



Livrable 1.3: Plan de Gestion des Données

Alejandra Leydy, Pierre-Antoine Versini

► To cite this version:

Alejandra Leydy, Pierre-Antoine Versini. Livrable 1.3: Plan de Gestion des Données. Ecole des Ponts Paris Tech, Paris-France. 2020. hal-04144757

HAL Id: hal-04144757

<https://enpc.hal.science/hal-04144757>

Submitted on 28 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Projet ANR 17-CE22-0002-01
EVNATURB
Programme JCJC 2017

Livrable 1.3: Plan de Gestion des Données



Leydy Alejandra CASTELLANOS DIAZ (HM&Co-ENPC)

Pierre-Antoine VERSINI (HM&Co-ENPC)

Juin 2020



Résumé de la démarche

L'objet de ce livrable est de présenter le Plan de Gestion des Données (PGD, ou Data Management Plan en anglais) mis en place dans le cadre du projet EVNATURB. Ce plan a pour objectif de présenter la manière dont les données générées et utilisées dans le projet sont gérées.

Plus spécifiquement, le PGD aborde les points suivants :

- La description des données collectées et/ou créées
- Les standards, formats et méthodologies appliqués sur le paquet de données
- Les questions d'ordre éthiques, de propriété intellectuelle et de restrictions (les éventuelles périodes d'embargo)
- Les prévisions pour le partage et l'ouverture des données
- La stratégie de la préservation à long terme (archivage)

Pour ce faire, l'outil OPIDoR (Optimisation du Partage et de l'Interopérabilité des Données de la Recherche, <https://opidor.fr/>) a été utilisé. Ce portail est mis à disposition de la communauté de l'Enseignement Supérieur et de la Recherche par l'Inist-CNRS

On trouvera donc à la suite le document formalisé produit sous OPIDoR.

EVNATURB - Evaluation des performances éco-systémiques d'une renaturation du milieu urbain

Plan de gestion de données créé à l'aide de DMP OPIDoR

Créateur du PGD : Pierre-Antoine Versini

Affiliation du créateur principal : Ecole des Ponts Paris Tech

Modèle du PGD : ANR - Modèle de PGD (français)

Dernière modification du PGD : 13/07/2020

Financier : Agence nationale de la recherche (ANR)

Numéro de subvention : ANR-17-CE22-0002-01

Résumé du projet :

Le projet ANR EVNATURB, prévu sur 4 ans, a pour objectif de développer une plateforme facilitant l'évaluation de certains des services éco-systémiques prodigués par les solutions bleues-vertes (BGS) à l'échelle du projet d'aménagement, et de promouvoir une renaturation de la ville. Pour tenter de résoudre les problèmes complexes liés à ce sujet, EVNATURB projette de mettre en œuvre une approche multidisciplinaire, multi-échelle et multi-physique pour :

- coupler hydrologie, thermique, urbanisme, biodiversité, en lien avec leurs services éco-systémiques,
- caractériser les variabilités spatio-temporelles des processus physiques concernés sur une large gamme d'échelles à l'aide de : (i) la mise en œuvre d'un système de mesures haute résolution, transportable et innovant, (ii) l'utilisation d'outils analytiques et de modélisations adaptées,
- développer un réseau scientifique dédié au suivi expérimental des BGS, et qui contribuera à améliorer nos connaissances sur le fonctionnement et le suivi de ces infrastructures,
- caractériser l'implantation de BGS en prenant en compte les enjeux socio-économiques locaux, et définir des indicateurs quantitatifs pertinents en regard du projet d'aménagement (certification, labellisation, respect des réglementations, continuité écologique).

Le projet EVNATURB doit faire le lien entre les méthodes et résultats produits par la recherche actuelle en instrumentation et en modélisation urbaine, les problèmes pratiques émanant des sciences de l'environnement et les besoins opérationnels de la planification urbaine. L'approche multi-échelle adoptée lui permettra de prendre en considération la variabilité spatio-temporelle des processus considérés, particulièrement forte en milieu urbain. En prenant l'hydrologie comme point de départ, elle s'étendra aux domaines connexes tels que la thermique et la biodiversité. Les activités expérimentales et de modélisation liées à l'estimation des flux d'évapotranspiration seront particulièrement pertinentes pour améliorer nos connaissances sur le comportement thermo-hydrigue en lien avec la végétation, et donc le rafraîchissement urbain.

