




Health Data Hub

**Guide de contribution et de
référencement d'un algorithme**
*Bibliothèque Ouverte d'Algorithmes en
Santé*



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 et la validation d'algorithmes ainsi que de programmes informatiques, contribuant à faciliter l'utilisation des données de santé.



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



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



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

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** à toutes les étapes de 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)

Filtrer

VALIDER

Type d'auteur

Objectif de l'algorithme

Domaine Médical

Langage de programmation

Données d'application

Validation

Maintenance

Rechercher un intitulé

10 par page

Les + récents

Algorithmes de ciblage de patients atteints d'hépatite virale B et C chronique, ainsi que de phénotypes associés à ces maladies (cirrheses, 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

T Top Diabete



☆ Ajouter aux favoris

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

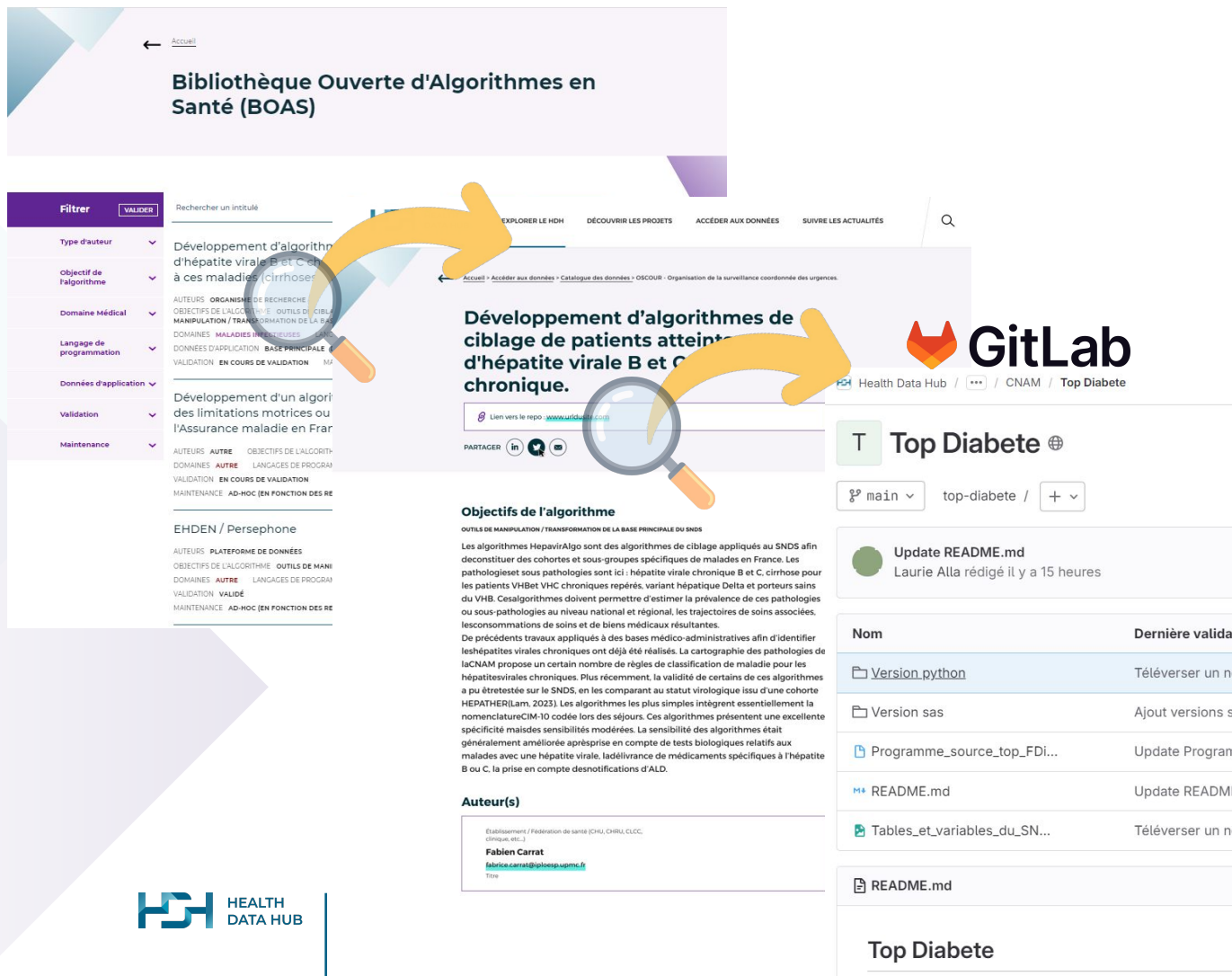
Le programme source en SAS de la CNAM tourne sur les données des années 2015 à 2019.

