

Mise en place d’un outil web-cartographique d’analyse et de diffusion d’observations environnementales

Vincent Thieu^{1*}, Marie Silvestre², Corentin Souton¹

¹ UPMC, UMR 7619 METIS, 4 place Jussieu, 75005 PARIS, France.

² FIRE, CNRS, UMR 7619 METIS, 4 place Jussieu, 75005 PARIS, France.

* personne à contacter : vincent.thieu@upmc.fr

Résumé

La recherche est nourrie par un ensemble de (bases de) données de plus en plus volumineuses qu’elle nourrit également en retour. Ce processus d’enrichissement est souvent limité dans le temps (durée d’une thèse, d’un programme de recherche, d’un CDD, etc.), et l’archivage de ces données extrêmement diversifiées est rarement pérenne.

Pourtant, l’accessibilité aux données environnementales géographiques est devenue une étape cruciale dans les travaux scientifiques. L’arrivée de l’Open-data dans un contexte déjà assez contraint (PSI sur réutilisation des données publiques, convention d’Aarhus sur l’accès à l’information environnementale, directive INSPIRE sur la diffusion en ligne des données, etc.) doit nous inciter à réfléchir à une politique de gestion des données plus efficiente avec pour double objectif : (1) de répondre sur le long terme à des obligations juridiques d’accès (publication, valorisation) aux données ; (2) de permettre, en interne, l’interopérabilité des données entre chercheurs en proposant des critères de standardisation, de mutualisation et de subsidiarité.

Initialement conçu et structuré pour valider le déploiement du modèle pyNuts-Riverstrahler à échelle de la façade atlantique Nord-Est (1 million de km² de bassin-versant et 350 000 km de réseau hydrographique depuis le Rhin jusqu’au Guadalquivir), le système d’information DoNuts (Database of Observed NUTrientS) s’appuie sur des choix technologiques (PostgreSQL + PostGIS) et des standards offrant une genericité suffisante pour proposer aux équipes du PIREN Seine et aux partenaires : (1) un accès standard et structuré à une sélection des données d’intérêts, homogénéisées et documentées ; (2) une solution d’archivage pérenne pour les données produites (càd acquises, analysées et encodées) (3) des possibilités de partage et/ou de (re-)distribution de certaines données (selon les droits associés et en contrôlant les accès).

Un outil web-cartographique (CocoNuts) a ainsi été développé pour offrir à l’ensemble des collaborateurs et partenaires du PIREN un accès simple (ne nécessitant pas de connaissances expertes) aux données de la base DoNuts. Basé sur des technologies libres, cet outil propose un ensemble de fonctionnalités facilitant la sélection, l’exploration, l’analyse et le téléchargement de données d’intérêt.

1 Introduction

1.1 Contraintes et enjeux liés aux données géographiques environnementales

La complexité du cadre juridique qui s’étioffe depuis bientôt 20 ans démontre l’enjeu grandissant autour des données qui revêtent une dimension environnementale, de surcroît si ces dernières disposent d’une composante spatiale (on parlera alors de donnée géographique environnementale).

En 1998, la convention internationale d’Aarhus établissait les bases légales des accès à l’information environnementale, afin de permettre à tous citoyens d’agir en « acteur informé ». Un peu plus tard la directive dite PSI (Public Sector Information, 2003/98/CE du 17 novembre 2003) précisait les conditions de réutilisation libre des données du secteur publique, permettant en théorie à toute personne de demander la totalité des données produites par une autorité publique. La transposition en droit français (2005), précisait alors l’obligation de créer des catalogues informant sur l’existence des données disponibles. En 2007 la directive INSPIRE (2007/2/CE du 14 mars 2007), sur le partage de l’information géographique entre autorités publiques, établissait un cadre technique pour favoriser le partage de l’information géographique avec des structures de données normalisées. Plus récemment la Charte du G8 sur l’open data (2013) impose une diffusion libre et gratuite de toutes les données (pas uniquement géographiques) de l’État et des établissements publics (Compte-rendu du séminaire : Les données accessibles dans le domaine environnemental – All Envi, 2014)

Dans ce maillage juridique complexe, les données issues de la recherche disposent d’un statut particulier. En effet, les enseignants-chercheurs possèdent un droit de propriété intellectuelle sur leur données (dites données élaborées), et restent ainsi propriétaire de ce qu’ils produisent, ce qui leurs permet de contrôler la diffusion de ces dernières. Néanmoins, la production de métadonnées et l’effort de catalogage (en ligne) informant sur l’existence de ces données restent en théorie exigés (Leobet, 2014).

La dynamique rapide de cette réglementation et les évolutions pressenties (notamment l’extension des données publiques *“aux documents détenus par des établissements d’enseignement et de recherche, y compris des organisations créées pour le transfert des résultats de la recherche, des écoles et des universités”*, PSI 2013), nous obligent à rester vigilants sur nos futures obligations en matière de gestion de données géographiques environnementales.

1.2 Cycle de vie de la donnée et bonnes pratiques pour le PIREN Seine

L’idée maitresse sous-jacente à l’ensemble de ces directives est le décroisement de l’information environnementale au bénéfice de collaborations plus efficaces autour de données partagées.

Cette idée d’ouverture et de partage de l’information n’est pas nouvelle au sein du PIREN. Néanmoins, l’analyse de ce cadre réglementaire permet d’identifier des moyens techniques pour aider les équipes du PIREN à mieux gérer sur le long terme la mutualisation et la pérennisation de leurs données.

Création, traitement, analyse, sauvegarde, diffusion et réutilisation constituent en théorie, et dans cet ordre, le cycle de vie d’une donnée scientifique. Dans la pratique, si nos recherches mobilisent et produisent des volumes importants de données, le fonctionnement contraint des projets de recherche (à durée déterminée) permet rarement d’achever ce processus cyclique, qui s’interrompt le plus souvent après l’étape d’« analyse » pour nourrir des besoins pressants de valorisation scientifique. La pérennisation des données utilisées, traitées et enrichies dans le cadre d’une thèse, d’un programme de recherche ou d’un CDD constitue donc un enjeu réel et doit inclure, au-delà du simple archivage, un effort de documentation (métadonnées) et une réflexion sur les formats de stockage et les droits associés. C’est la condition sine qua non pour éviter la perte d’information inéluctable au sein d’un groupement de recherche comme le PIREN qui utilise et produit des données depuis plus de 25 ans. C’est également un effort nécessaire pour permettre la réutilisation (à moindre coût) et l’interopérabilité des données entre chercheurs.

1.3 Le projet EMoSEM et la plateforme de modélisation pyNuts-Riverstrahler

Dans le cadre du projet européen EMoSEM (SEAS-ERA), les équipes de modélisation biogéochimique de METIS ont développé une infrastructure dédiée, pyNuts, qui a permis de faire évoluer le modèle biogéochimique Riverstrahler (Billen et al. 1994, Garnier et al. 1995) en le portant à une échelle continentale incluant l’ensemble des fleuves de la façade atlantique nord-est (soit plus de 350 000 km de réseau hydrographique du Rhin au Guadalquivir, (rapport EMoSEM : Desmit et al. 2015).

Une partie des développements réalisés durant ce projet ont concerné l’archivage des observations environnementales avec le recensement et la collecte de données sur la qualité et l’hydrologie des cours d’eaux auprès des réseaux de mesures dans une dizaine de pays (Tab. 1). Ces données étaient requises d’une part pour estimer certains forçages du modèle biogéochimique pyNuts-Riverstrahler (par exemple : l’hydrologie, la température, etc.) et d’autre part pour pouvoir valider les simulations produites par ce dernier pour différentes variables d’intérêts (nitrate, phosphate, silice, oxygène, etc.) (Fig.1).

