



HEALTH  
DATA HUB

## Health Data Hub

**Guide de contribution et de  
référencement d'un algorithme**


*Bibliothèque Ouverte d'Algorithmes en  
Santé*

# Sommaire

## **Guide de contribution et de référencement d'un algorithme à la Bibliothèque Ouverte d'Algorithmes en Santé (BOAS)**

1. Présentation de la Bibliothèque Ouverte d'Algorithme en Santé (BOAS)
2. Étapes et modalités de contribution à la Bibliothèque Ouverte d'Algorithme en Santé (BOAS)





# 1- Présentation de la Bibliothèque Ouverte d'Algorithme en Santé (BOAS)

# 1. Présentation de la Bibliothèque Ouverte d'Algorithme en Santé (BOAS)

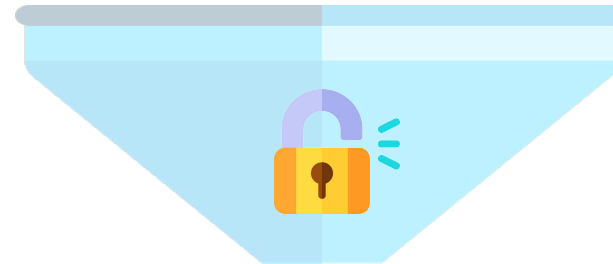
## Contexte



L'Appel à Manifestation d'Intérêt BOAS soutient et encourage le développement, la mise à jour et la validation d'algorithmes de ciblage sur le SNDS.



Codes et algorithmes développés dans le cadre de l'AMI BOAS



La BOAS est une bibliothèque centralisée d'algorithmes en santé partagés en Open Source



Codes et algorithmes facilitant l'étude des données de santé issues de tout type de projet

# 1. Présentation de la BOAS

## Objectifs

La Bibliothèque Ouverte d'Algorithme en Santé (BOAS) a un objectif double:

- Référencer les algorithmes de ciblage sur la base principale du SNDS et leur implémentation
- Référencer les algorithmes de traitement des données de santé pour faciliter leur réutilisation et leur implémentation

Les algorithmes **clairement documentés** et décrits en langage naturel, implémentés dans **différents langages de programmation, totalement ouverts** à la **revue** et **l'amélioration** par les pairs dans un **cadre transparent et collaboratif**, fourniront à la communauté une base de travail.

### Un projet collaboratif ouvert pour faciliter le traitement des données de santé



Une **bibliothèque centralisée** d'algorithmes en santé



Une **initiative collaborative** entre les différents **acteurs** de l'écosystème



Un **accompagnement** dans l'ouverture de codes sources.




Partage des **bonnes pratiques** de data management sur les données de santé



# 1. Présentation de la BOAS

## A quoi ressemble la bibliothèque ?

### Bibliothèque Ouverte d'Algorithmes en Santé (BOAS)

Rechercher un intitulé   10 par page  Les + récents

**Filtrer**

- Type d'auteur
- Objectif de l'algorithme
- Domaine Médical
- Langage de programmation
- Données d'application
- Validation
- Maintenance

Algorithmes de ciblage de patients atteints d'hépatite virale B et C chronique, ainsi que de phénotypes associés à ces maladies (cirrhoses, porteurs dits "sains", hépatite Delta).

AUTEURS ORGANISME DE RECHERCHE  
OBJECTIFS DE L'ALGORITHME Outils de ciblage dans la base principale du SNDS Outils de manipulation / transformation de la base principale du SNDS →  
DOMAINES MALADIES INFECTIEUSES LANGAGES DE PROGRAMMATION PYTHON R SQL  
DONNÉES D'APPLICATION BASE PRINCIPALE DONNÉES TABULAIRES  
VALIDATION EN COURS DE VALIDATION  
MAINTENANCE AD-HOC (EN FONCTION DES REMONTÉES DE PROBLÈMES, SUGGESTIONS)

Développement d'un algorithme de repérage des individus avec des limitations motrices ou organiques à partir des données de l'Assurance maladie en France.