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 ?



Accueil

Bibliothèque Ouverte d'Algorithmes en Santé (BOAS)

Rechercher un intitulé

Explorer le HDH | Découvrir les projets | Accéder aux données | Suivre les actualités

Filtrer [VALIDER]

Type d'auteur

Objectif de l'algorithme

Domaine Médical

Langage de programmation

Données d'application

Validation

Maintenance

Développement d'algorithme d'hépatite virale B et C à ces maladies (cirrhose)

AUTEURS ORGANISME DE RECHERCHE OBJECTIFS DE L'ALGORITHME OUTILS DE CIBLAGE MANIPULATION / TRANSFORMATION DE LA BASE PRINCIPALE DE DONNÉES D'APPLICATION BASE PRINCIPALE DE DONNÉES D'APPLICATION EN COURS DE VALIDATION MAINTENANCE

Développement d'un algorithme des limitations motrices ou l'Assurance maladie en France

AUTEURS AUTRE OBJECTIFS DE L'ALGORITHME DOMAINES AUTRE LANGAGES DE PROGRAMMATION VALIDATION EN COURS DE VALIDATION MAINTENANCE AD-HOC (EN FONCTION DES REQUIS)

EHDEN / Persephone

AUTEURS PLATEFORME DE DONNÉES OBJECTIFS DE L'ALGORITHME OUTILS DE CIBLAGE MANIPULATION / TRANSFORMATION DE LA BASE PRINCIPALE DE DONNÉES D'APPLICATION BASE PRINCIPALE DE DONNÉES D'APPLICATION EN COURS DE VALIDATION MAINTENANCE AD-HOC (EN FONCTION DES REQUIS)

Développement d'algorithmes de ciblage de patients atteints d'hépatite virale B et C chronique.

Accueil > Accéder aux données > Catalogue des données > OSCOUR : Organisation de la surveillance coordonnée des urgences.

Lien vers le repo : www.gitlab.com

PARTAGER

Objectifs de l'algorithme

OUTILS DE MANIPULATION / TRANSFORMATION DE LA BASE PRINCIPALE DU SNDS

Les algorithmes HepaVirAlgo sont des algorithmes de ciblage appliqués au SNDS afin de constituer des cohortes et sous-groupes spécifiques de malades en France. Les pathologies et sous pathologies sont ici : hépatite virale chronique B et C, cirrhose pour les patients VHB/VHC chroniques repérés, variant hépatite Delta et porteurs sains du VHB. Ces algorithmes doivent permettre d'estimer la prévalence de ces pathologies ou sous-pathologies au niveau national et régional, les trajectoires de soins associées, les consommations de soins et de biens médicaux résultantes.

De précédents travaux appliqués à des bases médico-administratives afin d'identifier les hépatites virales chroniques ont déjà été réalisés. La cartographie des pathologies de la CNAM propose un certain nombre de règles de classification de maladie pour les hépatites virales chroniques. Plus récemment, la validité de certains de ces algorithmes a pu être testée sur le SNDS, en les comparant au statut virologique issu d'une cohorte HEPATHERILAM 2023. Les algorithmes les plus simples intègrent essentiellement la nomenclature CIM-10 codée lors des séjours. Ces algorithmes présentent une excellente spécificité mais des sensibilités modérées. La sensibilité des algorithmes était généralement améliorée après prise en compte de tests biologiques relatifs aux malades avec une hépatite virale, l'adhésion de médicaments spécifiques à l'hépatite B ou C, la prise en compte des notifications d'ALD.