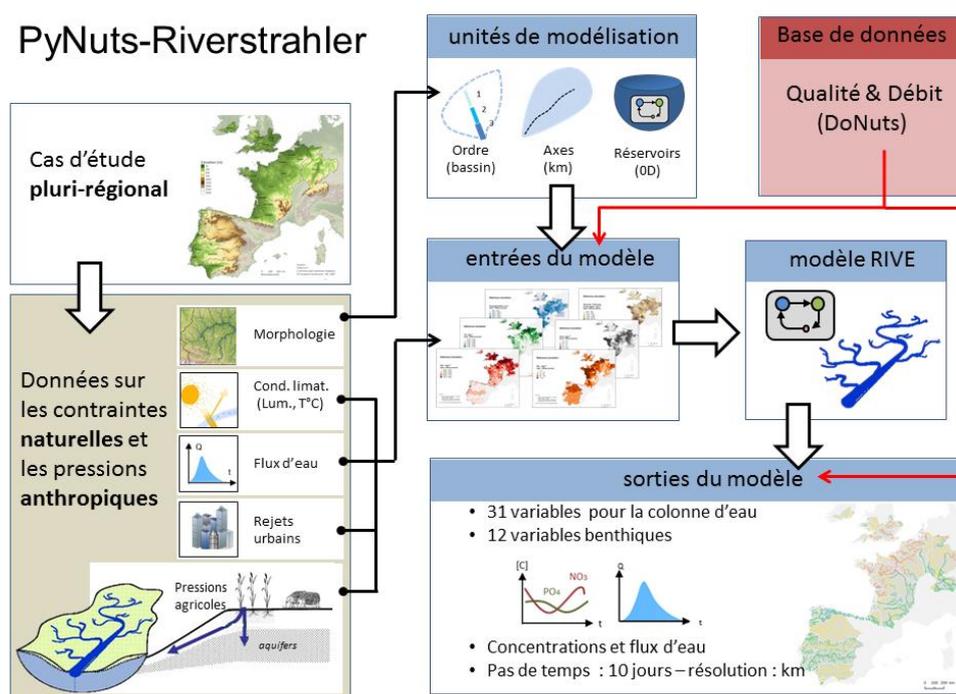


Figure 1 : Schéma conceptuel de l’approche de modélisation biogéochimique pyNuts-Riverstrahler, incluant (en rouge) la base de données DoNuts (Database of Observed NUTrientS)

La structuration de ces observations environnementales (volumineuses et hétérogènes) s’est appuyée sur (i) des technologies de gestion de base de données permettant le stockage et la manipulation des différentes informations de façon efficace et (ii) des développements méthodologiques pour assurer l’interopérabilité de ces données et permettre une intégration rigoureuse des mesures de terrain (concentrations et débit).

Au-delà du volume de donnée agrégé dans le cadre du projet EMoSEM, il semble intéressant de pouvoir restituer ici, une description de cette base de données DoNuts (Database of Observed NUTrientS) et des développements qui l’accompagnent (procédure d’intégration, de mise à jour et projet interfaçage) en vue d’une utilisation et d’une appropriation par les équipes du PIREN. En effet, la généricité de cette base permet de proposer une solution de partage et d’archivage des données ponctuelles produites par les équipes du PIREN.

2 Présentation de la base DoNuts

2.1 Structuration des informations dans DoNuts

DoNuts rassemble des mesures/observations sur la qualité et l’hydrologie des cours d’eau issues de multiples sources, qui sont pour le moment principalement des réseaux de surveillance des milieux mais qui pourraient également inclure des données issues de la recherche (et notamment des équipes PIREN). Les informations qui composent chaque observation environnementale (les formats, les unités de mesures, etc.) sont décrites et stockées de façon homogène.

Afin d’assurer une organisation optimale (non redondante) des informations stockées, la base DoNuts est structurée à l’aide du système de gestion de base de donnée relationnelle PostgreSQL, qui, grâce à l’extension spatiale PostGIS, permet d’associer aux fonctionnalités d’analyses de données numériques classiques, des fonctionnalités SIG (opérateurs topologiques, jointure-spatiale, géotraitements, etc.). Cette base de données permet l’accès multi-utilisateurs et peut être préparée à une diffusion de son contenu via le web (cf. section 3).

Les choix de conceptualisation et les regroupements d’informations sous forme de tables, reliées par des relations d’appartenance simple ou multiple, doivent permettre un accès standard et exhaustif à toute l’information qui accompagne systématiquement une observation environnementale (Fig. 2).

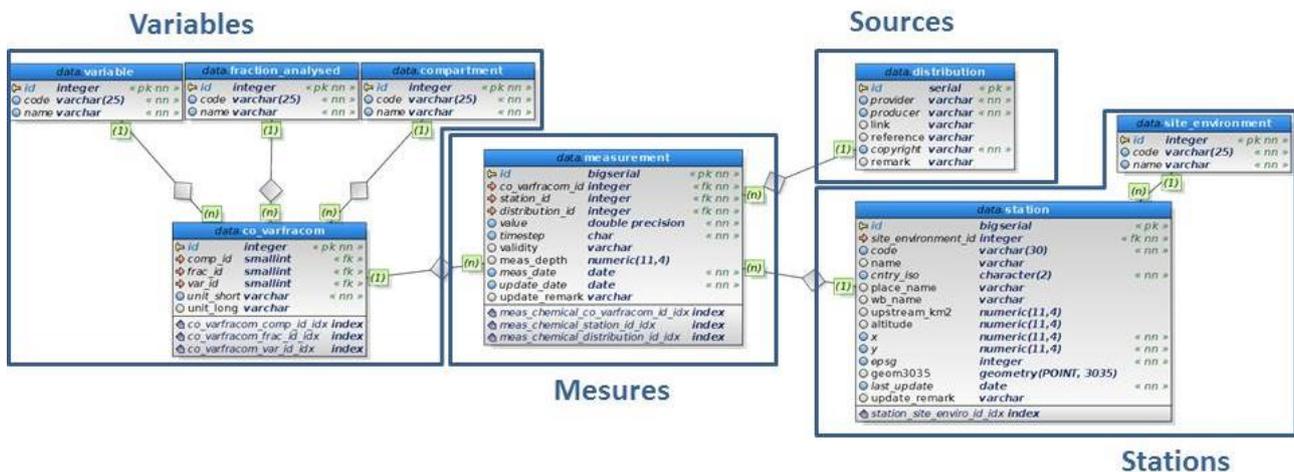


Figure 2 : Schéma relationnel de base de données DoNuts

Dans DoNuts, l’information relative à toute *observation ponctuelle* se répartie selon 4 grands thèmes décrivant à minima un lieu (Stations), une date et un pas de temps (Mesures), un acteur (Sources) et un élément observé (Variables). Ces 4 thèmes sont repris ci-après.