Les données collectées pendant le projet proviennent essentiellement du site de la Vague Verte de Champs-sur-Marne. Elles comprennent des mesures effectués sur le bilan hydrique (aptitude de la structure végétalisée à agir sur la gestion des eaux pluviales) et le bilan thermique (aptitude de la structure végétalisée à atténuer les îlots de chaleur).

D'autres données seront mises à disposition par les partenaires du projet (comme le Cerema par exemple).

Une plateforme unique permettra d'importer et d'exporter ces données.

Chercheur Principal : Pierre-Antoine Versini

Identifiant ORCID : 0000-0001-8041-2184

Contact pour les Données : Pierre-Antoine Versini

Produits de recherche :

1. Bilan hydrique BGW : Données du bilan hydrique réalisé sur la Vague Verte de Champs-sur-Marne (Jeu de données)
2. Bilan thermique BGW : Données du bilan énergétique réalisé sur la Vague Verte de Champs-sur-Marne (Jeu de données)
3. Campagne ETP : Données d'évapotranspiration mesurées à l'aide d'une chambre à transpiration sur la Vague Verte de Champs-sur-Marne (Jeu de données)

EVNATURB - Evaluation des performances éco-systémiques d'une renaturation du milieu urbain

1. Description des données et collecte ou réutilisation de données existantes

Bilan hydrique BGW : Données du bilan hydrique réalisé sur la Vague Verte de Champs-sur-Marne

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les données issues de la Vague Verte de Champs-sur-Marne sont le résultat d'une instrumentation débutée en 2013 à travers le projet Européen Blue Green Dream de la KIC Climate et le projet RadX-IdF. Cette instrumentation se poursuit aujourd'hui dans le cadre du projet ANR EVNATURB. C'est à travers ce projet que sont maintenant collectées une partie des données relatives au bilan hydrique de la Vague Verte.

Il est à noter que les mesures de précipitation sont réalisées sous la responsabilité d'Auguste Gires dans le cadre de l'observatoire TARANIS de la plate-forme Fresnel d'HM&Co.

Le bilan hydrique représente la relation entre les entrées (précipitations), le stockage (rétention dans le substrat) et les sorties (débit en sortie d'ouvrage) en eau de la Vague Verte. Celui-ci ne tient pas compte du processus d'évapotranspiration qui a généralement lieu en dehors des périodes de précipitations.

La mesure du bilan hydrique comprend :

- La précipitation : Un disdromètre installé sur le toit du Bâtiment Carnot de l'ENPC donne accès à des mesures de tailles et de vitesse des gouttes de pluie toutes les 30 s. Les données sont collectées à l'aide d'un raspberry puis transférées sur les serveurs de stockage de l'ENPC
- La teneur en eau : Un réseau de 16 sondes TDR Campbell mis en place sur la surface végétalisée fournit des mesures de permittivité diélectrique toutes les 4 minutes. Les données sont collectées à l'aide d'une station d'acquisition CR6 connectée au serveur de stockage de l'ENPC
- Le débit : des capteurs de hauteur d'eau sont placés (i) dans une conduite reprenant 1143 m² de la Vague Verte (toutes les 30 secondes), (ii) dans une unité de stockage reprenant 3500 m² de la Vague Verte - dont les 1143 m² précédents (toutes les 15 secondes). Les premières sont collectées à l'aide d'une station d'acquisition CR6 connectée au serveur de stockage de l'ENPC, les secondes à l'aide d'un micro-contrôleur Arduino. Pour ce dernier la collecte et le transfert sur le serveur de stockage de l'ENPC se font manuellement.

