

Micro Projet VALDOM 2021-2022

Eaux de surfaces

Data

- Pas une vérité absolue, (<https://www.tylervigen.com/spurious-correlations>) des bonnes pratiques bien sûr, mais l'apport du métier est primordial tout au long de la mise en place d'un projet data driven :
 - └ Dans la phase de définition/cadrage d'un projet :
 - Quelles entrées, que représentent elles, quelles informations métiers sont elles susceptibles de porter
 - Avec quelle qualité (questions sur la régularité des collections, la complétude des données, le traitement des manques...)
 - Quel « domaine » de couverture : comment traiter des aspect catégoriel 1 seul modèle ou plusieurs
 - Quelles sortie attendues : quelle grandeur sera exploitable du point de vue du métier avec quelle fréquence de rafraichissement
 - Avec quelles performances possibles ou attendues :
 - Comment évaluer/ caractériser ces performances
 - └ Dans la phase de mise en œuvre :
 - Support à la formalisation de la problématique métier en approche data driven
 - Généralement cette phase est assez itérative avec des proposition et des ajustements
 - └ Dans la phase de qualification des résultats
 - Est-ce que le formalisme des sorties correspond aux attentes du métiers
 - Analyse des performances :
 - formaliser les résultats observé d'un point de vue métiers : interprétabilité (compréhension des éléments impactant)
 - Adéquation avec les attentes et incidence de mauvaise détection s (FP/FN), avec **quelles conséquences**.
 - └ Dans la phase d'industrialisation :
 - Quelles surveillances à considérer dans la mise en œuvre (déploiement opérationnel) d'une telle approche

Organisation & Déroulement du micro-projet

— Objectifs :

- └ Manipulation de données et appréhension des aspects métiers à considérer lorsque l'on adresse un projet data driven

— Découpage en 4 session :

- └ Session 1 : contexte, objectifs et prise en main des données
- └ Session 2 : ébauche (consolidation?) des uses cases : revue « collaborative » des cas d'usage, critiques constructive & support à consolidation
- └ Session 3 : Support aux problématiques rencontrées, aide à formaliser le problème
- └ Session 4 : présentations de la démarche, de l'avancements des points durs rencontrés

Peu d'écart entre session 1 et 2.

La surveillance de l'eau en france

- La qualité des eaux de surface (rivières, lacs, etc.) est réglementée et contrôlée par les services publics ou par des prestataires privés (bureaux d'étude) qui sont mandatés pour effectuer ces prélèvements.
- Surveillance de la qualité des eaux de surface :
 - └ Dans le droit français, la loi sur l'eau et les milieux aquatiques (Lema) (transcrite de la directive cadre européenne sur l'eau ([DCE](#))) impose une série de normes et d'objectifs en termes de qualité et de quantité de la ressource en eau.
 - └ Des programmes de [surveillance des eaux](#) de surface sont donc en action pour contrôler régulièrement différents paramètres tels que le [débit](#), l'état chimique (présence de [molécules](#), concentrations), l'état écologique (composition, abondance et [biomasse](#) de divers organismes), la [morphologie](#), etc.
 - └ Les contrôles sont effectués dans chaque grand [bassin versant](#) par l'Office national de l'eau et des milieux aquatiques ([Onema](#)) avec l'aide des [agences de l'eau](#) et de leurs délégations territoriales.
- L'ensemble des données sont disponibles sur le [portail eaufrance](#) du Système d'information sur l'eau (SIE). <https://www.eaufrance.fr/>

Données sur l'eau

- <https://www.eaufrance.fr/les-donnees-des-sites-eaufrance>
- **Eaux de surface :**
 - **Naïades** est l'interface nationale officielle permettant d'accéder aux données de qualité de l'eau des rivières et des lacs. <http://www.naiades.eaufrance.fr/>
- **Eaux souterraines**
 - **ADES** est le portail national d'**Accès aux Données sur les Eaux Souterraines** pour la France métropolitaine et les départements d'outre-mer : données quantitative et qualitative.
<https://ades.eaufrance.fr/Spip?p=presentation>
- Le Sandre a pour mission, d'établir et de mettre à disposition le référentiel * des données sur l'eau du SIE (*dispositif créé par l'État pour le partage et la mise à disposition des données sur l'eau du secteur public.*) . Ce référentiel *, composé de spécifications techniques et de listes de codes libres d'utilisation, décrit les modalités d'échange des données sur l'eau à l'échelle de la France. D'un point de vue informatique, le Sandre garantit l'interopérabilité des systèmes d'information * relatifs à l'eau.
- La traçabilité des données du producteur à l'utilisateur garantit un bon niveau d'informations pour une meilleure utilisation et interprétation des données consultées, conformément aux préconisations du Service d'Administration Nationale des Données et Référentiels sur l'Eau (SANDRE) : origine des données, indication du niveau de validité des données...