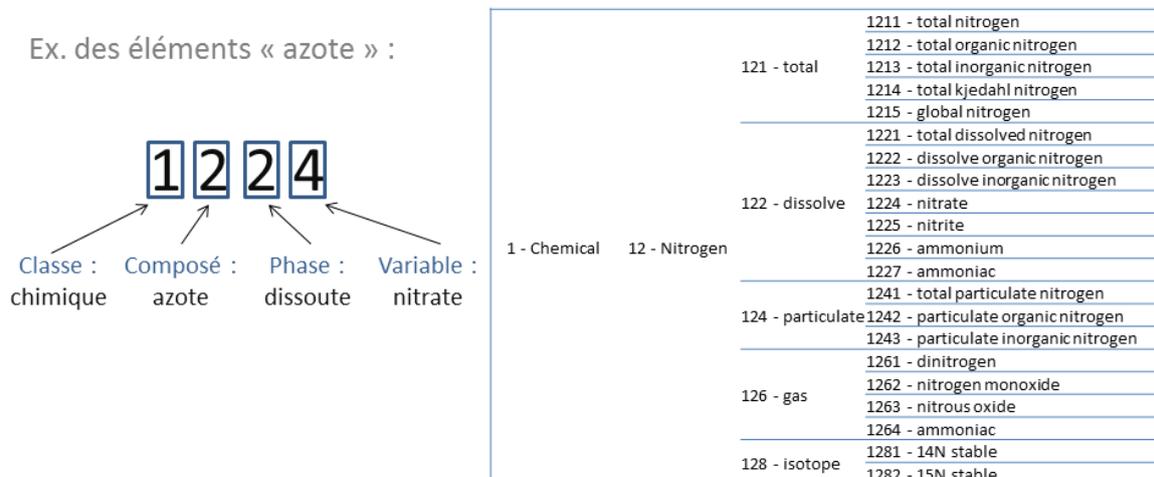
2.1.1 Les variables

Il s’agit en premier lieu de décrire la *variable observée* : cette information combine l’identification d’un compartiment prélevé (par exemple : colonne d’eau, sédiment benthique, sol, atmosphère, etc.) mais aussi la caractérisation de la fraction analysée (par exemple : eau filtrée, mes < 2mm, gaz brute etc.) et la dénomination de l’élément d’intérêt (par exemple : nitrate, phosphate, etc.).

Dans un compartiment environnemental, chaque élément (espèce chimique, physique ou biologique) peut être analysé sur une fraction (substrat) particulière, constituant ainsi une combinaison unique, à laquelle est associée une unité de référence spécifique.

Chaque compartiment, chaque fraction d’analyse et chaque élément d’intérêt dispose d’un identifiant numérique unique permettant d’assurer sans erreur possible les associations nécessaires pour caractériser une *variable observée*. Une classification de ces identifiants numériques a été établie afin de permettre des accès rapide à certains sous-ensembles.

La figure 3 illustre ce principe de classification pour les « éléments », avec par exemple la possibilité de sélectionner toutes les observations portant sur les formes gazeuses de l’azote (avec le code 126*)



**Figure 3 : Classification des identifiants numériques pour des éléments d’intérêt.
Exemple des différentes formes de l’azote.**

La classification proposée dans DoNuts est évolutive, et permet aisément l’ajout de nouvelles classes (pour décrire les compartiments, les fractions d’analyse ou les éléments d’intérêt). Cette classification se veut suffisamment générique pour prévoir la possible observation de certaines variables, indépendamment de leur présence dans la base DoNuts. Ainsi, actuellement, la base DoNuts dispose d’enregistrements (d’observations) pour 105 variables (càd 105 combinaisons compartiment-fraction-élément).

Tous projets d’intégration de nouvelle famille d’éléments (ex : métaux, HAP, Pesticide, etc.) devra (i) intégrer une étape de réflexion sur l’identification des combinaisons possibles caractérisant les variables observables (compartiments/fractions/éléments) et (ii) proposer une classification générique adaptée.

2.1.2 Les sources de données

Le volet « sources » est constitué d’une seule et unique table regroupant des informations sur le producteur et l’organisme diffusant la donnée. On y trouve également mention de références bibliographique (ou autres liens web), des copyrights et des restrictions éventuelles liées à la rediffusion de cette donnée. DoNuts rassemble actuellement 25 producteurs sur 9 pays (Tab. 1).

Dans sa version actuelle, la quasi-totalité des observations ponctuelles intégrées à la base DoNuts sont issues de producteurs externes aux équipes de recherche du PIREN. Dans une perspective de mutualisation de cette base de données, il sera nécessaire de faire apparaître derrière le diffuseur de données « PIREN », les différentes équipes de recherche produisant des données ainsi que les éventuelles restrictions d’accès/diffusion.

Tableau 1 : Producteurs - diffuseurs de données sollicités dans le cadre du projet EMoSEM (« Qualité » recouvre les principaux paramètres physico-chimique de la qualité des eaux)

Pays	Producteurs - diffuseurs	Débits	Qualité
BE	Hydronet - Vlaamse Milieumaatschappij	X	X
	Aquaphyc DGO3		X
	Aqualim - Région wallonne	X	
	Voies hydrauliques - Région wallonne	X	
CH	Federal Office for the Environment (FOEN)	X	
DE	GRDC (Global Runoff Data Centre)	X	
	German Federal Institute of Hydrology (BfG)	X	
	International Commission for the Protection of the Rhine		X
ES	CEDEX (Centro de Estudios y Experimentacion de Obras Publicas)	X	
	ICA Duero		X
	ICA Guadiana		X
	CH Norte		X
FR	Banque Hydro	X	
	6 Agences de l'eau		X
IE	Hydronet - Environmental Protection Agency (EPA)	X	X
	Office of Public Work (OPW)	X	
NL	Waterbase - Ministry of infrastructure and the Environment	X	X
PT	SNIRH (Serviço Nacional de Informação de Recursos Hídricos)	X	X
	INAG (Instituto da água)		
UK	CEH (Centre for Ecology and Hydrology)	X	
	UK Environment Agency		X

2.1.3 Les stations de mesure

Sous le terme de station, il s’agira de renseigner un lieu d’observation, caractérisé à minima par des coordonnées (X, Y, Z), un code identifiant la station de mesure, un pays d’appartenance, ainsi qu’un type de milieu. Cette dernière information vise à proposer un accès rapide permettant de filtrer par exemple les stations rivière, eaux-souterraine, lac-réservoir, etc. (tout comme les autres informations, la classification des milieux est évolutive et pourra être raffinée).

La table station présente la particularité de stocker une information de type géométrie ponctuelle permettant de visualiser la localisation des sites de prélèvements-mesures via un logiciel SIG. Tous les points sont positionnés (éventuellement re-projeté) dans le système de coordonnées de référence EPSG 3035 (ETRS89-LAEA), en utilisant les coordonnées originales fournies par le producteur de données. La figure 4 indique la répartition spatiale de l’ensemble des stations présentes dans la base DoNuts.

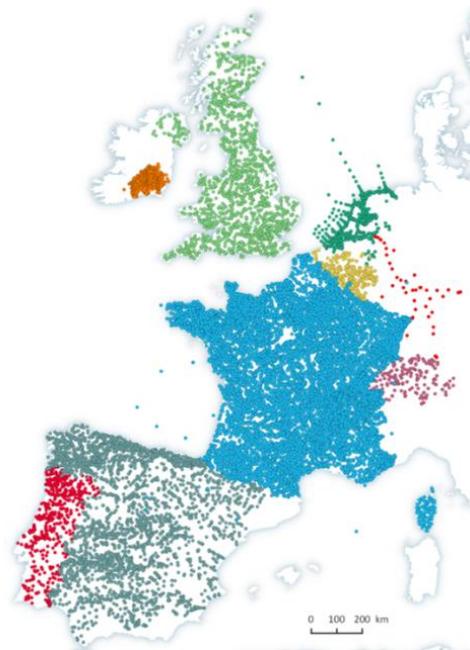


Figure 4 : localisation des stations DoNuts sur les 9 pays inclus dans le domaine du projet EMOSEM. Au total 35 000 stations.