Des références aux travaux de mesures cités sont à retrouver dans les data paper suivants :

Gires, A., Tchiguirinskaia, I., Schertzer, D.: Two months of disdrometer data in the Paris area, Earth Syst. Sci. Data, 10, 941–950, <https://doi.org/10.5194/essd-10-941-2018>, 2018

Versini, P.-A., Stanic, F., Gires, A., Schertzer, D., Tchiguirinskaia, I.: Data for "Measurement of the water balance components of a large green roof in Greater Paris Area", <https://doi.org/10.5281/zenodo.3467300>, 2019

Lors de l'exportation de données, les informations sur la provenance (propriétaire, lieu, période, remarques) des données seront rappelées soit dans le titre des fichiers, soit dans un document complémentaire.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Il s'agit ici de mesures en continue, c'est à dire que pour un pas de temps donné, une valeur numérique de pluie, de teneur en eau ou de hauteur d'eau est estimée. Ces données sont mises à disposition sous format de fichiers texte (.csv) où chaque ligne correspond à un pas de temps (les exemples sont donnés ici pour le mois de janvier 2020).

- Pluie (2020_0101_0201_Data_rainfall.csv) : ce sont des mesures réalisées à l'aide d'un disdromètre situé sur le toit de l'ENPC. Pour un pas de temps de 30 secondes, l'intensité de la pluie est exprimée en mm/h.
- Débit : les valeurs de débits sont données pour deux emplacements. Le premier (2020_0101_0201_Data_Q_conduit.csv) correspond à une mesure de hauteur d'eau dans une conduite reprenant 1143 m² de la Vague Verte (30 secondes). Le second (2020_0101_0201_Data_Q_storage.csv) correspond à une mesure de hauteur d'eau dans une unité de stockage reprenant 3500 m² de la Vague Verte - dont les 1143 m² précédents (15 secondes). Les données fournies sont

exprimées en mV. Elles sont ensuite converties en hauteur d'eau (mm), puis en débit (l/s).

- Teneur en eau (2020_0101_0201_Data_vwc.csv) : ce sont des mesures de permittivité diélectrique réalisées à l'aide d'un réseau de 16 sondes TDR. Elles sont réalisées à une fréquence de 4 minutes et sont sans unité. Elles sont ensuite converties en teneur en eau (sans unité).

Un programme Python Read_Data_BGW.py (développé au sein du laboratoire HM&Co) est mis à disposition pour convertir ces données brutes en mesures hydrologiques et les représenter graphiquement.

Bilan thermique BGW : Données du bilan énergétique réalisé sur la Vague Verte de Champs-sur-Marne

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Le bilan thermique ou énergétique représente la relation entre les quantités d'énergie surfacique (exprimée en W/m²) reçues, émises et stockées par une surface (ici végétalisée). Cela comprend les flux de chaleur sensible (convection), de chaleur latente (évapotranspiration), de conduction dans le sol (ainsi que son stockage), et de rayonnement (incident et réfléchi). L'hypothèse est faite ici d'un footprint nul (flux de chaleur anthropique).

Réalisées lors de campagnes ponctuelles de quelques heures (6-7h), ces données comprennent des mesures continues (série de données moyennées sur 10 minutes) de rayonnement net (CNR4, fréquence de 5 minutes), de températures dans le sol (4 thermocouples Type-K, fréquence de 5 minutes) et d'indice de réfraction de l'air (scintillomètre LAS, fréquence de 10 minutes). Ces données sont collectées à l'aide d'une station d'acquisition CR3000 qui est ensuite connectée à un ordinateur. Ces données sont par la suite traitées par différents logiciels pour estimer les différentes composantes du bilan énergétique.

Un transfert est enfin réalisé sur un serveur de stockage de l'ENPC pour sauvegarder l'ensemble des données mesurées et calculées.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données des thermocouples, du radiomètre et du scintillomètre sont collectées grâce au programme « CR3000_LAS_CNR4_T.CR3 » installé sur la station d'acquisition. Ces données sont respectivement enregistrées pour chaque campagne sur deux fichiers « 2019_09_30_CR3000_cnr4.dat » et « 2019_09_30_CR3000_LAS.dat ». Les données suivantes sont enregistrées :

Radiomètre CNR4 :

- Rayonnement incident d'onde courte (W/m²)
- Rayonnement sortant/réfléchi d'onde courte (W/m²)
- Rayonnement net d'onde courte (W/m²)
- Rayonnement incident d'onde longue (W/m²)
- Rayonnement sortant/réfléchi d'onde longue (W/m²)
- Rayonnement net d'onde longue (W/m²)
- Température de l'air (°C)
- Température de l'air (°K)
- Albedo
- Rayonnement net (W/m²)

