*trouver, voir, télécharger*) (3c Figure 1)

Les sites Web éditoriaux (trouver)

Plusieurs sites internet font références aux données de l'observatoire AgrHyS et en pointent l'accès afin de maximiser les chances pour un utilisateur de trouver ces données.

L'observatoire AgrHyS propose bien sûr un site institutionnel (www.inra.fr/ore_agrhys) qui est développé en eZpublish. Le site présente les activités, objectifs et l'organisation de l'ORE, ils pointent notamment vers les publications associées, les thèses, et autres productions en lien avec les recherches conduites sur l'observatoire. Dans un onglet « données », il permet l'accès aux différents outils (portail thématique, VIDAE, Catalogue...) accompagné de leur présentation et tutoriels, en français et en anglais. Un lien vers ce site existe aussi depuis le site institutionnel de l'UMR SAS (www.rennes.inra.fr/umrsas), dans la rubrique “nos dispositifs”, ainsi que depuis le site internet GéoSAS, portail de l'information géographique de l'UMR (geowww.agrocampus-ouest.fr/web).

Le catalogue de métadonnées GéoSAS (trouver) (geowww.agrocampus-ouest.fr/geonetwork/apps/georchestra)

Les métadonnées du catalogue sont moissonnées, le plus souvent quotidiennement, par différents portails de métadonnées régionaux nationaux et internationaux (Figure 6) :

- GéoBretagne : Catalogue de la région Bretagne,
- Agroenvgeo : Catalogue mis en place par le Centre automatisé de traitement de l'information GEDEOP (GEstion des Données d'Expérimentations, d'Observations et de Pratiques sur les agro-socio-écosystèmes) de l'Inra,
- OSURIS : Catalogue de l'OSUR (*Observatoire des Sciences de l'Univers de Rennes*),
- GeoCatalogue : Catalogue national des données géospatiales publiques,
- geo.data.gouv.fr : plateforme dédiée au référencement et à la valorisation des données géographiques,
- INSPIRE Geoportal : Catalogue européen,
- RBV : Catalogue du Réseau des Bassins versants.

