

Plan de gestion des données du Service National d'Observation des Tourbières

Jean-Baptiste Paroissien, Sébastien Gogo, Fatima Laggoun

10 avril 2018

1 Introduction

Le Service National d'Observation des Tourbières (**SNO Tourbières**) a acquis, acquière et va acquérir un nombre important et varié de données provenant à la fois de site et d'équipes de recherches différentes. Son objectif est d'observer sur le long terme les flux de gaz à effet de serre et ses variables explicatives pour comprendre les effets du changement climatique sur le fonctionnement des tourbières et les modéliser. La pérennité, l'accessibilité, la compréhension et l'interopérabilité des données avec d'autres système d'information sont des notions qui doivent être respectées tout au long du cycle de vie des données du **SNO Tourbières**. Pour parvenir à cet objectif, un ensemble de documents « métiers » accompagne la vie des données du **SNO Tourbières** durant ces principales étapes :

Acquisition L'acquisition des données correspond à l'étape

Traitement et modélisation Le traitement et la modélisation des données du **SNO Tourbières**

Diffusion La diffusion des données représente l'étape ...

Maintenance et mise à jour Ici, il faudra prendre en compte l'évolution du SI **SNO Tourbières** à travers un ensemble de mode opératoire pour maintenir le SI à jour et assurer son évolution (ajout d'un nouveau site, nouveaux instruments, nouveaux types de variables...)

Ce document est un plan de gestion de données qui regroupe la démarche globale de la gestion des données du **SNO Tourbières**. Il est accompagné de lien vers des documents techniques et métiers qui détaillent la démarche de gestion des données.

Le trame du document s'appuie la traduction française du gestion de données de la commission européenne consultable à cette [adresse](#).

2 Résumé descriptif des données collectées dans le SNO Tourbières

2.1 Objectif de la collecte des données

Les données collectées dans le **SNO Tourbières** ont pour objectif de répondre aux problématiques associées aux changements climatiques et aux répercussions possibles dans les tourbières.

Estimer le bilan de carbone et hydrologique dans les tourbières : Les sites du **SNO Tourbières** sont équipés d'instruments de mesures qui collectent à un pas de temps de 30 minutes les émissions de Gaz à Effet de Serre (GES), la concentrations en carbone organique dissous à l'exutoire, les variables hydrologiques et des données environnementales permettant l'établissement à l'échelle de l'écosystème du bilan de carbone et d'eau.

Mettre en évidence les rétroactions entre le changement climatique et les GES des tourbières :

En étant représentatif de différents contextes climatiques, les données collectées du **SNO Tourbières** permettront de modéliser les émissions de GES et d'alimenter et affiner les modèles de changement climatiques globaux. En évaluant les effets du changement du climat sur les tourbières et les effets de ces changements sur le climat, le **SNO Tourbières** se propose de déterminer quelles rétroactions vont se mettre en place,

Développer des expérimentations : Le **SNO Tourbières** a pour vocation à accueillir des dispositifs expérimentaux portés sur l'étude des forçages climatiques et anthropiques, qu'ils soient *in situ* ou au laboratoire. Ces expérimentations sont réalisées dans le cadre de projet.

2.2 Description des données collectées

2.2.1 Origine et volume des données collectées

Les données collectées du **SNO Tourbières** proviennent d'un ensemble de 4 sites de tourbières instrumentés en France dans des contextes climatiques variés :

- Bernadouze (Ariège),
- Frasne (Doux),
- La Guette (Cher),
- Landemarais (Ille-et-Vilaine).

Le niveau d'équipement des sites est hétérogène mais un ensemble de données communes aux sites est collecté. Un maximum théorique de 100 stations est prévue, mais concrètement, le nombre de stations d'acquisition par site devrait se situer entre 10 et 20 par site. En l'état actuel des équipements, le volume des données collectées est estimé à environ XX/an et devrait être de XX/an lorsque l'ensemble des sites seront complètement équipés.

2.2.2 Types et formats des données

Les données collectées sur les sites du **SNO Tourbières** sont regroupés en 4 types :

Biogéo : Ensemble des données biogéochimique et physique des eaux à l'exutoire des sites,

Bioveg : Données sur la biodiversité,

GES/Météo-sol : Ensemble des données provenant des tours à flux de GES ainsi que des stations collectant des données sur la météorologie de l'atmosphère et du sol,

Hydro/Carto : Données sur l'hydrologie et l'hydrogéologie.

Les données collectées sont au format seront stockées au format `csv` (coma separated values) avec les caractéristiques suivantes :

- séparateur de colonne : virgule,
- séparateur décimal : point,
- encodage : UTF-8.

2.2.3 Documents

Voir si rajouter un lien vers le protocole du SNO Tourbières

2.3 La réutilisation des données

3 Gestion des données selon le principe « FAIR »

Cette section aborde le point de la gestion des données selon le principe « FAIR », c'est à dire selon le principe que les données doivent être trouvables, accessibles, interopérables et réutilisables. Pour plus d'information sur ce concept, le lecteur intéressé peut consulter cet [article](#).