Thermocouples :

- Température du sol (°C) x 4

Scintillomètre LAS

- Logarithme du paramètre de structure de l'indice de réfraction de l'air (V) - Ucn2
- Ecart-type du logarithme du paramètre de structure de l'indice de réfraction de l'air (V)
- Force du signal or intensité (mV)
- Ecart-type de la force du signal or intensité (mV)
- Paramètre d'échelle de l'indice de réfraction de l'air (m-2/3)
- Ecart-type du paramètre d'échelle de l'indice de réfraction de l'air (m-2/3)

Ces 2 fichiers sont ensuite traités par un programme développé en python « Scintillometer.py », qui a l'aide des données

météorologiques de METEO-France (ex : synop.201809.txt) créé le fichier d'entrée nécessaire au logiciel EVATION. Ce fichier contient l'indice de réfraction de l'air (Cn2), la force du signal, les paramètres météorologiques (vent, température, pression...), le flux de conduction (Qg) et rayonnement net avec une fréquence de 10 minutes. Ce fichier prend la forme suivante : 2019_09_30_LAS_EVATION.txt.

Le logiciel EVATION, fourni avec le scintillomètre, calcule les termes manquants du bilan énergétique (flux de chaleur sensible (Qh) et latente (Qe)). Il synthétise tous les résultats (l'indice de réfraction de l'air, paramètres météorologiques, bilan énergétique) dans un fichier texte (2019_09_30_Output_EnergyBalance.txt) avec une fréquence de 10 minutes.

Les variables les plus significatives du fichier de sortie d'EVATION sont décrites ci-dessous :

Température de l'air (°C), humidité absolue (%), pression de (hPa), vitesse du vent (m/s), direction du vent (°), rayonnement net (W/m²), Indice de réfraction de l'air (Cn2- (m-2/3)), Force de la signal, intensité (Udemod- (mV)), flux de conduction dans le sol (Gsoil-(W/m²)), flux de chaleur sensible (H- (W/m²)), Flux de chaleur latente (LvE- (W/m²)) et le rapport de Bowen.

Campagne ETP : Données d'évapotranspiration mesurées à l'aide d'une chambre à transpiration sur la Vague Verte de Champs-sur-Marne

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

La mesure d'évapotranspiration (ETP) représente la quantité d'eau évaporée par le complexe sol/végétation, ainsi que celle transpirée par les plantes.

Ce sont des données ponctuelles réalisées pendant une campagne de mesures de quelques heures (6-7h) par temps sec. Elles proviennent de mesure par chambre de transpiration d'un m² que l'on pose pendant une durée de 2 minutes sur une couverture végétale (l'emplacement peut changer suivant la campagne). Elles comprennent des mesures d'humidité absolue (LI-COR LI-7500 en 2018, puis LI-COR LI-850 à partir de 2019), de température de l'air (thermocouple Type-K) et de rayonnement net (CNR4). Les données sont collectées à l'aide d'une station d'acquisition CR800 connectée à un ordinateur, afin de vérifier instantanément les mesures par l'utilisateur. Elles sont ensuite traitées à l'aide d'un logiciel adapté pour estimer la chronique d'évapotranspiration correspondante.

Un transfert est enfin réalisé sur un serveur de stockage de l'ENPC pour sauvegarder l'ensemble des données mesurées et calculées.

La chambre à transpiration et les différents capteurs qu'elle comprend appartiennent au Céréma Ile de France qui est partenaire du projet EVNATURB.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données enregistrées sont détaillées ci-dessous. Notons que 2 appareils différents ont été utilisés pour effectuer ces mesures : un LI-COR LI-7500 en 2018 et un LI-COR LI-850 à partir de 2019.