NAIADES : Données sur la qualité de l'eau

- <http://naiades.eaufrance.fr/bienvenue-naiades>



Données sur la qualité des eaux de surface



Espace perso

Vous êtes ici : Accueil

 Présentation

 À propos des données

 Boîte à outils

 Accès aux données

 Bienvenue sur NAIades

Accédez simplement aux relevés d'observation sur la qualité des cours d'eau et des plans d'eau...



Recherchez toutes les données sur la qualité des eaux de votre vallée

en effectuant une recherche par aire hydrographique

Actualités

12 janvier 2021
Recherche de données sur le gros paramètres Phytosanitaires

23 décembre 2020
Différence entre données en ligne pour le bassin Loire-Bretagne

26 novembre 2020
Recherche de paramètre PC : utiliser la préférence les codes Sandre

18 novembre 2020
Etat de la mise à jour des fichiers France entière

18 novembre 2020
Recherche par bassin dysfonctionnel

Toutes les actualités

Thématiques

— <http://naiades.eaufrance.fr/thematiques>



Données sur la qualité des eaux de surface



Espace persoContactGlossaire

Vous êtes ici : Accueil > À propos des données > Thématiques

Présentation

À propos des données

Origine des données

Thématiques

Organisation des données

Validation des données

Boîte à outils

Accès aux données

Thématiques

La surveillance de la qualité des eaux de surface couvre différentes natures de données appelées thématiques. Dans Naiades, ces thématiques sont représentées par des pictogrammes sélectionnables en haut du [formulaire de recherche de données](#).

Hydrobiologie

En hydrobiologie, quel que soit le **sout**port (poisson, invertébrés, diatomées etc) et la catégorie d'eau (cours d'eau, plan d'eau), l'acquisition des données biologiques consiste à déterminer le peuplement d'un milieu à un moment donné (pour un support donné), autrement dit à inventorier l'ensemble des espèces (**taxons**) présentes sur le terrain d'étude le jour de l'observation.

Ces peuplements sont consignés sous la forme de **listes** faunistiques ou floristiques. Pour chaque **taxon répertorié** la liste permet d'enregistrer :

- Un **taxon** : espèce vivante observée dans le milieu et décrite à un niveau particulier de la systématique (famille, genre, espèce etc.).
- Un **type de mesure** : selon les études, les inventaires pourront répertorier différentes grandeurs. Par exemple, l'étude des invertébrés consiste à identifier des taxons, puis leur *effectif* alors que l'étude des macrophytes consiste à identifier des taxons (algues, bryophytes, plantes vasculaires etc.) puis une estimation de leur *recouvrement*. Le type de mesure permettra de savoir si le résultat est un *effectif* ou un *recouvrement*.
- Un **résultat** : valeur du type de mesure pour le taxon d'intérêt.