2.1.4 Les mesures

La table « Mesure » centralise les identifiants numériques correspondant à la « Variable observée » d’intérêt, la « Station » de mesure, et la « Source » de la donnée. Chaque mesure est alors caractérisée par une date (jour et heure de la mesure) et doit impérativement faire mention d’une résolution temporelle. Il est ainsi possible de stocker des acquisitions instantanées, journalières, décadaire (10 jours), mensuelle ou annuelle.

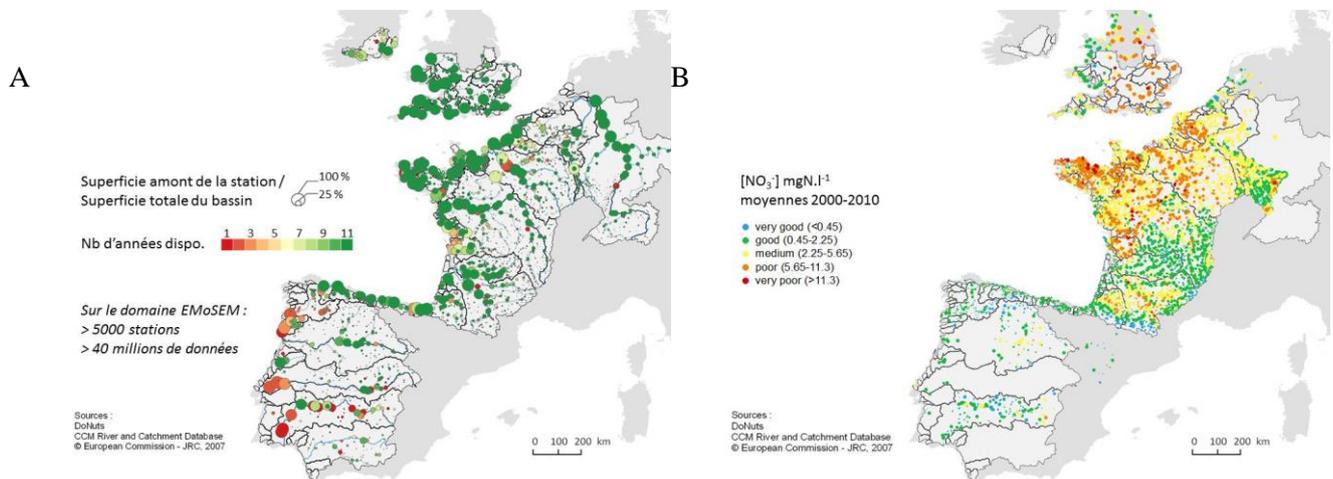


Figure 5 : disponibilité des mesures de débits (A) et contamination nitrique moyenne (B) en rivière sur la période 2000-2010

Dans le cadre du projet EMOSEM, c’est plus de 70 millions de mesures qui ont ainsi été collectées sur l’ensemble des bassins versants étudiés. À titre d’exemple la figure 5 permet d’apprécier le volume de données disponibles pour deux variables d’intérêt : les débits et les concentrations nitriques en rivière.

2.2 Enrichissement des observations disponibles dans DoNuts

2.2.1 Spatialisation des observations « rivière » sur un réseau hydrographique

L'intérêt d'un système de gestion de base de données disposant d'une extension « spatiale » est de pouvoir facilement combiner à des analyses statistiques classiques des critères géographiques. Il sera par exemple aisé de croiser la localisation des stations de mesures avec des informations relatives à l'occupation des sols, au contexte lithologique, etc.

Il est également extrêmement utile dans nos recherches de pouvoir analyser ces observations ponctuelles à l'aide de fonctions topologiques comme les parcours de réseaux hydrographiques et permettre ainsi la sélection directionnelle de stations se situant à l'aval d'un point ou au contraire dans un sous-bassin amont.

Ce dernier type d'analyse s'appuie sur le choix préalable d'un réseau hydrographique de référence (donnée SIG). La précision relative des coordonnées des stations de mesure, l'approximation géométrique des données SIG représentant les réseaux de drainage par ailleurs souvent très complexe (zones de confluences multiples, diffuence, canalisation, etc.) sont autant de facteurs qui rendent l'association d'une station de mesure à un tronçon de rivière potentiellement inexacte.

Ce travail a été réalisé pour l'ensemble du domaine EMoSEM, en se basant sur la base européenne CCM 2.1 (Catchment Characterisation and Modelling) produite par le JRC (Vogt et al. 2007). Disponible pour l'ensemble de l'Europe continentale cette base conserve une précision relative lorsqu'il s'agit d'étudier un bassin individuel. À titre d'exemple, le bassin de la Seine y est décrit avec 5793 bassins versant élémentaires d'une taille moyenne de 13 km² (contre 7123 bassins de 10.7 km² en moyenne dans les données AESN utilisées en modélisation). L'accrochage des 35000 stations DoNuts, se base sur le calcul de plusieurs indicateurs : les distances aux trois cours d'eau les plus proches, la comparaison des superficies amont, le changement éventuel de bassin-versant, l'analyse des toponymes. L'ensemble de ces indicateurs permet de détecter des erreurs relatives à une mauvaise localisation des stations, ou une représentation erronée d'une rivière.

DoNuts offre la possibilité de travailler avec plusieurs réseaux hydrographiques de référence, afin de s'adapter à l'emprise et à la résolution des différents projets de recherche. Un travail similaire sur le réseau hydrographique couvrant le territoire de l'AESN et développé pour les besoins du PIREN-Seine est par ailleurs en cours de réalisation.

2.2.2 Des protocoles établis pour l'intégration et l'amélioration en continu des données

La chaîne d'intégration conçue pour la base DoNuts (Fig. 6) s'initie avec la collecte d'informations auprès du producteur de données. Il est à ce stade crucial d'obtenir une extraction standard de données « brutes », quel qu'en soit le format, ceci afin d'assurer la répétabilité des traitements qui seront ensuite mis en œuvre en vue d'une intégration dans la base DoNuts.

Deux éléments constituent la base d'une intégration de nouvelles données et excluent tous prétraitements manuels des données brutes :

- une grille de lecture et de conversion doit fournir, pour une sélection de variables d'intérêts, la correspondance entre l'information brute fournie par le producteur de données et les possibles combinaisons « compartiment – fraction – élément » propres à DoNuts.
- un script d'importation des données, gérant les cas d'ajout simple ou de mise jour, gère les conflits et convertit les observations (unités et identifiants des variables) en s'appuyant sur la grille de lecture préalablement définie.

Ce protocole permet d'élaborer des procédures d'intégration spécifiques et assure la possibilité de corriger à la source (et non en post-traitement) certains problèmes de qualité de données propres à chaque producteur de données. Ainsi, la détection d'erreurs (unités mal renseignées, valeurs extrêmes ou aberrantes) doit en premier lieu conduire à une révision des « scripts d'importation » ou des « grilles de lecture et de conversion » concernés.