AUTEURS AUTRE OBJECTIFS DE L'ALGORITHME Outils de ciblage dans la base principale du SNDS →  
DOMAINES AUTRE LANGAGES DE PROGRAMMATION SAS DONNÉES D'APPLICATION BASE PRINCIPALE  
VALIDATION EN COURS DE VALIDATION  
MAINTENANCE AD-HOC (EN FONCTION DES REMONTÉES DE PROBLÈMES, SUGGESTIONS)

Algorithme pour construire le top diabète de la cartographie de la CNAM (version G8) pour l'année 2019 en langage SAS et Python à partir des données synthétiques du HDH

AUTEURS ADMINISTRATIONS ET MINISTÈRE  
OBJECTIFS DE L'ALGORITHME Outils de cartographie des pathologies Outils de ciblage dans la base principale du SNDS →  
DOMAINES DIABÈTE LANGAGES DE PROGRAMMATION PYTHON SAS  
DONNÉES D'APPLICATION BASE PRINCIPALE VALIDATION NON VALIDE  
MAINTENANCE AD-HOC (EN FONCTION DES REMONTÉES DE PROBLÈMES, SUGGESTIONS)

Prise en charge des allergies aux pollens dans la région Normandie

AUTEURS PLATEFORME DE DONNÉES  
OBJECTIFS DE L'ALGORITHME ALGORITHMES DE REQUÊTE À LA DEMANDE →  
DOMAINES MALADIES RESPIRATOIRES LANGAGES DE PROGRAMMATION SAS SQL  
DONNÉES D'APPLICATION BASE PRINCIPALE VALIDATION VALIDE  
MAINTENANCE PAS DE MAINTENANCE

## Algorithme pour construire le top diabète de la cartographie de la CNAM (version G8) pour l'année 2019 en langage SAS et Python à partir des données synthétiques du HDH

 Lien vers le repo : [Gitlab](#)

### Objectifs de l'algorithme

OUTILS DE CARTOGRAPHIE DES PATHOLOGIES Outils de ciblage dans la base principale du SNDS

#### Objectifs et périmètre

L'algorithme ici présenté a pour objectif de cibler les personnes prises en charge pour un diabète dans la base principale du SNDS afin de créer le « Top Diabète » de la cartographie des pathologies créée et maintenue par la CNAM (version G8). Il s'appuie sur le programme source partagé par la CNAM et a été conçu pour le périmètre suivant :

- Caractéristique(s) ciblée(s) : Diabète, quel que soit son type
- Perspective : Construction du top diabète de la cartographie de la CNAM
- Périmètre géographique : France entière
- Périmètre historique programme source: Années 2015 à 2019 (incluses)
- Périmètre historique programmes adaptés : Années 2018-2019
- Régimes retenus : Ensemble des régimes d'assurance maladie

Version de la cartographie des pathologies : G8

Le programme source en SAS de la CNAM tourne sur les données des années 2015 à 2019.

Les versions Python et SAS adaptés de ce programme portent sur des données synthétiques pour les années 2018-2019 mais peuvent être étendues à d'autres années.

# 1. Présentation de la BOAS

*A quoi ressemble la bibliothèque ?*



Lien vers le repo : [Gitlab](#)

Health Data Hub / CNAM / Top Diabete

Top Diabete

main top-diabete Historique Rechercher un fichier Modifier Code

Update README.md Laurie Alla rédigé il y a 15 heures 0f4d7b03

Nom	Dernière validation	Dernière mise à jour
Version python	Téléverser un nouveau fichier	il y a une semaine
Version sas	Ajout versions sas	il y a une semaine
Programme_source_top_FDi...	Update Programme_source_-_top_FDi...	il y a une semaine
README.md	Update README.md	il y a 15 heures
Tables_et_variables_du_SN...	Téléverser un nouveau fichier	il y a une semaine