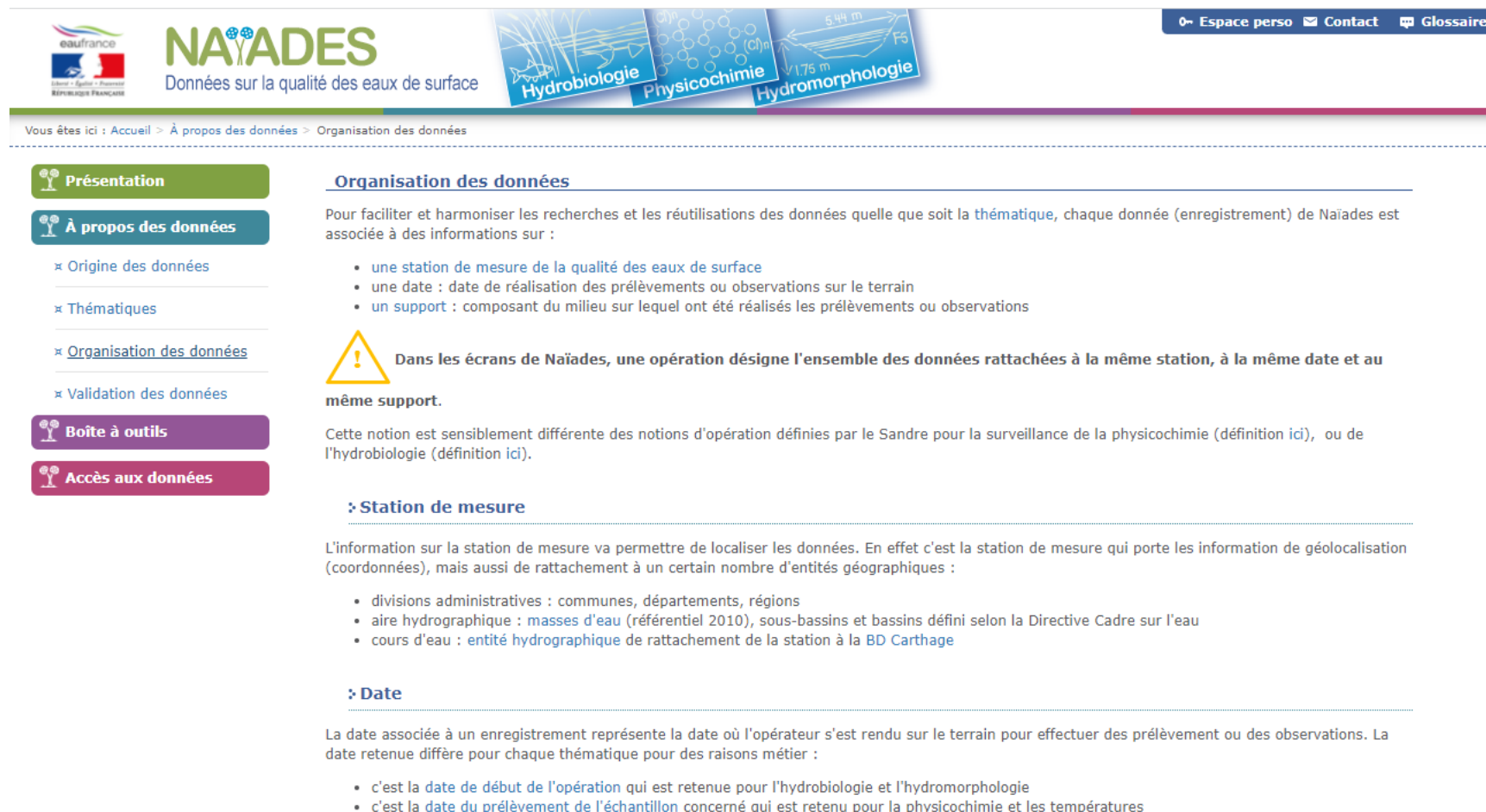
Les peuplements variant, l'expertise sur la valeur écologique d'un peuplement est spécifique à chaque typologie de cours d'eau (localisation, taille, régime hydrologique etc.). Pour pouvoir comparer différents peuplements dans l'espace et dans le temps, des indices (et métriques sous-jacentes) ont été définis. Les indices permettent d'associer à chaque peuplement une note traduisant l'état du cours d'eau. Les valeurs prises par ces indices sont enregistrées dans les **résultats biologiques**. Chaque résultat biologique est décrit par :

- Un **paramètre** : paramètre hydrobiologique correspondant à l'indice ou la métrique étudiée. La liste des paramètres est issue du [référentiel des paramètres](#).
- Un **résultat** : valeur calculée de l'indice ou de la métrique.
- Une **unité de mesure** : unité d'expression du résultat. La liste des unités est issue du [référentiel des unités de référence](#).
- Un **code remarque** : code permettant d'apporter des précisions sur le résultat.