En 2018 (LI-COR LI-7500) :

- Humidité absolue (g/m³),
- Dioxyde de carbone (mg/m³),
- Températures de l'air à l'intérieur (°C) et à l'extérieur (°C) de la chambre
- Radiation nette (W/m²)
- Pression (kPa)

Depuis 2019 (LI-COR LI-850) :

- Humidité absolue (g/m³)
- Humidité absolue (mmol/mol)
- Dioxyde de carbone (µg/m³)
- Températures de l'air à l'intérieur (°C) et à l'extérieur (°C) de la chambre
- Radiation nette (W/m²)

Deux fichiers sont créés par campagne :

- Le premier fichier (ex : 2019_08_08_ChamberData.dat) contient les données enregistrées par les appareils de mesure à raison d'une valeur toutes les secondes.
- Le second fichier (ex : 2019_08_08_Vague_poseChambre.txt) contient les dates des pas de temps de pose et dépose de

la chambre sur la couverture végétale.

Un programme de lecture interactive (visualisation de la courbe d'humidité absolue pour chaque pose/dépose de la chambre) et de traitement des données collectées a été développé en python. Ce programme a deux versions, une pour les données collectées avec le LI-COR LI-7500 (2018) et une autre pour le LI-COR LI-850 (2019). Ils portent les noms suivants : « Lecturevague_chambreEvap_2018.py » (2018) et « Lecturevague_chambreEvap_2019.py » (2019). Chaque programme lit les données issues des 2 fichiers précédents, vérifie la bonne correspondance entre les pas de temps de pose et dépose de la chambre avec l'accroissement de la mesure d'humidité absolue, et détermine pour chacun d'eux une valeur d'évapotranspiration. Les données d'évapotranspiration sont synthétisées pour chaque campagne dans un fichier (ex : 2019_08_08_EvapotranspirationData.txt) qui contient également les valeurs de température intérieure et extérieure à la chambre, ainsi que le rayonnement net mesuré à l'intérieur de celle-ci pour chaque pas de temps de mesure.

2. Documentation et qualité des données

Bilan hydrique BGW : Données du bilan hydrique réalisé sur la Vague Verte de Champs-sur-Marne

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

Les jeux de données du bilan hydriques comprenant 4 fichiers (pluie, teneur en eau, débit dans la conduite, débit dans l'unité de stockage) sont regroupés dans un dossier par période d'un mois sur le serveur de stockage de l'ENPC. Ce dossier comprend un document (2020_0101_0201_Bilan_hydrique.png) présentant succinctement cette période de mesure à l'aide de graphiques (représentation des chroniques) et de valeurs numériques (cumul de pluie, coefficient de ruissellement, variation de la rétention). Les entêtes des 4 fichiers de données rappellent la mesure réalisée, l'unité, la période d'étude.

Le descriptif précis de ces mesures est à retrouver dans le data paper suivant :

Versini, P.-A., Stanic, F., Gires, A., Schertzer, D., Tchinguirinskaia, I.: Data for "Measurement of the water balance components of a large green roof in Greater Paris Area", <https://doi.org/10.5281/zenodo.3467300>, 2019

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Lors de la constitution du dossier mensuel, une analyse est effectuée pour s'assurer que le nombre de données manquantes est minime. Le bilan hydrique est aussi évalué pour s'assurer que le volume précipité est bien inférieur à la somme des volumes ruisselés (déterminé à l'aide du débit) et stocké dans le substrat (déterminé à l'aide de la teneur en eau). Une représentation graphique de l'évolution dans le temps de ces différentes données est aussi réalisée pour s'assurer de la pertinence de ces mesures. Le code « NAN » est synonyme d'absence de mesure dans les différents fichiers.

Bilan thermique BGW : Données du bilan énergétique réalisé sur la Vague Verte de Champs-sur-Marne

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

Pour chaque campagne de mesure (ex : 2019_09_30_EnergyBalance_report.txt pour la campagne du 30 septembre 2019), un fichier décrit brièvement le protocole suivi (emplacement, durée, problème rencontré).