README.md

## Top Diabete

### Titre du projet :

Algorithme pour construire le top diabète de la cartographie de la CNAM (version G8) pour l'année 2019 en langage SAS et Python à partir des données synthétiques du HDH

Algorithme pour construire le top diabète de la cartographie de la CNAM (version G8) pour l'année 2019 en langage SAS et Python à partir des données synthétiques du HDH

## Objectifs de l'algorithme

OUTILS DE CARTOGRAPHIE DES PATHOLOGIES OUTILS DE CIBLAGE DANS LA BASE PRINCIPALE DU SNDS

### Objectifs et périmètre

L'algorithme ici présenté a pour objectif de cibler les personnes prises en charge pour un diabète dans la base principale du SNDS afin de créer le « Top Diabète » de la cartographie des pathologies créée et maintenue par la CNAM (version G8). Il s'appuie sur le programme source partagé par la CNAM et a été conçu pour le périmètre suivant :

- Caractéristique(s) ciblée(s) : Diabète, quel que soit son type
- Perspective : Construction du top diabète de la cartographie de la CNAM
- Périmètre géographique : France entière
- Périmètre historique programme source: Années 2015 à 2019 (incluses)
- Périmètre historique programmes adaptés : Années 2018-2019
- Régimes retenus : Ensemble des régimes d'assurance maladie

Version de la cartographie des pathologies : G8

Le programme source en SAS de la CNAM tourne sur les données des années 2015 à 2019.

Les versions Python et SAS adaptés de ce programme portent sur des données synthétiques pour les années 2018-2019 mais peuvent être étendues à d'autres années.



# 1. Présentation de la BOAS

## A quoi ressemble la bibliothèque ?

The screenshot shows the BOAS website interface. On the left, there is a 'Filtrer' (Filter) sidebar with categories like 'Type d'auteur', 'Objectif de l'algorithme', 'Domaine Médical', 'Langage de programmation', 'Données d'application', 'Validation', and 'Maintenance'. The main content area displays a search bar and a list of algorithm entries. One entry is highlighted with a magnifying glass: 'Développement d'algorithmes de ciblage de patients atteints d'hépatite virale B et C chronique'. Below this, there is a detailed view of the algorithm, including its objectives, author information (Fabien Carrat), and a list of files in a GitLab repository. The GitLab interface shows a repository for 'Top Diabete' with files like 'Version python', 'Version sas', 'Programme\_source\_top\_FDI...', 'README.md', and 'Tables\_et\_variables\_du\_SN...'. A magnifying glass is also placed over the GitLab repository name.

Sur le site internet du Health Data Hub, chaque **fiche de description sur la BOAS** renvoie au **dépôt de l'algorithme** sur une **plateforme ouverte et collaborative** (telle que GitLab ou GitHub) permettant de :

- **mettre à disposition** des algorithmes et codes sources à la communauté **quelque soit le langage de programmation**,
- **travailler en collaboration** avec la communauté avec un système de tickets,
- **faire évoluer** les codes et algorithmes afin de les perfectionner,
- **documenter les changements** de version et ainsi **historiser les versions** des codes sources,
- **faire valider** (ou approuver) les propositions de modification,
- **comparer** les différentes versions.





## 2 - Étapes et modalités de contribution à la bibliothèque

## 2. Étapes et modalités de contribution à la bibliothèque

*Étapes successives*



### Faire une review du code

La **relecture du code à ouvrir par un pair** permet notamment de :

- Détecter d'éventuelles **erreurs**,
- Améliorer la **qualité** du code,
- Inciter le rédacteur à **commenter** son code,
- **Consolider** les expertises,
- Rendre le code **interopérable**.

La réalisation d'un code review contribue à la création de logiciels **robustes** et de haute **qualité**.

## 2. Étapes et modalités de contribution à la bibliothèque

*Étapes successives*



Faire une  
review du  
code

Fournir une documentation  
détaillée

Un **format standardisé** de  
documentation contenant :

- Titre du projet et auteurs,
- Objectifs de l'algorithme,
- Méthodologie,
- Données utilisées,
- Validation,
- Date de mise à jour,
- Maintenance,
- Comment installer et utiliser l'algorithme,
- etc.