Hydromorphologie

Organisation des données

— <http://naiades.eaufrance.fr/organisation-donnees>



NAIADES
Données sur la qualité des eaux de surface

Vous êtes ici : Accueil > À propos des données > Organisation des données

Présentation

À propos des données

- ✕ Origine des données
- ✕ Thématiques
- ✕ Organisation des données
- ✕ Validation des données


Boîte à outils

Accès aux données

Organisation des données

Pour faciliter et harmoniser les recherches et les réutilisations des données quelle que soit la **thématique**, chaque donnée (enregistrement) de NAIADES est associée à des informations sur :

- une station de mesure de la qualité des eaux de surface
- une date : date de réalisation des prélèvements ou observations sur le terrain
- un support : composant du milieu sur lequel ont été réalisés les prélèvements ou observations

 **Dans les écrans de NAIADES, une opération désigne l'ensemble des données rattachées à la même station, à la même date et au même support.**

Cette notion est sensiblement différente des notions d'opération définies par le Sandre pour la surveillance de la physicochimie (définition [ici](#)), ou de l'hydrobiologie (définition [ici](#)).

✧ Station de mesure

L'information sur la station de mesure va permettre de localiser les données. En effet c'est la station de mesure qui porte les informations de géolocalisation (coordonnées), mais aussi de rattachement à un certain nombre d'entités géographiques :

- divisions administratives : communes, départements, régions
- aire hydrographique : masses d'eau (référentiel 2010), sous-bassins et bassins défini selon la Directive Cadre sur l'eau
- cours d'eau : entité hydrographique de rattachement de la station à la [BD Carthage](#)

✧ Date

La date associée à un enregistrement représente la date où l'opérateur s'est rendu sur le terrain pour effectuer des prélèvements ou des observations. La date retenue diffère pour chaque thématique pour des raisons métier :

- c'est la **date de début de l'opération** qui est retenue pour l'hydrobiologie et l'hydromorphologie
- c'est la **date du prélèvement de l'échantillon** concerné qui est retenu pour la physicochimie et les températures

Données disponibles

— <http://naiades.eaufrance.fr/donnees-disponibles>



Données sur la qualité des eaux de surface



Espace persoContactGlossaire

Vous êtes ici : Accueil > Accès aux données > Données disponibles

Présentation

À propos des données

Boîte à outils

Accès aux données

Recherche

France Entière

Données disponibles

Données disponibles

L'organisation de la centralisation des données varie d'un acteur à un autre. Pour le portail Naiades, la fréquence d'actualisation minimale des données qui a été retenue est la suivante : à minima à l'automne de l'année n sont mises à disposition les données qualifiées de l'année n-1.

Le tableau suivant récapitule les données disponibles le 12/03/2021 :

Thématique	Zone géographique	Catégorie d'eau (*)	Date de la première observation	Date de la dernière observation	Nombre d'opérations	Nombre d'observations (**)	Date de mise à jour des données	Code Banque de référence
Physicochimie	Bassin Adour-Garonne	CE / PE	03/01/1971	31/12/2019	311 429	15 250 848	27/10/2020	PC AEAG
	Bassin Artois-Picardie	CE / PE	16/03/1960	12/02/2020	87 924	5 808 387	29/10/2020	PC AEAP
	Bassin Loire-Bretagne	CE / PE	02/01/1971	15/10/2020	369 233	39 476 342	18/11/2020	PC AELB
	Bassin Rhin-Meuse	CE / PE	01/02/1970	26/12/2019	181 902	17881188	21/04/2021	PC AERM
	Bassin Rhône-Méditerranée-Corse	CE / PE	14/01/1969	30/12/2020	219087	42 853 267	01/02/2021	PC AERMC
	Bassin Seine-Normandie	CE / PE	03/09/1970	27/12/2019	235 139	58 752 400	25/11/2020	PC AESN
	Bassin Guadeloupe	CE / PE	08/01/2014	20/12/2017	706	86 780	14/10/2019	PCOEGUA
	Bassin Guyane	CE / PE	17/06/2007	01/08/2020	1181	130 950	14/01/2021	PC OEGUY
	Bassin Martinique	CE / PE	29/07/1993	27/11/2019	4 123	673 549	26/10/2020	PC OEMAR