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Les données du scintillomètre sont soumises à une procédure de contrôle de qualité très rigoureuse. Elles doivent satisfaire les conditions suivantes :

- Intensité du signal du scintillomètre (Demod) < -50 mV (réalisé par le logiciel EVATION)
- Logarithme du paramètre de structure de l'indice de réfraction de l'air (V) - UCn2 compris entre -5 to 0 V (réalisé par le programme python Scintillometer.py)
- Indice de réfraction de l'air- Cn2 compris entre 10-17 to 10-11 m-2/3 (réalisé par le programme python Scintillometer.py et le logiciel EVATION)
- Conditions atmosphériques (instables) propices à la mesure : les données mesurées dans des zones saturées, caractérisées par une faible turbulence, ou pendant des événements de précipitation sont refusés (réalisé par le programme python Scintillometer.py)
- Vérification de la colonne « Status » sur le fichier de sortie d'EVATION pour s'assurer d'éventuels problèmes de données et les conditions atmosphériques de mesure.

Les données de température du sol sont comparées avec les mesures effectuées par les capteurs de teneur en eau TDR installés sur la Vague Verte.

Enfin, les données du fichier de sortie d'EVATION sont analysées à l'aide du même programme développé pour la lecture des données brutes « Scintillometer.py ». A cette fin, les paramètres d'entrée sont modifiés afin de réaliser cette analyse. Ainsi configuré, ce programme permet de :

- représenter graphiquement les différents composants du bilan énergétique
- comparer les variations de l'indice de réfraction air vis à vis de ce bilan énergétique

Campagne ETP : Données d'évapotranspiration mesurées à l'aide d'une chambre à transpiration sur la Vague Verte de Champs-sur-Marne

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

Pour chaque campagne de mesure, un fichier (ex : 2019_09_30_Evapotranspiration_report.txt) décrit brièvement le protocole suivi (emplacement, durée, problèmes rencontrés).

Les entêtes des fichiers rappellent la mesure réalisée, l'unité, la période d'étude.

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Le contrôle des données s'effectue manuellement par l'utilisateur à l'aide du programme python (Lecturevague_chambreEvap_2018.py ou Lecturevague_chambreEvap_2019.py) qui permet notamment de vérifier que les pas de temps de pose et dépose de la chambre coïncide bien avec le début d'augmentation de l'humidité absolue (visualisation de la courbe d'humidité absolue).

3. Stockage et sauvegarde pendant le processus de recherche

Bilan hydrique BGW : Données du bilan hydrique réalisé sur la Vague Verte de Champs-sur-Marne

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Différents protocoles de stockage sont mis en œuvre suivant le type de données :

- Précipitation, teneur en eau et débit dans la conduite : les données sont collectées automatiquement toutes les nuits sur le serveur de stockage de l'ENPC
- Débit dans l'unité de stockage : les données sont collectées manuellement tous les mois. Cette différence est due à un éloignement trop important entre l'unité de stockage et la connexion internet.

Les données sont ensuite regroupées mensuellement dans un même dossier (AAAA_MMJJ_MMJJ) où l'on retrouve les fichiers de données et celui présentant la période de mesure. Exemple pour le mois de Janvier 2020 :

- 2020_0101_0201_Bilan_hydrique.png
- 2020_0101_0201_Data_Q_storage.csv (débit dans l'unité de stockage)
- 2020_0101_0201_Data_Q-conduit.csv (débit dans la conduite)
- 2020_0101_0201_Data_rainfall.csv
- 2020_0101_0201_Data_vwc.csv

Les données sont stockées sur les serveurs de stockage du service informatique de l'ENPC qui disposent des outils standards et mis à jour en terme de récupération de données. Le service informatique ENPC assure la sécurisation technique des données et le processus de sauvegarde. Ce dernier dispose notamment d'outil de réplication sur deux sites différents.

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

Les données collectées ici ne sont pas particulièrement sensibles.

L'accès aux répertoires contenant ces données est contrôlé. Une demande est faite à la DSI pour toute nouvelle personne interne à l'ENPC qui souhaiterait y avoir accès. Pour le moment les personnes extérieures à l'ENPC n'ont pas accès à ces données.

Pour les recherches collaboratives utilisant ces données, une convention de collaboration de recherche est signée entre les différentes parties. Cela concerne la mise à disposition des données produites par l'ENPC à un partenaire, la mise à disposition des données produites par un partenaire à l'ENPC et la production scientifique qui en découle.