Un **template** de **documentation standardisé** est proposé pour vous guider dans la rédaction à cette étape.





## 2. Étapes et modalités de contribution à la bibliothèque

*Étapes successives*



Faire une  
review du  
code

Fournir une  
documen-  
tation

Attribuer un DOI au code  
pour garantir la parentalité

**DOI : Digital Object Identifier**

- Fournit un **lien stable** vers un objet scientifique et sa description
- Il s'agit d'un **identifiant numérique unique et pérenne** pour un objet scientifique que l'on souhaite rendre **citable**

**Un DOI peut être attribué à l'algorithme à ouvrir via la plateforme fabrica de DataCite sur laquelle le HDHD a un compte**

# 2. Étapes et modalités de contribution à la bibliothèque

*Étapes successives*



La licence définit **les modalités de réutilisation** et de **distribution** des versions modifiées. Nous recommandons les licences :

- **Apache 2.0**
- **Licence MIT**

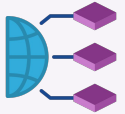
Ces licences sont **permissives**, autorisées aux **administrations** et permettent une **réutilisation libre, gratuite et sans restriction** dans le **respect de la citation des auteurs**



**Documentation :** ✓ Outil de la Commission Européenne

## 2. Étapes et modalités de contribution à la bibliothèque

*Modalités d'ouverture : deux cas de figures*



Les **modalités** d'ouverture dépendent de si le porteur de projet utilise ou non **une plateforme de dépôt**

- Cas de figure 1** : Le porteur a déjà un compte sur une plateforme de dépôt sur lequel l'algorithme sera référencé  
Dans ce cas, la fiche de présentation de l'algorithme sur la BOAS renverra vers le dépôt du porteur
- Cas de figure 2** : Le porteur n'utilise pas de plateforme de dépôt et souhaite utiliser le GitLab du Health Data Hub  
Dans ce cas, la fiche de présentation de l'algorithme sur la BOAS renverra vers le GitLab du Health Data Hub



## 2. Étapes et modalités de contribution à la bibliothèque

*Modalités d'ouverture :*

Cas de figure 1 :



**Le porteur a déjà un compte sur une plateforme de dépôt sur laquelle l'algorithme est référencé**



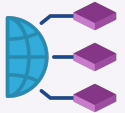
**Etape 1 : Création de la page de description de l'algorithme sur le dépôt du porteur via le template du readme.**



**Etape 2 : Référencement du dépôt sur la page de présentation de l'algorithme dans la BOAS sur le site internet du Health Data Hub**


## 2. Étapes et modalités de contribution à la bibliothèque

*Modalités d'ouverture : deux cas de figures*



Les **modalités** d'ouverture dépendent de si le porteur de projet utilise ou non **une plateforme de dépôt**

- Cas de figure 1** : Le porteur a déjà un compte sur une plateforme de dépôt sur lequel l'algorithme sera référencé  
Dans ce cas, la fiche de présentation de l'algorithme sur la BOAS renverra vers le dépôt du porteur

- 
- Cas de figure 2** : Le porteur n'utilise pas de plateforme de dépôt et souhaite utiliser le GitLab du Health Data Hub  
Dans ce cas, la fiche de présentation de l'algorithme sur la BOAS renverra vers le GitLab du Health Data Hub

# 2. Étapes et modalités de contribution à la bibliothèque

Modalités d'ouverture :



Cas de figure 2 :

**Le porteur de projet n'utilise pas de plateforme de dépôt et souhaite utiliser le GitLab du Health Data Hub**



**Etape 1 : Création d'un compte GitLab pour le porteur : étape nécessaire pour accéder au dossier du projet créé sur le GitLab du Health Data Hub**