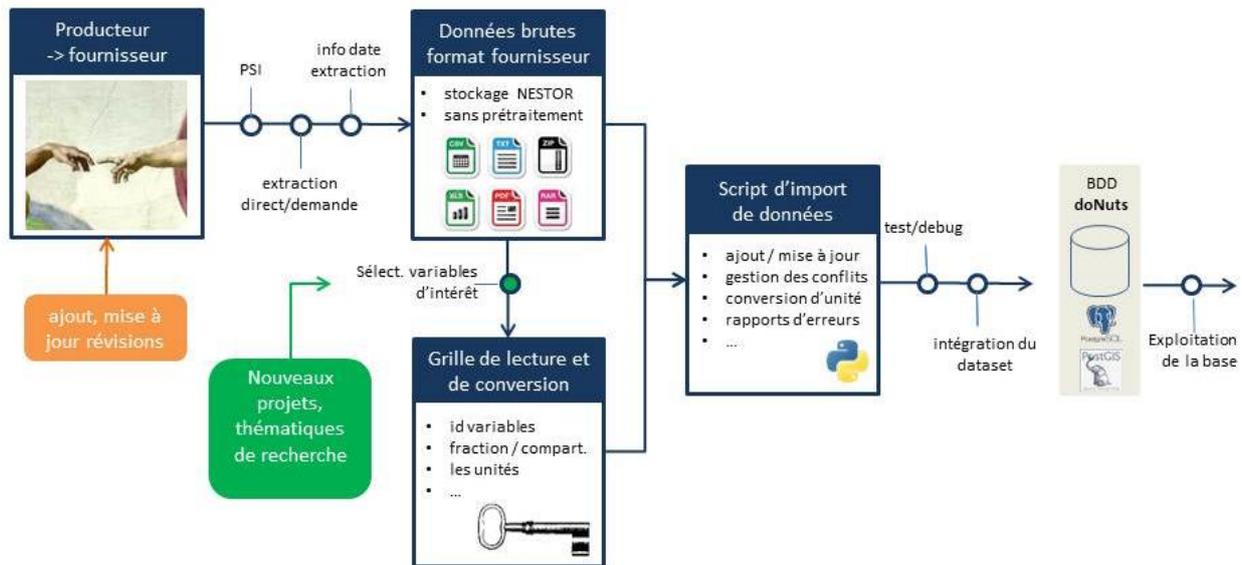


Figure 6 : Principales étapes d’intégration de données dans la base DoNuts

Les codes qualité fournis par le producteur de la donnée font partie des données stockées dans la table « Mesures ». L’option retenue dans DoNuts en cas de mise à jour (ou de correction) est de conserver en plus de la nouvelle donnée, la date et la dernière valeur avant la mise à jour.

L’investissement important, nécessaire à la mise en place de nouveaux protocoles d’intégration pour des données provenant de nouveaux producteurs est rapidement compensé par :

- l’absence d’erreur inhérente au prétraitement manuel des données ;
- la répétabilité inévitablement nécessaire, pour opérer des corrections, des mises à jour, des compléments sur des chroniques d’observation, etc. ;
- la facilité d’importer une variable d’intérêt supplémentaire dans le cas d’un producteur de données disposant déjà d’un protocole d’intégration.

2.2.3 Quelles opportunités pour les équipes du PIREN ?

La proposition faite par DoNuts est celle d’un système structuré et pérenne d’archivage des données relatives aux observations environnementales utilisées, enrichies ou produites par les équipes du PIREN.

Le premier objectif est de ne pas perdre l’information au fil des actions de recherche et de répondre aux besoins de stockage, de sauvegarde, de diffusion et de partage des données, qui restent les seuls et uniques moyens de ne pas laisser une donnée tomber dans l’oubli (cf. cycle de vie).

À cette notion de continuum temporel, s’ajoute l’idée d’un continuum spatial, structuré en premier lieu autour des réseaux hydrographiques, mais intégrant également des observations des milieux souterrains, des milieux terrestres ou encore atmosphériques. En rassemblant au sein d’une structure homogène des données issues d’entités administratives différentes (agences de bassin, réseaux de surveillance nationale, données fédérales, etc.), la base DoNuts permet également de s’extraire du canevas géographique classique des producteurs de données et de travailler sur de nouveaux territoires en faisant abstraction des limites administratives.

C’est aussi une opportunité d’accéder à de nouvelles variables interopérables, de pouvoir croiser au sein d’une même base de données et autour d’objet ou de territoire communs, des chroniques d’observations environnementales relevant de thématiques scientifiques différentes et pouvant potentiellement générer de nouvelles questions de recherche, des data-papers, etc.

3 Développement d’un outil web-cartographique (CocoNuts)

Si l’administration d’une base de données comme DoNuts nécessite des connaissances expertes, il semble néanmoins indispensable de pouvoir offrir aux scientifiques et aux partenaires du PIREN, un accès simple et interfacé, permettant l’interrogation et la visualisation des chroniques d’observations environnementales disponibles. Un outil web-cartographique (CocoNuts) a ainsi été développé pour permettre la sélection, l’exploration, l’analyse et le téléchargement de données d’intérêt. Cette solution présente l’avantage d’offrir un accès immédiat à n’importe quel utilisateur, sans installation logicielle préalable.

CocoNuts vient, au côté du modèle pyNuts et de la base de données DoNuts, compléter le champ lexical des développements gastronomiques de l’équipe historiquement baptisée C-Nuts. « Coco » acronyme de « CartO and Charts for environmental Observations » a été initialement développé par Corentin Souton dans le cadre de son stage de 2ème année de Master STIC BDIA.

3.1 Présentation fonctionnelle de l’application web-carto-graphique CocoNuts :

3.1.1 Principe général de l’exploration des données avec CocoNuts

L’interface web CocoNuts propose de progresser en deux temps dans l’exploration des observations environnementales de la base DoNuts.

La première étape permet au visiteur (i) de visualiser, sur un fond cartographique paramétrable, les stations de mesures/prélèvements disponibles, (ii) d’accéder à une sélection d’informations simples permettant, (iii) d’établir une liste d’une ou plusieurs stations d’intérêts pour lesquelles la base DoNuts sera interrogée.

La seconde étape amène à une interface de visualisation de graphiques dans laquelle il est possible d’explorer (sous formes de chroniques temporelles) les observations attenantes aux stations préalablement sélectionnées.

Les technologies et les fonctionnalités des deux interfaces carto- et graph-ique sont reprises ci-après, avant de présenter une vision plus complète de l’architecture mise en place et son administration

3.1.2 La sélection de stations d’intérêts

Le volet cartographique propose à l'utilisateur de sélectionner des stations d'observations d'intérêt. Cette première interface se base sur le client web Lizmap qui permet la publication sur internet de projets cartographiques interactifs. Elle fournit alors les outils classiques de tous logiciels SIG : déplacement, zoom, diverses mesures, recherche d'entité dans une couche de données, etc. L'interface intègre également la gestion de comptes authentifiés qui permettent alors de définir des groupes d'utilisateurs auxquels sont associés des droits spécifiques : accès restreint à certains projets cartographiques et à certaines fonctionnalités (interrogation, export de données, etc.).

En tant que simple viewer SIG, Lizmap ne dispose pas d’une fonctionnalité permettant la sélection d’entités (simples ou multiples) au sein d’une couche d’information. Le développement de cette fonctionnalité a donc fait partie des missions de Corentin Souton au cours de son stage de Master 2.

La partie client de Lizmap prévoit l'ajout de code en Javascript qui agit en surcouche du code de base et permet d'étendre les fonctionnalités de l'application sans modifier son cœur. Ce système facilite les procédures de mise à jour éventuelle du client web.