3.1 Rendre les données et les métadonnées trouvables

Les données et les métadonnées relatives au SNO Tourbières seront organisées dans une base de données relationnelle connectée à un moteur de génération de service web de type Open Geospatial Consortium (OGC). A terme, ce système d'information (SI) permettra :

- de générer des fiches de métadonnées et de les diffuser à travers des services web interopérables OGC de type Catalog Service for the Web (CSW).
- de diffuser les données des stations d'acquisition avec des services web interopérables OGC de type Sensor Observation Service (SOS).

3.1.1 Les métadonnées du SNO Tourbières

Le SI du SNO Tourbières permettra de générer différents type de métadonnées pour les jeux de données du SNO Tourbières .

Format OZCAR-Pivot L'infrastructure de recherche OZCAR auquel le SNO Tourbières appartient diffuse les métadonnées de ses membres sous la base d'un standard dit hybride appelé « OZACR-Pivot ».

Format INSPIRE Pour répondre aux exigences de la directive INSPIRE, des métadonnées au format INSPIRE seront générées et moissonnées sur un ensemble de plateforme pas encore définies.

Format SensorML Chaque jeux de données du SNO Tourbières est associé à des instruments qui seront également documentés au travers de métadonnées. Ces métadonnées respecteront le format SensorML de l'OGC.

3.1.2 Procédure d'identification

Pour le moment, la procédure d'identification des jeux de données du SNO Tourbières n'est pas encore définie. *A priori*, elle s'appuiera sur un identifiant pérenne unique de type DOI. Pour mettre en place ces identifiants, le SNO Tourbières s'appuiera sur les services de l'INIST-CNRS.

3.1.3 Diffusion des métadonnées

Plusieurs services (à identifier) moissonneront ces métadonnées pour faciliter la découverte des jeux de données proposés par le SNO Tourbières . Le site web du SNO Tourbières (en construction) proposera également un lien vers les métadonnées.

3.2 Rendre les données librement accessibles

Toutes les données du SNO Tourbières seront librement disponibles

3.2.1 Méthode de diffusion

La mise à disposition des données du SNO Tourbières sera réalisée à travers le développement de services web interopérables OGC de type SOS.

Site web du SNO Tourbières : le site web du SNO Tourbières donne un accès à une interface web proposant le libre téléchargement des données à travers un formulaire de requête.

Service web interopérable OGC (SOS) : Un moteur de génération de flux de données interopérable de type (Sensor Observation Service) permettra à n'importe quel utilisateur d'accéder aux données du SNO Tourbières .

3.2.2 Accès aux codes et documentation

L'ensemble du code de développement du système d'information du SNO Tourbières est stocké à un serveur de versionnement `git` connecté à une forge redmine. L'accès à la forge et au code est pour le moment restreint. Les demandes seront étudiées selon une procédure qui reste à définir. Une documentation en ligne basée sur [mkdocs](#) sera mise en place et accessible sur le site web du SNO Tourbières pour guider l'utilisateur dans l'extraction et l'utilisation des données.

3.3 Rendre les données interopérables

L'interopérabilité du SI du SNO Tourbières est recherché sur le niveau de la sémantique et sur le niveau technique. La recherche de l'interopérabilité du SNO Tourbières est un point de vigilance qui s'appuie notamment sur les recommandations du [Référentiel Général d'Interopérabilité \(RGI\)](#)

3.3.1 Démarche d'interopérabilité au niveau sémantique

Le vocabulaire employé pour décrire les données du SNO Tourbières s'appuie sur plusieurs sources :

- Le code des variables du SNO Tourbières respecte les recommandations de la discipline et notamment du réseau [ICOS](#)
- Un tableau de correspondance sera mis en place entre le taxonomie [Global Change Master Directory](#) de la NASA et les variables du SNO Tourbières .

D'autre part, l'organisation et la description des données sera consultable sur le site web du SNO Tourbières à travers la génération d'une documentation basée sur [schemaspy](#).

3.3.2 Démarche d'interopérabilité au niveau technique

Au niveau technique, la démarche d'interopérabilité s'applique sur les standards et les protocoles d'échanges des données et de leur métadonnées ainsi que sur la syntaxe utilisée.

Données (brutes, élaborées) Des métadonnées (données, jeux de données et capteurs) Encodage du texte (utf-8) Type de document (docx, odf, pdf) Structuration des données (XML)

L'ensemble des variables du SNO Tourbières s'appuie sur les recommandations du réseau européen d'acquisition de données de flux de gaz à effet de serre ICOS. D'autres part, deux types de métadonnées seront générées :

Les métadonnées aux formats INSPIRE

Les métadonnées dans un format spécifique. Ce format est proposé par l'infrastructure de recherche OZCAR. Dans ce cas, le vocabulaire employé s'appuie sur le thesaurus Global Change Master Directory de la NASA (<https://gcmd.nasa.gov/>).

3.4 Accroître la réutilisation des données

Les données seront accessibles pour être ré-utiliser dès lors qu'elles seront diffusées. Il n'y aura pas d'embargo sur les données, la diffusion des données sera réalisé après un processus de vérification et de traitement automatisé. Pour le moment, aucune licence n'a été identifié, ce travail reste à faire.

Développer la partie ré-utilisation...!

4 Allocation des ressources

5 Sécurité des données

6 Aspects éthiques

7 Autres