**Etape 2 : Création des accès administrateurs au dossier du projet sur le GitLab du Health Data Hub pour le porteur**



**Etape 3 : Dépôt du contenu sur le dossier du projet sur le GitLab du Health Data Hub par le porteur**



**Etape 4 : Référencement du dépôt sur la page de présentation de l'algorithme dans la BOAS sur le site internet du Health Data Hub**



## 2. Étapes et modalités de contribution à la bibliothèque

*Modalités d'ouverture :*



Cas de figure 2 :

**Le porteur de projet n'utilise pas de plateforme de dépôt et souhaite utiliser le GitLab du Health Data Hub**

Autre option :



**Etape 1 : Envoi du contenu à ouvrir à l'équipe Open Science du Health Data Hub au format définitif prêt à l'emploi pour les réutilisateurs**



**Etape 2 : Dépôt du contenu dans le dossier du projet sur le GitLab du Health Data Hub par un membre de l'équipe Open Science**



**Etape 3 : Référencement du dépôt sur la page de présentation de l'algorithme dans la BOAS sur le site internet du Health Data Hub**

# Conclusion

Exemple de ressource externe disponible dans la BOAS

## Algorithmes de ciblage validés dans le cadre du projet RISH de l'IRDES (lauréat AMI BOAS 3)

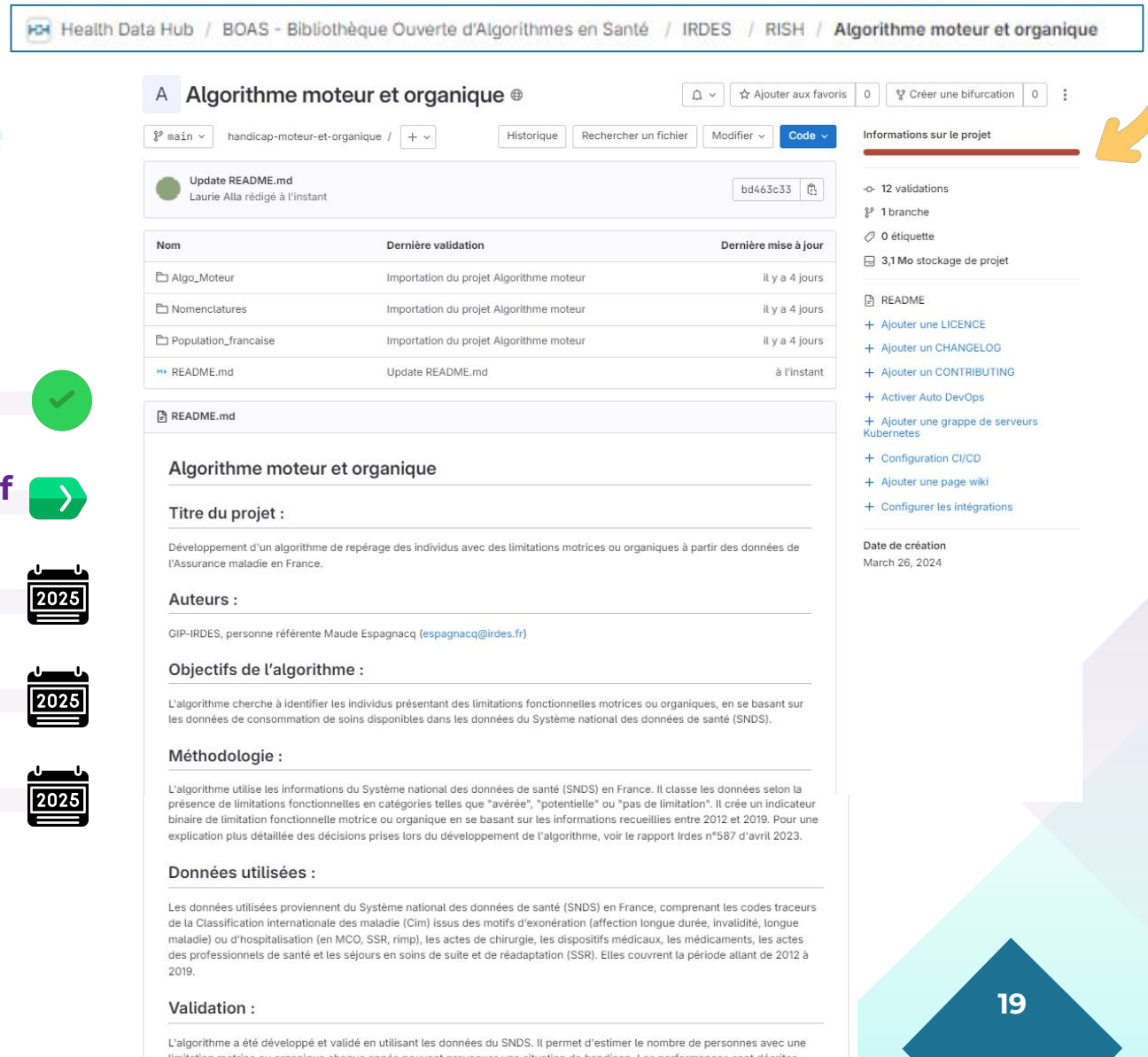
Ouverture algorithme moteur et organique 