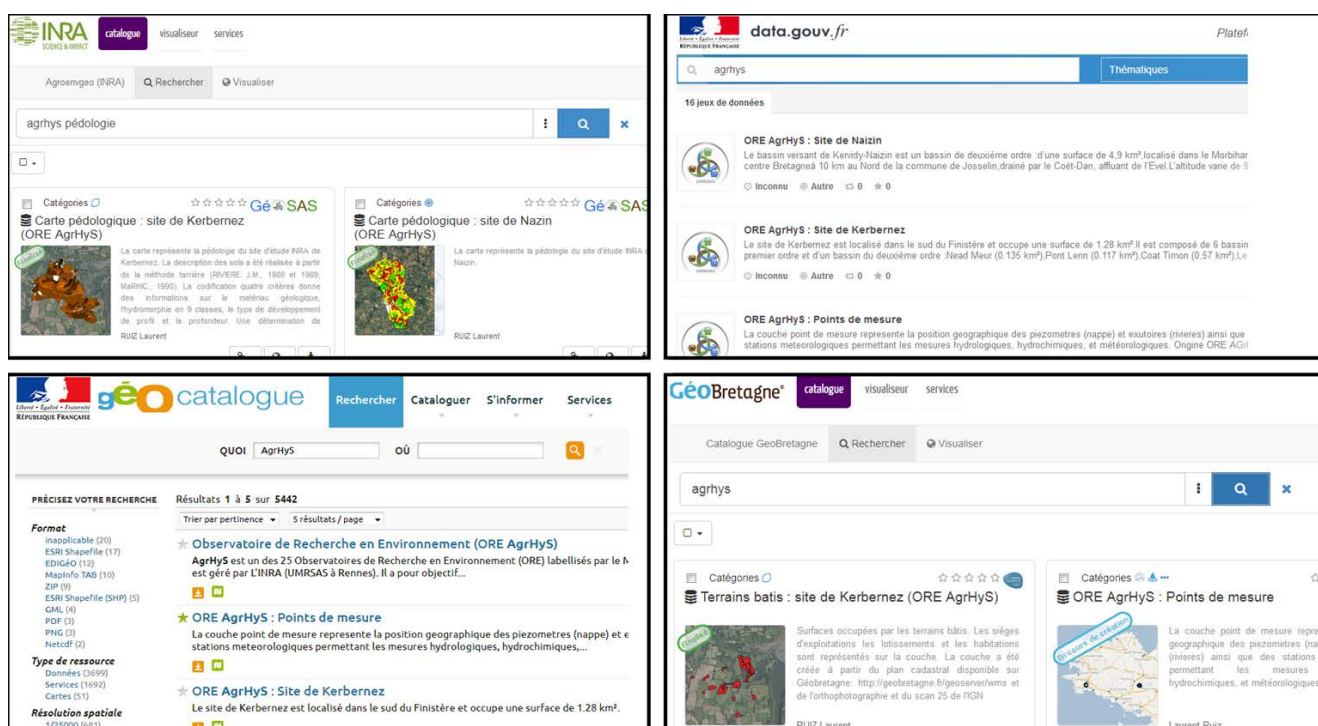


Figure 6. Illustration du moissonnage multiple des fiches de métadonnées d'AgrHyS par quelques captures d'écran sur différents portails de métadonnées.

[Le portail cartographique thématique d'AgrHyS sur GéoSAS \(trouver, voir\)](http://geowww.agrocampus-ouest.fr/hportails/?portail=vidae) (geowww.agrocampus-ouest.fr/hportails/?portail=vidae)

Ce portail propose plusieurs onglets correspondant à trois contextes permettant d'accéder à l'ensemble de l'Observatoire "AgrHyS" ou à chacun des sites de l'observatoire "Naizin" et "Kerbernez". Le portail présente également deux modes d'accès pour chaque contexte sélectionné :

- ❑ "Rapide" permet de visualiser pour chaque contexte, quelques couches sélectionnées par défaut et les points de mesure. Grâce à la publication en mode WMS temporel, un clic sur un point de mesure fait apparaître une fenêtre avec un résumé des informations disponibles sur ce point de mesure, une photo si elle existe, et un graphe d'un type de donnée prédéfini. Les liens actifs permettent d'accéder au grapheur dynamique VIDAE.
- ❑ "Expert" permet de visualiser sur chaque contexte mais aussi de rajouter des couches à partir des différents catalogues disponibles (GéoSAS, GéoBretagne...).

[Le grapheur dynamique VIDAE \(voir, télécharger\)](http://agrhyss.fr/BVE/vidae) (agrhyss.fr/BVE/vidae)

Le grapheur dynamique VIDAE (**Figure 7**) permet de visualiser les données temporelles, de les superposer, et de les télécharger au format csv. Il s'agit d'un outil Web interactif développé en JavaScript avec la bibliothèque ExtJS. Le grapheur VIDAE est composé de trois parties. À gauche, se trouve l'arbre de sélection des points de mesure. À droite, s'effectue le choix de l'origine de la donnée (brutes ou traitées), et en dessous l'arbre de sélection des types de données organisés par thème. Tous ces outils de sélection ont été développés en PHP. Au centre, se trouve l'outil d'affichage du graphe de données journalières correspondant aux sélections. Les graphes, dynamiques, sont construits avec la bibliothèque JavaScript Highcharts. Pour certains points de mesure, les exutoires notamment, le graphe peut afficher les données brutes très récentes, collectées la veille. En haut au centre de ce graphe, se trouve un bouton de téléchargement qui permet, *via* un développement en PHP, d'obtenir des extractions sous forme de fichiers csv, le tout disponible dans une archive compressée. Les types d'extraction possibles par point de mesure sont, pour une période sélectionnée,

pour les données capteur :

- toutes les données instantanées brutes
- toutes les données instantanées traitées (chaque enregistrement est enrichi d'un code traitement et d'un code fiabilité, **Tableau 2**)
- les données instantanées traitées et fiables uniquement
- les données journalières (moyenne ou cumul des données fiables, selon le type de données)

et pour les données de prélèvements :

- les données instantanées brutes
- les données instantanées traitées (chaque enregistrement est enrichi d'un code traitement et d'un code fiabilité, **Tableau 2**).