Données disponibles sur le bassin Adour-Garonne

Hydrobiologie : 10-2015 à aujourd'hui

— Bassin adour-garonne

- └ Description des données :

<https://drive.google.com/file/d/1GQtdWMoVsgghquLNrVzF8PkPe31TNLRDq/view?usp=sharing>

- └ Resultats Hydrobiologiques :

<https://drive.google.com/file/d/1A7IMg6kbgxDSka2nhd4HikJqdsGrgcA9/view?usp=sharing>

- └ Stations Hydrobiologiques :

· <https://drive.google.com/file/d/15czH1wtQheIy-AGokQJp45zor8vDRRJq/view?usp=sharing>

- └ Operations hydrobiologiques :

· https://drive.google.com/file/d/1lpOZceblFkMiPzwE-SDZWmy9_6jLXxSM/view?usp=sharing

- Conditions environnementales :

· <https://drive.google.com/file/d/1kbb9PoGHwCb8F9XAsqLhL02Fofk5Cer5/view?usp=sharing>

- Liste faune flore :

· <https://drive.google.com/file/d/1cOJQ9GrV9S7cFIBJvaKb0emIJYp6dske/view?usp=sharing>

Données disponibles sur le bassin Adour-Garonne

Physicochimie : 10-2015 à aujourd'hui

— Bassin adour-garonne

- └ Description des données :

- https://drive.google.com/file/d/1pB8BZGVFXFXqJ09palVIImZn_Ao1wQZs/view?usp=sharing

- └ Analyses Physicochimiques :

- https://drive.google.com/file/d/1OGg9P27v_-NrAUTyvdc4tqAAar9zqqZz/view?usp=sharing

- └ Stations Physicochimiques :

- <https://drive.google.com/file/d/1CmCYfmkgQrd8sCw8-S4WUqSmIVx1tyxZ/view?usp=sharing>

- └ Operations Physicochimiques :

- https://drive.google.com/file/d/1dk_9ahapnmJxFTHeQxLyjOtxeoXS5FJA/view?usp=sharing

- └ Conditions environnementales :

- https://drive.google.com/file/d/1ZN_4jQWfdwDoYG6rwQa80iVjidNvIAR-/view?usp=sharing

Objectifs

— Définir un use case Data, si possible avec une application ML

- └ Compréhension des données : quelle valeur métiers sont-elles susceptibles de porter ?
- └ Quels usages peuvent en être fait avec les données telles que disponibles actuellement ?
- └ Comment pourrait on donner de la valeur à l'utilisation de ces données ?
- └ Identifier les « manques », ou ce qui permettraient d'aller plus loin par rapport à l'état actuel
 - Couverture (spatiale, temporelle...)
 - Complétions : quelles informations additionnelles « internes » propre au domaine pourraient être pertinentes
 - Quels croisements : quelle données connexes pourraient permettre d'enrichir le use case
- └ Formaliser le problème sous un angle data-science / Machine Learning : type de problème
- └ Autant que faire se peut, ébaucher cet analyse data :
 - Identifier les complexités rencontrées, toujours sous un angle métier
- └ Se projeter : cette approche, si elle devait être industrialisée, quelles en seraient les contraintes (rafraichissement des données, pérennité des données...)

Organisation

— Collaborative :

- └ Travail en équipe : 5/6 personnes
- └ Présentation aux autres équipes pour critique constructive
- └ Définir une use case Data, Datascience, IA si possible avec une application ML

— Identifier les questions métier qui surviennent dans la définition de ce use case :

- └ Lors de l'analyse des données (représentativité d'un paramètre par rapport à un autre, incidence d'un manque...)
 - Quels sont les manques éventuels ?
- └ Lors de la phase de complétion des données, lors de l'adoption d'une stratégie ou une autre
 - Notamment en identifiant ce qui serait améliorable dans la collecte des différentes données (fréquence, représentation)
- └ Lors du croisement des données (quelles informations sont pertinentes, échelle des jointures...)
- └ Lors de la définition du use case :
 - Remettre les enjeux opérationnels du métier au cœur de la démarche : pour qui, pour quoi avec quelle « granularité »
 - Quelle utilisation pourrait être effectuée si cette démarche donnait des résultats
 - Ou seraient à mettre les efforts pour qualifier cette donnée