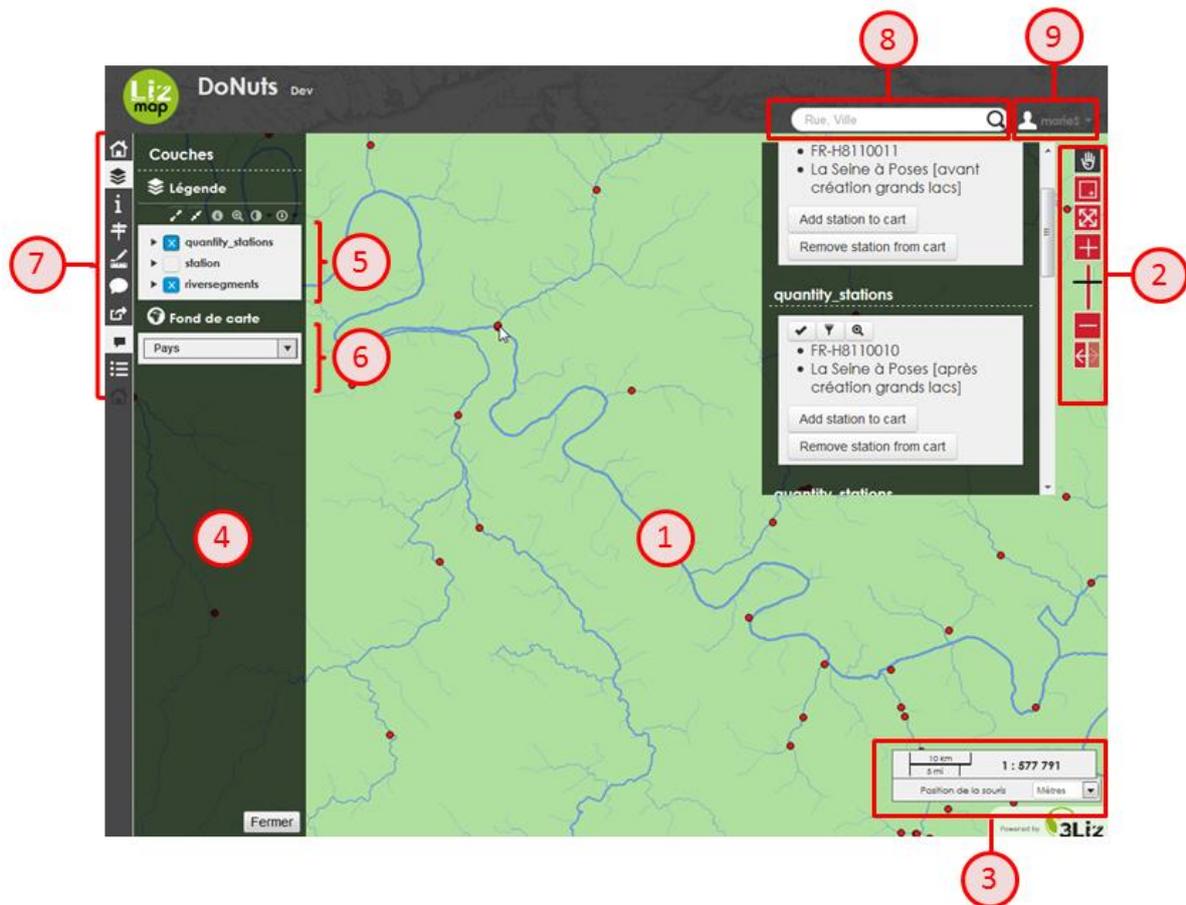


Figure 7 : Volet cartographique (Lizmap) de l’interface CocoNuts. (1) Espace de visualisation cartographique (2) Outils de déplacement et zoom (3) Echelle et unité (4) Table des matières (5) Couches SIG interrogeables et sélectionnables (6) Fonds de carte (7) Outils de mesures, info-bulles, accès aux attributs (8) Géolocalisation (9) Accès authentifié

L’outil de sélection de stations développé en Javascript génère du code HTML qui ajoute un bouton à l’interface web et y associe des événements. Au clic sur une entité géographique, une infobulle apparaît et affiche le nom et le code de la ou les stations (cette information est paramétrable). Cette petite fenêtre intègre des boutons permettant d’ajouter les entités au « panier » ou de les retirer si elles sont déjà présentes. L’ensemble des stations de ce panier de sélection apparaît en surbrillance sur la carte. Le panier est accessible à tout moment lors de la session et peut être modifié. Une fois la sélection effectuée, la liste des identifiants des stations est envoyée comme formulaire HTML au deuxième volet de visualisation de graphiques.

Il est à noter que si l’entité sélectionnée est un tronçon de rivière, une opération de calcul du bassin amont sera effectuée lors du transfert vers la partie graphique pour obtenir les identifiants des stations uniquement situées dans l’emprise spatiale de ce bassin. Ces calculs de bassin versant amont se font actuellement sur les données issues de la base CCM 2.1 (cf. 2.2.1 Spatialisation des observations « rivière » sur un réseau hydrographique) mais d’autres référentiels hydrographiques pourraient être intégrés.

3.1.3 L'analyse des chroniques d'observation DoNuts

Une fois la sélection des stations d'intérêts validée, l'ensemble des observations correspondantes est visualisable dans le deuxième volet de l'application (Figure 8).

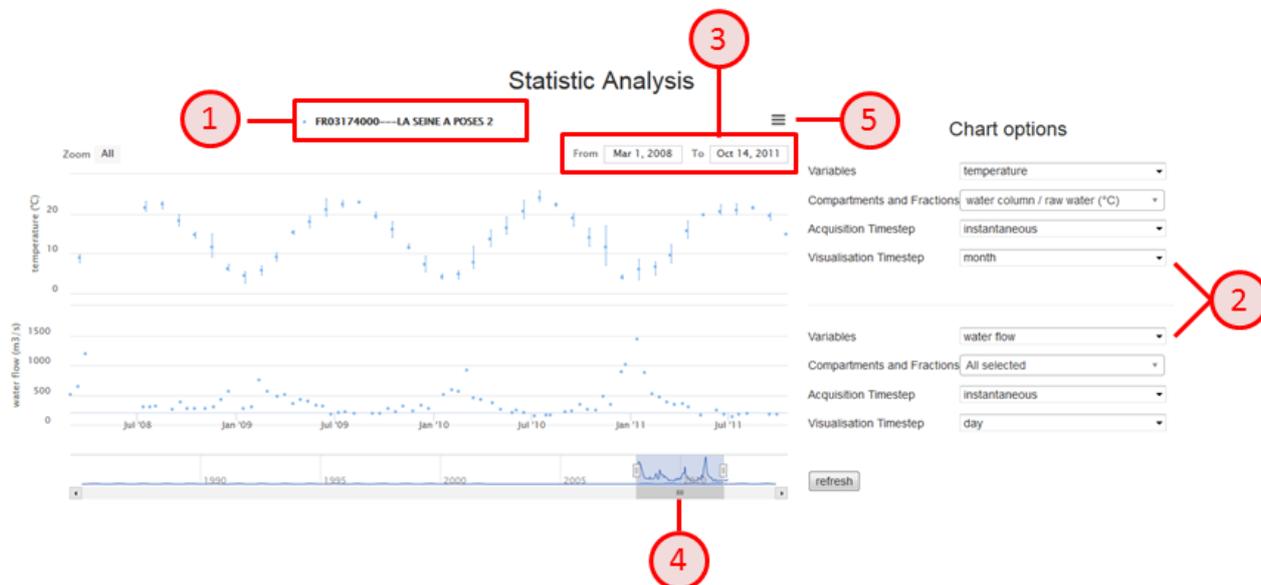


Figure 8 : Volet d'exploration graphique (HighChart) de l'interface CocoNuts.