Ouverture algorithme psychique intellectuel et cognitif 

Ouverture algorithme auditif 

Ouverture algorithme visuel 

Ouverture algorithme synthèse 



The screenshot shows the GitHub-like interface for the 'Algorithme moteur et organique' project. The breadcrumb trail is: Health Data Hub / BOAS - Bibliothèque Ouverte d'Algorithmes en Santé / IRDES / RISH / Algorithme moteur et organique. The page title is 'A Algorithme moteur et organique'. It features a commit history table, a README file, and a detailed project description.

Nom	Dernière validation	Dernière mise à jour
Algo_Moteur	Importation du projet Algorithme moteur	il y a 4 jours
Nomenclatures	Importation du projet Algorithme moteur	il y a 4 jours
Population_francaise	Importation du projet Algorithme moteur	il y a 4 jours
README.md	Update README.md	à l'instant

**Algorithme moteur et organique**

**Titre du projet :**

Développement d'un algorithme de repérage des individus avec des limitations motrices ou organiques à partir des données de l'Assurance maladie en France.

**Auteurs :**

GIP-IRDES, personne référente Maude Espagnacq (espagnacq@irdes.fr)

**Objectifs de l'algorithme :**

L'algorithme cherche à identifier les individus présentant des limitations fonctionnelles motrices ou organiques, en se basant sur les données de consommation de soins disponibles dans les données du Système national des données de santé (SNDS).

**Méthodologie :**

L'algorithme utilise les informations du Système national des données de santé (SNDS) en France. Il classe les données selon la présence de limitations fonctionnelles en catégories telles que "avérée", "potentielle" ou "pas de limitation". Il crée un indicateur binaire de limitation fonctionnelle motrice ou organique en se basant sur les informations recueillies entre 2012 et 2019. Pour une explication plus détaillée des décisions prises lors du développement de l'algorithme, voir le rapport Irdes n°587 d'avril 2023.

**Données utilisées :**

Les données utilisées proviennent du Système national des données de santé (SNDS) en France, comprenant les codes traceurs de la Classification internationale des maladies (Cim) issus des motifs d'exonération (affection longue durée, invalidité, longue maladie) ou d'hospitalisation (en MCO, SSR, rimp), les actes de chirurgie, les dispositifs médicaux, les médicaments, les actes des professionnels de santé et les séjours en soins de suite et de réadaptation (SSR). Elles couvrent la période allant de 2012 à 2019.

**Validation :**

L'algorithme a été développé et validé en utilisant les données du SNDS. Il permet d'estimer le nombre de personnes avec une limitation motrice ou organique chaque année pouvant provoquer une situation de handicap. Les performances sont décrites



**Avez-vous des questions ?**



HEALTH  
DATA HUB