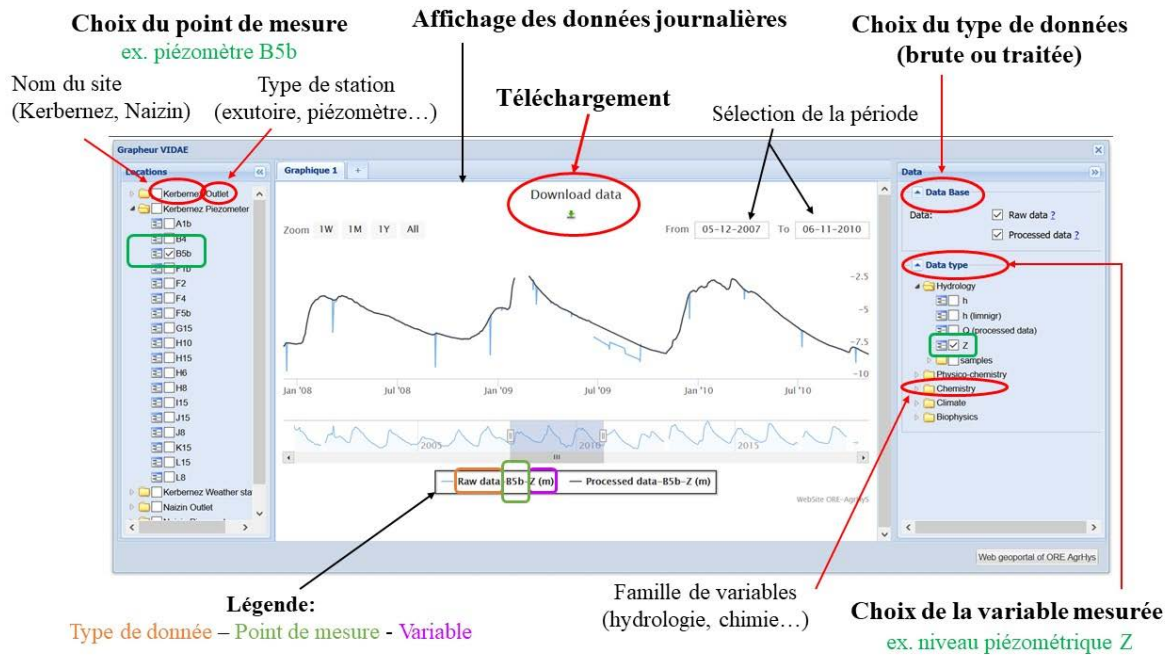


Figure 7. Présentation du grapheur dynamique VIDAE : affichage d'un extrait d'une série de hauteur piézométrique de nappe au point de mesure B5b. La double sélection ici des séries brutes et traitées (courbes bleu et noir respectivement) met en évidence l'expertise apportée par le traitement.

Interrogation du FLUX WMS à partir d'un logiciel SIG Desktop

Voici un exemple de réutilisation relativement inattendue mais qui démontre bien l'interopérabilité effective de l'approche. Il consiste à charger la couche "Points de mesure" de l'ORE AgrHyS dans le logiciel SIG desktop QGIS via un flux WMS, puis d'interroger un des points de mesure. Au lieu d'afficher de façon classique la table attributaire, QGIS affiche la photo, le tableau synthétique et le grapheur dynamique comme décrit dans la section précédente (Figure 8). On peut alors accéder au grapheur dynamique VIDAE via les liens actifs à partir du logiciel SIG également.