La liste des stations est récapitulée (Figure 8 - 1), une couleur est attribuée à chacune pour l'affichage des données sur les graphiques et l'utilisateur a la possibilité de masquer une à une les stations en cliquant sur leur nom.

L'utilisateur peut choisir de représenter les variations temporelles d'une ou plusieurs variables (combinaison entre un compartiment prélevé et une fraction analysée pour un élément d'intérêt). Les observations peuvent être sélectionnées en fonction de leur pas de temps d'acquisition (instantané pour la grande majorité, journalier, décadaire, etc.) et visualisées à des résolutions temporelles égales ou supérieures (dans ce dernier cas, chaque valeur moyenne sera systématiquement encadrée par les min et max observés) (Figure 8 – 2).

La plage temporelle visualisée correspond dans un premier temps à la disponibilité de la ou les variables sélectionnées. L'utilisateur peut ensuite choisir de restreindre cette emprise, soit en entrant directement les dates de début et de fin (Figure 8 - 3) soit en manipulant les curseurs de la barre de temps située sous les graphiques (Figure 8 – 4). Si plusieurs variables sont sélectionnées (et donc s'il y a plusieurs graphiques), ce paramétrage est automatiquement appliqué à tous les graphiques.

Au survol de la souris sur les points représentés sur les graphiques, une infobulle apparaît et affiche les informations détaillée correspondantes :

- pour des données instantanées : date et heure, valeur,
- pour des données moyennées : plage temporelle, valeur moyenne, minimum, maximum et nombre de valeurs.

L'ensemble des données affichées sur les graphiques ainsi que les informations relatives aux stations (nom complet, coordonnées de localisation, etc.) et aux métadonnées (producteur de données, unités, etc.) peuvent être exportées par l'utilisateur (Figure 8 - 5).

Ce volet a été développé dans le cadre du stage de Master 2 de Corentin Souton. À la suite d'une analyse comparative détaillée des différentes solutions existantes pour l'affichage de graphiques interactifs sur le web, le choix s'est porté sur la librairie Javascript libre HighCharts. Son faible niveau de dépendances, la grande variété de type de graphiques proposés ainsi que la gestion de grands volumes de données en font un outil puissant et parfaitement adapté aux besoins de ce volet. Toutefois, HighCharts ne gère que les données au format JSON, ce qui a nécessité un traitement sur les mesures obtenues depuis la base de données DoNuts. D'autre part, cette bibliothèque graphique propose des types de graphiques par défaut mais ils ont dû être adaptés.

Ce volet suit une architecture Modèle-Vue-Contrôle (MVC), indépendante du volet cartographique. Pour faciliter le déploiement d'une telle architecture, le framework CodeIgniter a été utilisé. Les développements effectués sur la base de ce framework ont consisté à créer des objets PHP qui génèrent le code Javascript et HTML de la page.

Le contrôleur reçoit et traite les informations venant de la sélection du volet cartographique et crée les modèles de données : un objet de type HighCharts va définir la structure du graphique à l'aide de différents attributs comme les axes, le titre, la représentation des données, etc. Les mesures provenant de la base de données sont alors intégrées dans cet objet via PHP. Les vues, qui structurent la page, génèrent d'une part, des blocs HTML contenant le graphique et les différents boutons d'interactions et, d'autre part, du code Javascript pour permettre notamment l'affichage des graphiques et la gestion des événements, interactivité avec l'utilisateur.

3.1.4 L'architecture de CocoNuts : un chaînage de technologies libres

L'architecture mise en place est basée sur un chaînage QGIS-Server / LizMap (pour la partie web-cartographique) et couplée à la technologie Javascript HighCharts (pour la création de graphiques et l'analyse statistique). L'ensemble de ces logiciels sont sous licence libre. Ils ont été installés sur un serveur hébergé et administré par le service informatique de l'UMR METIS dont l'accès est pour le moment restreint à l'équipe de développement. La vue d'ensemble de l'architecture CocoNuts et son fonctionnement sont schématisés sur la Figure 9.

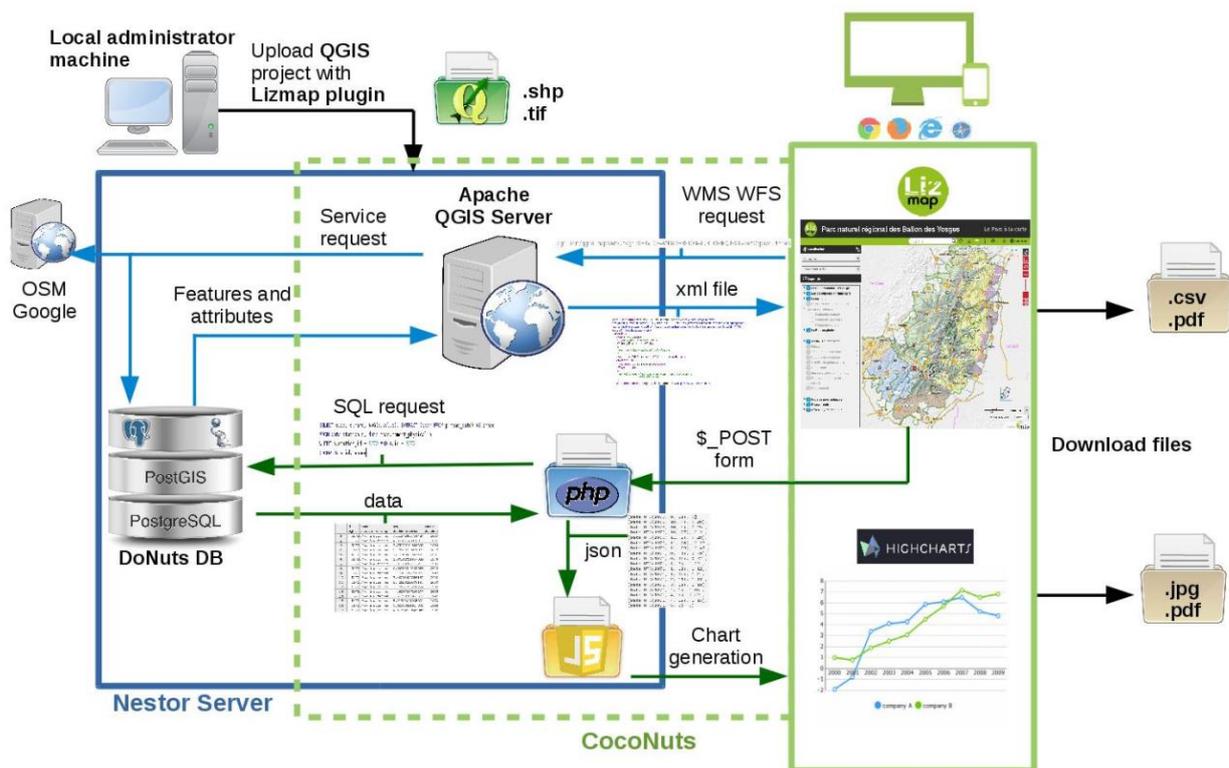


Figure 9 : Architecture client-serveur du chaînage web-carto-graphique CocoNuts

Le point de départ du fonctionnement de CocoNuts se situe sur l'ordinateur d'un administrateur de l'interface qui va gérer la liste des projets cartographiques à diffuser, leurs contenus ainsi que les groupes d'utilisateurs et leurs droits associés. Un projet cartographique se crée donc d'abord sur un ordinateur en local, via le logiciel QGIS. Le projet QGIS va permettre de définir la liste des couches à afficher, leur style, les étiquettes, les plages d'échelles de visibilité de chaque donnée, les alias des noms des couches et des champs des tables attributaires, etc. Les couches peuvent provenir soit de bases de données PostgreSQL (par ex. la

couche des stations de la base DoNuts), soit de données vecteur ou raster en local, soit de données externes type WMS/WFS (par ex. fonds Openstreetmap, Google Maps). Une extension du logiciel QGIS, LizMap Plugin, va générer un fichier de configuration nécessaire au client web. Ce fichier accompagné du fichier de projet QGIS et des données locales éventuelles doivent ensuite être chargés sur le serveur.