Bilan thermique BGW : Données du bilan énergétique réalisé sur la Vague Verte de Champs-sur-Marne

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Les données sont stockées par campagne de mesure dans un répertoire dédié sur le serveur de stockage de l'ENPC. Chaque dossier comprend le fichier décrivant le jour de mesure, ainsi que les données brutes et calculées à l'aide du logiciel EVATION (ex du 30 septembre 2019):

- 2019_09_30_EnergyBalance_report.txt
- 2019_09_30_CR3000_cnr4.dat
- 2019_09_30_CR3000_LAS.dat
- 2019_09_30_LAS_EVATION.txt
- 2019_09_30_Output_EnergyBalance.txt
- synop.201909.txt

Un programme Python de lecture de ces données est aussi mis à disposition (Scintillometer.py)

Les données sont stockées sur les serveurs de stockage du service informatique de l'ENPC qui disposent des outils standards et mis à jour en terme de récupération de données. Le service informatique ENPC assure la sécurisation technique des données et le processus de sauvegarde. Ce dernier dispose notamment d'outil de réplication sur deux sites différents.

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

Les données collectées ici ne sont pas particulièrement sensibles.

L'accès aux répertoires contenant ces données est contrôlé. Une demande est faite à la DSI pour toute nouvelle personne interne à l'ENPC qui souhaiterait y avoir accès. Pour le moment les personnes extérieures à l'ENPC n'ont pas accès à ces données.

Pour les recherches collaboratives utilisant ces données, une convention de collaboration de recherche est signée entre les différentes parties. Cela concerne la mise à disposition des données produites par l'ENPC à un partenaire, la mise à disposition des données produites par un partenaire à l'ENPC et la production scientifique qui en découle.

Campagne ETP : Données d'évapotranspiration mesurées à l'aide d'une chambre à transpiration sur la Vague Verte de Champs-sur-Marne

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Les données sont stockées par campagne dans le répertoire dédié sur le serveur de stockage de l'ENPC (ex : 20180927). Chaque dossier comprend le fichier décrivant le jour de mesure, ainsi que les données brutes et calculées (ex du 30 septembre 2019):

- 2019_09_30_Evapotranspiration_report.txt
- 2019_08_08_ChamberData.dat
- 2019_08_08_Vague_poseChambre.txt
- 2019_08_08_EvapotranspirationData.txt

Un programme Python de lecture de ces données est aussi mis à disposition (Lecturevague_chambreEvap_2018.py et Lecturevague_chambreEvap_2019.py)

Les données sont stockées sur les serveurs de stockage du service informatique de l'ENPC qui disposent des outils standards et mis à jour en terme de récupération de données. Le service informatique ENPC assure la sécurisation technique des données et le processus de sauvegarde. Ce dernier dispose notamment d'outil de réplication sur deux sites différents.

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

Les données collectées ici ne sont pas particulièrement sensibles.

L'accès aux répertoires contenant ces données est contrôlé. Une demande est faite à la DSI pour toute nouvelle personne interne à l'ENPC qui souhaiterait y avoir accès. Pour le moment les personnes extérieures à l'ENPC n'ont pas accès à ces données.

Pour les recherches collaboratives utilisant ces données, une convention de collaboration de recherche est signée entre les différentes parties. Cela concerne la mise à disposition des données produites par l'ENPC à un partenaire, la mise à disposition des données produites par un partenaire à l'ENPC et la production scientifique qui en découle.

4. Exigences légales et éthiques, codes de conduite

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Pas concerné. Sans objet.

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Les données collectées au cours du projet EVNATURB sont la propriété de l'ENPC. L'ENPC a une politique proactive en matière de partage des données. Pour se faire, une convention de partage de données est signée par les établissements intéressés d'utiliser ces données. Une convention peut aussi être établie avec un autre établissement fournisseur de données si l'ENPC souhaite les utiliser.

Habituellement, le fournisseur de données reste l'unique propriétaire de celles-ci. Les personnes qui les utilisent ont l'obligation de : (i) ne les communiquer sous aucune forme à des tiers sauf accord préalable et écrit entre les Parties, (ii) transmettre une demande d'accord préalable et écrit au propriétaire pour toute publication ou communication incorporant les données, (iii) faire mention explicite des Parties et des laboratoires propriétaires des données dans toute publication ou communication les incorporant.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Sans objet.