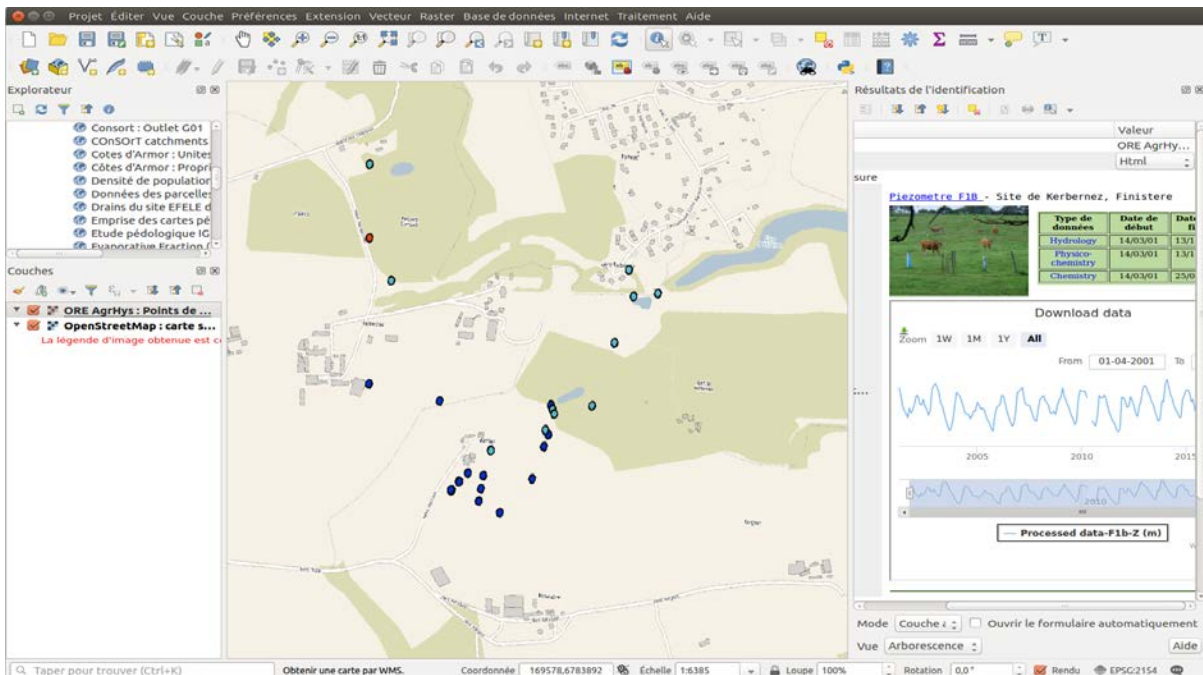


Figure 8. Capture d'écran illustrant un exemple de réutilisation de la couche "Points de Mesure" dans le logiciel Desktop QGIS 3.4.

Diffusion des données journalières via le Flux WFS temporel

La publication des chroniques des données journalières via des flux WFS a été pensée pour une réutilisation directe à partir de modèles ou par des utilisateurs experts. On se rapproche alors des services de type REST tout en utilisant un protocole d'échange standard interopérable, le Web Feature Service (WFS). L'interrogation du service WFS se fait par des requêtes HTTP GET ou POST. Il est possible de paramétrer ces requêtes de telle sorte que ne soient sélectionnées que certaines mesures. Pour cela on peut utiliser et combiner différents filtres pour sélectionner une période, un point de mesure, un type de mesure, et une emprise spatiale. À titre de démonstration, nous proposons deux exemples de requête: l'une 1) récupère toutes les données de toutes les stations hydrologiques à partir du 1 juillet 2018 au format CSV, et l'autre 2) récupère uniquement les données débit du point de mesure exutoire E3 depuis le début de l'acquisition au format CSV.

- 1) http://geowww.agrocampus-ouest.fr/geoserver/ore/wfs?service=wfs&version=2.0.0&request=getFeature&typeName=ore:agrhys_donnees_hydro_jour&cql_filter=jour>='2018-07-01'&outputFormat=CSV
- 2) [http://geowww.agrocampus-ouest.fr/geoserver/ore/wfs?service=wfs&version=2.0.0&request=getFeature&typeName=ore:agrhys_donnees_hydro_jour&cql_filter=name='E3'&PropertyName=\(name.jour,q\)&outputFormat=CSV](http://geowww.agrocampus-ouest.fr/geoserver/ore/wfs?service=wfs&version=2.0.0&request=getFeature&typeName=ore:agrhys_donnees_hydro_jour&cql_filter=name='E3'&PropertyName=(name.jour,q)&outputFormat=CSV)

Monitoring et reporting

Imposé par la directive INSPIRE, le monitoring consiste à mettre en place un suivi qualitatif des services de publication en terme de conformité des données et des métadonnées, et de continuité de service avec un seuil de 99% du temps. Le **Tableau 3** fournit un exemple des informations recueillies dans le cadre de ce suivi. Destiné aux gestionnaires ou responsables des observatoires, le reporting consiste à fournir des éléments chiffrés sur l'utilisation effective des services de diffusions des données et des métadonnées (3d **Figure 1**). Pour cela, nous utilisons le logiciel libre Graylog (**Tableau 4**).