Les fichiers de configuration sont lus par QGIS Server qui joue le rôle de serveur de données pour le web et produit des flux de données cartographiques. Ces flux sont intégrés à l'interface web générée par LizMap Web Client qui dispose des fonctionnalités classiques d'un logiciel SIG enrichies des outils de sélection présentés précédemment.

La sélection de station faite depuis le volet cartographique est envoyée via un bouton au volet d'exploration graphique sous forme de requête POST. Ce deuxième volet utilise donc la liste des identifiants des stations demandées pour envoyer une requête SQL à la base DoNuts et récupérer les mesures associées aux stations ainsi que diverses métadonnées (informations détaillées sur les stations, les producteurs, les unités). Le traitement de ces données (conversion au format JSON, etc.) permet enfin à la librairie Javascript HighCharts d'afficher les graphiques.

3.2 Perspectives d'évolution et d'utilisation par le PIREN :

L'une des fonctionnalités offertes par LizMap est la possibilité de générer de multiples projets cartographiques accessibles depuis une même page web et dont les droits d'accès sont gérés au cas par cas. Comme précisé précédemment, la diffusion de nouveaux projets web-cartographiques passe par la création et la configuration de ces projets via le logiciel QGIS mais ne nécessite aucun développement supplémentaire. Ainsi, grâce aux capacités de rendu cartographique de QGIS et la simplicité de création des fichiers de configuration, il est aisé de concevoir des projets web-cartographiques complexes et « sur-mesure » centrés sur un territoire d'intérêt, avec une sélection d'observations pertinentes accompagnées de fonds de carte et d'analyse thématiques adaptées.

Avec l'intégration de données produites par les équipes de recherche du PIREN à la base DoNuts, il est donc envisageable de créer des projets PIREN thématiques et/ou focalisés sur un site atelier tel que la Bassée, l'Orgeval, etc. La création de tels projets serait alors adossée à la gestion de groupes d'utilisateurs spécifiques qui permettent ainsi une gestion fine des droits d'accès.

À l'heure actuelle, certaines fonctionnalités de l'architecture CocoNuts sont encore en développement : l'export de données et de métadonnées, la possibilité de visualiser des profils en longs pour lesquels l'utilisateur choisit les points amont et aval. D'autre part, des documentations utilisateur et développeur sont en cours de rédaction.

La gestion des accès est à préciser : seuls les membres de l'équipe de développement ont les droits de visualisation (et de modification) aux projets cartographiques tests actuellement en ligne.

Un groupe d'utilisateurs « PIREN » est donc à créer ainsi qu'un projet générique couvrant l'ensemble de la Zone Atelier réunissant des informations pertinentes telles qu'une couche géographique sur les sites atelier avec une description courte des travaux de recherche, des photos, etc. Pour une intégration cohérente avec le site internet du PIREN, une discussion devra avoir lieu afin de déterminer dans quelle mesure les éléments de charte graphique du site PIREN pourront être réutilisés et des liens créés.

Conclusion

Si le cadre réglementaire (arrivée de l'Open-data, PSI sur la réutilisation des données publiques, convention d'Aarhus sur l'accès à l'information environnementale, directive INSPIRE sur la diffusion en ligne des données, etc.) s'établissant autour des données environnementale est de plus en plus contraignant, il nous incite à réfléchir collectivement pour mettre en place des solutions concrètes permettant de gérer les données issues de la recherche de façon plus efficiente.

Développé par l'équipe de modélisation C-Nuts de l'UMR METIS et la FIRE, DoNuts permet de regrouper en une seule base un ensemble de données d'observations ponctuelles en harmonisant les formats de données,

les variables, les unités de mesures et la localisation des stations. Déjà fonctionnelle et couplée au modèle PyNuts-Riverstrahler, son utilisation pourrait aisément être étendue. En effet, de par sa généralité et les technologies libres qu'elle emploie, DoNuts propose une structure de données opérationnelle pour intégrer de nouvelles bases, par exemple produites par les équipes de recherche du PIREN Seine.

L'architecture CocoNuts, dont les développements ont été initiés par le stage de Master 2 de Corentin Souton et qui poursuit actuellement son travail en CDD, propose une interface web de visualisation, d'exploration et d'analyse du contenu de la base DoNuts via deux volets. Le premier volet, cartographique, associé à la localisation des stations d'observation des données géographiques thématiques pertinentes et, grâce à différents outils SIG (recherche dans les données, sélection de bassin amont, etc.), permet d'effectuer une sélection experte de stations. Le deuxième volet permet l'analyse des données associées à cette sélection en laissant le choix à l'utilisateur des variables et des pas de temps de visualisation ainsi que leur téléchargement.

La généralité des développements produits permettra d'envisager, à terme, une déclinaison facilitée de ces outils en différents projets. Ainsi, l'architecture CocoNuts pourra héberger un projet web-carto-graphique PIREN Seine général et des sous-projets sur des thématiques ou des sites ateliers tout en gardant le contrôle des accès aux données grâce à la gestion de groupes d'utilisateurs associés à des droits spécifiques.

Référence :

- Alliance nationale de recherche pour l'environnement (2014) Compte-rendu du séminaire : Les données accessibles dans le domaine environnemental. Quels freins ? Co-organisé par Michel Guiraud et Cécile Callou (MNHN) - 21 mai 2014
- Billen, G., Garnier, J., and Hanset, P. (1994). Modelling phytoplankton development in whole drainagenetworks: the RIVERSTRAHLER Model applied to the Seine river system. *Hydrobiologia* 289, 119-137.
- Desmit et al.(2015). EMOSEM Final Report - Ecosystem Models as Support to Eutrophication Management In the North Atlantic Ocean. <http://archimer.ifremer.fr/doc/00292/40301/>
- Garnier J., Billen G. & Coste M. (1995). Seasonal succession of diatoms and Chlorophyceae in the drainage network of the river Seine: Observations and modelling. *Limnol. Oceanogr.* 40: 750-765.
- Leobet M. (2014) Conférence en ligne : directive INSPIRE, où en est la France en 2014 ? - <http://professionnels.ign.fr/conference-inspire>
- Vogt et al. (2007). A pan-European River and Catchment Database. EC-JRC (Report EUR 22920 EN) Luxembourg, 120 p.
- Thieu, V., Silvestre, M., Billen, G., Garnier, J., Passy, P., Lassaletta, L. (2015) Transferts et exports de nutriments dans les continuums rivières – zones côtières - Vers une application générique du modèle biogéochimique Riverstrahler aux grands fleuves de la façade Atlantique Nord-Est. 2nd international conference IS-Rivers, Lyon – France