5. Partage des données et conservation à long terme

5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

Un délai d'un an est prévu entre la collecte/analyse des données et leur mise à disposition. Cette durée doit permettre au laboratoire d'avoir le temps de les valoriser en termes de productions scientifiques (publications, modèles...).

Pour les données relatives aux bilans hydrique et énergétique, ainsi que celles issues des campagnes d'évapotranspiration, un tableau récapitulatif des périodes de disponibilité sera mis en ligne via le site du laboratoire (pages dédiées au projet EVNATURB et de la Vague Verte).

La demande de mise à disposition se fera par cette même interface.

Cette mise à disposition des données se fera dans le cadre des conventions de collaboration déjà mentionnées.

Il est aussi envisagé de rédiger des data paper avec dépôt de données sur un entrepôt de confiance chaque fois que cela sera possible. C'est d'ores et déjà le cas pour les mesures de bilan hydrique :

Versini, P.-A., Stanic, F., Gires, A., Schertzer, D., Tchinguirinskaia, I.: Data for "Measurement of the water balance components of a large green roof in Greater Paris Area", Earth Syst. Sci. Data, <https://doi.org/10.5281/zenodo.3467300>, 2019

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

Les données seront conservées sur le long terme sur les serveurs de l'ENPC.

5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

Concernant le bilan hydrique, les données sont toutes au format .csv, ce qui permet une utilisation par tout un chacun sans

requérir d'outils ou logiciels particuliers. Un programme en Python développé au sein du laboratoire HM&Co est aussi mis à disposition pour lire et visualiser ces données (Read_Data_BGW.py).

Concernant le bilan énergétique, un programme Python « Scintillometer.py » (développé au sein du laboratoire HM&Co) est mis à disposition pour : (i) représenter graphiquement les différents composants du bilan énergétique, (ii) comparer les variations de l'indice de réfraction air vis à vis de ce bilan énergétique.

Concernant les campagnes d'évapotranspiration, un programme Python de lecture de ces données (développé au sein du laboratoire HM&Co) est mis à disposition (Lecturevague_chambreEvap_2018.py et Lecturevague_chambreEvap_2019.py) pour vérifier que les pas de temps de pose et dépose de la chambre coïncide bien avec le début d'augmentation de l'humidité absolue.

5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

Un DOI sera attribué à chaque jeu de données ayant fait l'objet d'une publication de type data paper. C'est d'ors et déjà le cas pour 3 mois de mesure de bilan hydrique : <https://doi.org/10.5281/zenodo.3687775>

6. Responsabilités et ressources en matière de gestion des données

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

Le coordinateur du projet ANR EVNATURB, Pierre-Antoine Versini, chargé de recherche au sein du laboratoire Hydrologie, Météorologie et Complexité de l'Ecole des Ponts ParisTech, sera en charge de la gestion de ces données (collecte, analyse et stockage mensuels, puis mise à disposition). Il est prévu une demi-journée mensuelle pour cette opération de collecte et d'analyse.

C'est lui qui sera le point d'entrée pour les demandes de mise à disposition et la mise en oeuvre de collaborations de recherche avec d'autres laboratoires. Ces collaborations seront ensuite validées au niveau de la direction du laboratoire.

Il sera épaulé pour les aspects techniques par un responsable informatique qui sera en charge de la gestion des données produites par le laboratoire (dont les données radar). Ce recrutement est prévu par le laboratoire, mais sera nécessairement financé sur fond propre.

Concernant les aspects juridiques relatifs aux conventions de partenariat, le service idoine de la Direction de la Recherche sera sollicité.

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

La gestion des données fait partie intégrante du rôle de porteur de projet pour lequel Pierre-Antoine Versini consacra 50% de son temps.

Le Pôle IST de la Direction de la Doc est partie prenante du plan Science Ouverte de l'École et constitue un support auprès des chercheurs tout au long du projet, tant sur le partage des publications que sur la mise à disposition des données. Il peut donner notamment des pistes sur le choix des entrepôts et les contraintes juridiques.