Tableau 3. Informations recueillies dans le cadre du monitoring, interrogation février 2019.

| Service health | Last 7 days | Year 2019 |
|------------------------------------|-------------|-----------|
| availability | 100% | 100% |
| total downtime | 0.1 hours | 0.1 hours |
| preannounced service interruptions | 0 hour | 0 hour |
| unexpected downtime | 0.1 hours | 0.1 hours |

Tableau 4 page suivante

Tableau 4. Informations recueillies dans le cadre du reporting via le logiciel Graylog, interrogation février 2019

| | |
|---------------------------------|--------|
| Visualisations en WMS en 1 mois | 33.876 |
| IP différents* WMS en 1 mois | 101 |
| Interrogations en WMS en 1 mois | 240 |
| Téléchargements WFS en 1 mois | 305 |

* La liste des IP et leurs géolocalisations sont également fournies.

Généricité et transposition à d'autres observatoires

Les piliers au cœur de la démarche développée ici (section Introduction) sont les clés assurant la possibilité de transposer l'approche à d'autres observatoires. Cette transposition a été réalisée en particulier pour d'autres observatoires portés par l'UMR SAS ou auxquels elle contribue. Ainsi, la chaîne de traitement-diffusion présentée ici a également été transposée aux autres bassins versants sur lesquels l'UMR SAS conduit des observations hydro-chimiques sur le long terme à savoir : l'ORE Petits Fleuves Côtiers (bassins de Moulinet et de l'Oir), et le bassin de la Sélune, ainsi que la Zone Atelier Armorique (ZAA, site atelier de Pleine Fougères, Thomas et al., 2019).

Un autre exemple notable est celui du site EFELE (Effluents d'Elevage et Environnement) du Système d'Observation et d'Expérimentation pour la Recherche en Environnement PRO, un réseau de sites expérimentaux en plein champs de longue durée ayant pour objectif d'observer les effets de l'apport de Produits Résiduaires Organiques (PRO) sur les différents compartiments d'un système de grandes cultures. Il s'agit donc non plus de bassins versants mais de parcelles expérimentales instrumentées pour le suivi des trois matrices Sol, Air et Eau. L'observation sur ce site est donc également basée sur des capteurs, des prélèvements (en particulier de sols mais aussi de solutions du sol), et des mesures *in situ* (de gaz et composés organiques volatiles). La chaîne de traitement développée pour l'ORE AgrHyS a donc été transposée au réseau de capteurs d'humidité et température du sol implanté sur ce site, en particulier l'automatisation de l'insertion des données brutes d'une part, et l'intégration des outils de visualisation (VIDAE) d'autre part. La transposition de ces développements est permise aujourd'hui par le partage sur une forge publique des scripts développés pour l'ORE AgrHyS comprenant le portail cartographique thématique, le grapheur VIDAE, les scripts d'insertion, de qualification et d'exports des données. La publication des données spatiales et des métadonnées peut se faire en s'appuyant sur l'IDS de son unité (ici GéoSAS), ou de son OSU (Observatoire des Sciences de l'Univers) de rattachement (ici OSURIS pour la ZAA), ou de celle du Centre Automatisé de Traitement de l'Information Inra (CATI) GEDEOP pour tout autre SOERE.

En parallèle des mesures pérennisées sur le long-terme sur les sites-observatoires, des projets de recherche sont également conduits sur ces sites. Dans le cadre de ces projets d'une durée de 2 à 5 ans, des mesures complémentaires peuvent être effectuées pour répondre aux objectifs spécifiques du projet. Tous les projets de recherche s'accompagnent aujourd'hui d'une stratégie de gestion des données produites dans leur cadre (Data Management Plan). La chaîne de traitement-diffusion élaborée est aussi transposable à ces données produites en dehors de l'activité d'observation au sens strict. Ainsi, le système d'information a été remobilisé pour diffuser les données produites dans le cadre de l'ANR MOSAIC (2013-2017) qui s'appuyait sur l'un des sites de l'ORE AgrHyS. Une entrée spécifique au projet a été créée dans le portail cartographique de l'ORE afin d'y accéder.

L'évolution des observatoires vers la très haute résolution temporelle va générer une augmentation importante du flux de données du capteur à la diffusion. Pour la partie diffusion, notre démarche est compatible avec les

nouvelles contraintes liées au "Big Data". D'une part, le module d'affichage de graphes dynamiques Highcharts utilisé par VIDAE permet une montée en charge vers la très haute résolution temporelle potentiellement illimitée, grâce à son module Highstock. D'autre part l'évolution du Standard OGC SOS V2.0, implémenté dans la dernière version du logiciel ISTSOS V3, augmente significativement les performances afin de répondre à des scénarios de réutilisation massifs en terme de volume de donnée et de simultanéité des requêtes (Massimiliano, 2018).

Conclusions et perspectives

La réalisation de cette chaîne automatique et sécurisée retraçant le chemin de la donnée en passant par les étapes de l'acquisition, de l'expertise, de la sauvegarde, de la publication est transposable à toute autre collection de données, notamment aux données d'autres observatoires. Cette mise en œuvre a été possible grâce à la collaboration étroite entre les acteurs de terrain, les scientifiques et les informaticiens.

La démarche initiée dans le développement de l'outil VIDAE en 2009 consistant à appliquer les principes d'INSPIRE (2007) aux données spatiales et temporelles de l'ORE constituait donc un pari. Ce dernier a été remporté dans la mesure où l'infrastructure de données développée propose l'accès à des données qualifiées, actualisées (pseudo-temps réel) et répondant aux dernières exigences de l'Open Data (Wilkinson, 2016) puisque :

- Trouvables (Findable) sur différents catalogues internationaux, nationaux, régionaux, thématiques et institutionnels,
- Accessibles par tous, conformément aux licences associées, via à un navigateur Web et téléchargeables directement dans les interfaces développées,
- Interopérables grâce aux standards de diffusions utilisés, et comme en atteste leur moissonnage par d'autres portails nationaux (ex. : portail RBV) ou régionaux (ex. : GéoBretagne),
- Réutilisables car liées à des métadonnées permettant leur description, mais aussi d'un point de vue technique puisqu'elles sont également pré-visualisables et téléchargeables dans les Interfaces développées ou directement accessibles sous forme de flux de données facilitant ainsi leur réutilisation par des modèles hydrologiques, par exemple.

Comme perspective, nous avons la volonté de suivre l'évolution des bonnes pratiques et des standards de diffusion des données temporelles des observatoires. Pour cela, nous allons tout d'abord mettre en production un serveur SOS v2.0, en test à ce jour. Nous allons aussi effectuer une veille autour du thème émergent de l'Internet des objets (en anglais *Internet of Things*, ou IoT) qui a pour ambition de connecter directement les capteurs sur internet. Le nouveau standard OGC SensorThings semble adapté à la mise en œuvre de ce nouveau concept.

En aval de la démarche présentée, une perspective prometteuse consiste à approfondir la réutilisation des données, ainsi produites et diffusées, en entrée des modèles. Cette réutilisation peut prendre différentes formes en fonction de leur niveau d'automatisation et d'interopérabilité. La forme la plus simple consiste à récupérer les mesures et/ou des analyses de l'observatoire sous la forme de séries temporelles *via* un téléchargement à partir de VIDAE. Les données sont ensuite importées dans le modèle en amont de son exécution. Une forme beaucoup plus aboutie consiste à développer une approche de modélisation interopérable où le modèle, ses entrées et ses sorties, sont intégrés à l'Infrastructure de Données Spatiales et sont conformes aux standards OGC. L'import des données spatiales et temporelles se fait *via* des flux OGC WFS ou SOS, le modèle est conforme au standard WPS (Web Processing Service) (Squidant et al., 2015) et les sorties sont publiées sur GeoServer ou ISTSOS et sont de fait, elles aussi, interopérables.

Remerciements

Nous remercions Yannick Fauvel, Béatrice Trinkler, Laurence Carteaux du Pôle analytique de l'UMR SAS, Patrice Petitjean et Patricia Madec du Laboratoire d'analyse de Géosciences-Rennes de l'Université de Rennes I.

Références bibliographiques

- Bera R, Squidant H, Le Henaff G, Pichelin P, Ruiz L, et al. (2015) GeoSAS: A modular and interoperable Open Source Spatial Data Infrastructure for research. *IAHS-AISH publication = International Association of Hydrological Sciences-Association Internationale des Sciences Hydrologiques publication*, International Association of Hydrological Sciences, 368, pp.9-14. [〈10.5194/piahs-368-9-2015〉](#) . [〈hal-01222416〉](#)
- Cannata M, Antonovic M (2010) ISTSOS: Investigation of The Sensor Observation Service. WebMGS 2010: 1st International Workshop on Pervasive Web Mapping, Geoprocessing and Services. 26.-27. August 2010. Como, Italy.
- Doc GeoServer GetFeatureInfo. Personnalisation de la requête WMS GetFeatureInfo de GeoServer. <https://docs.geoserver.org/latest/en/user/tutorials/GetFeatureInfo/index.html>
- Fovet O et al. (2018) AgrHyS: An Observatory of Response Times in Agro-Hydro Systems. *Vadose Zone J.* 17(1). DOI:10.2136/vzj2018.04.0066
- Gaillardet J et al. (2018) OZCAR: The French Network of Critical Zone Observatories. *Vadose Zone J.* 17(1). DOI:10.2136/vzj2018.04.0067
- Gascuel-Oudou C, Fovet O, Gruau G, Ruiz L, Merot P (2018) Evolution of scientific questions over 50 years in the Kervidy-Naizin catchment: from catchment hydrology to integrated studies of biogeochemical cycles and agroecosystems in a rural landscape. DOI:10.18172/cig.3383
- Massimiliano C, Antonovic M, Strigaro D, Cardoso M (2018) istSOS Version 3, *Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings*: Vol. 18 , Article 5. DOI: 10.7275/zdfm-4d10
- MIG 2013. Mission de l'Information Géographique du MEDDE (2013) « Trouver, voir et télécharger » Petite incursion du côté des métadonnées de service lien: www.geoinformations.developpement-durable.gouv.fr/fichier/pdf/trouver-voir-telecharger_cle5a19eb.pdf?arg=177829947&cle=85f531d409c7f6dc1367b3c0365e33e9075464c7&file=pdf%2Ftrouver-voir-telecharger_cle5a19eb.pdf
- Molénat J et al. (2018) OMERE: A Long-Term Observatory of Soil and Water Resources, in Interaction with Agricultural and Land Management in Mediterranean Hilly Catchments. *Vadose Zone J.* 17(1). DOI:10.2136/vzj2018.04.0086
- SAS, U. M. R., & UMR SAS. (2010). Observatoire de Recherche en Environnement sur les Agro-Hydrosystèmes (ORE AgrHyS). INRA. DOI : 10.15454/1.5499682911557678E12
- Squidant H, Bera R, Arousseau P, Cudennec C (2015) Online watershed boundary delineation: sharing models through Spatial Data Infrastructures. *IAHS-AISH publication = International Association of Hydrological Sciences-Association Internationale des Sciences Hydrologiques publication*, International Association of Hydrological Sciences, 368, pp.144-149. [〈10.5194/piahs-368-144-2015〉](#) . [〈hal-01215784〉](#)

Thomas Z et al. (2019) Long-term ecological observatories needed to understand ecohydrological systems in the Anthropocene: a catchment-scale case study in Brittany, France. *Regional Environmental Change*, 19(2): 363-377. DOI:10.1007/s10113-018-1444-1

Wilkinson, Mark D et al. (2013) The FAIR Guiding Principles for scientific data management and stewardship. *Nature*.

Cet article est publié sous la licence Creative Commons (CC BY-SA).



<https://creativecommons.org/licenses/by-sa/4.0/>

Pour la citation et la reproduction de cet article, mentionner obligatoirement le titre de l'article, le nom de tous les auteurs, la mention de sa publication dans la revue « Le Cahier des Techniques de l'INRA », la date de sa publication et son URL).