Auteur(s)

Établissement : Fédération de santé (CHU, CHRL, CLCC, clinique, etc.)

Fabien Carrat

fabien.carrat@psoeap.umc.fr

Titre

GitLab

Health Data Hub / CNAM / Top Diabète

Top Diabète

main top-diabète / +

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

Nom	Dernière validation
Version python	Téléverser un nouveau fichier
Version sas	Ajouter versions saines
Programme_source_top_FDI...	Update Programme_source_-_top_FDI...
README.md	Update README.md
Tables_et_variables_du_SN...	Téléverser un nouveau fichier
README.md	

Top Diabète

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.

1. Présentation de la BOAS

Etat des lieux



- 41 ressources sont disponibles dans la BOAS depuis sa mise en ligne en avril 2024



Algorithmes de ciblage dans la base principale du SNDS

Top Diabète

RISH : limitations cognitives

Requête type : affection longue durée

RISH : limitations motrices/handicap

Requête type : prestation dans le DCIR

DIONISOS : sclérose en plaques

Requête type : séjours hospitaliers

HepaVirAlgo : hépatite B et C

MICI : rectocolite hémorragique

Insuffisance rénale chronique

BIS : Asthme de l'enfant

Distinction diabète T1 vs T2



Requêtes à la demande

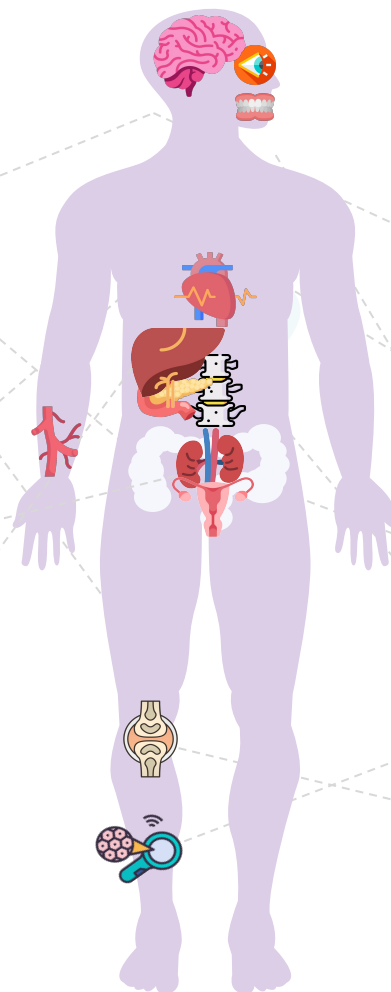
Requête à la demande : femmes ayant droits

Requête à la demande : dépistage des cancers féminins

Requête à la demande : socio-démographie

Requête à la demande : recours à l'IVG

Requête à la demande : allergies en Normandie



Algorithmes de traitement des données de santé



ALIA : sous-types d'AVC hospitalisés



Allergologie : prédiction des allergènes

Maladies rares dentaires, orales et faciales

Cancer du col de l'utérus : détection de lésions

Grefe pulmonaire : prédiction du rejet

Mélanome : prédiction de la récurrence

Cytologie : classification de leucocytes



Cartographie des pathologies



Outils de manipulation, transformation des données

OMOP : Standardisation de la base principale

Magic Loop : génération des tables de consommation de soin

Deep.piste : pseudonymisation de fichiers DICOM

Extract metadata : extraction de métadonnées (données tabulaires et d'imagerie)

Macro SAS : d'extraction de données SNDS


Librairies de manipulation de données sur les l'EDS

Outil de conversion de format de lame d'anatomopathologie



En moyenne 380 visites/mois





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

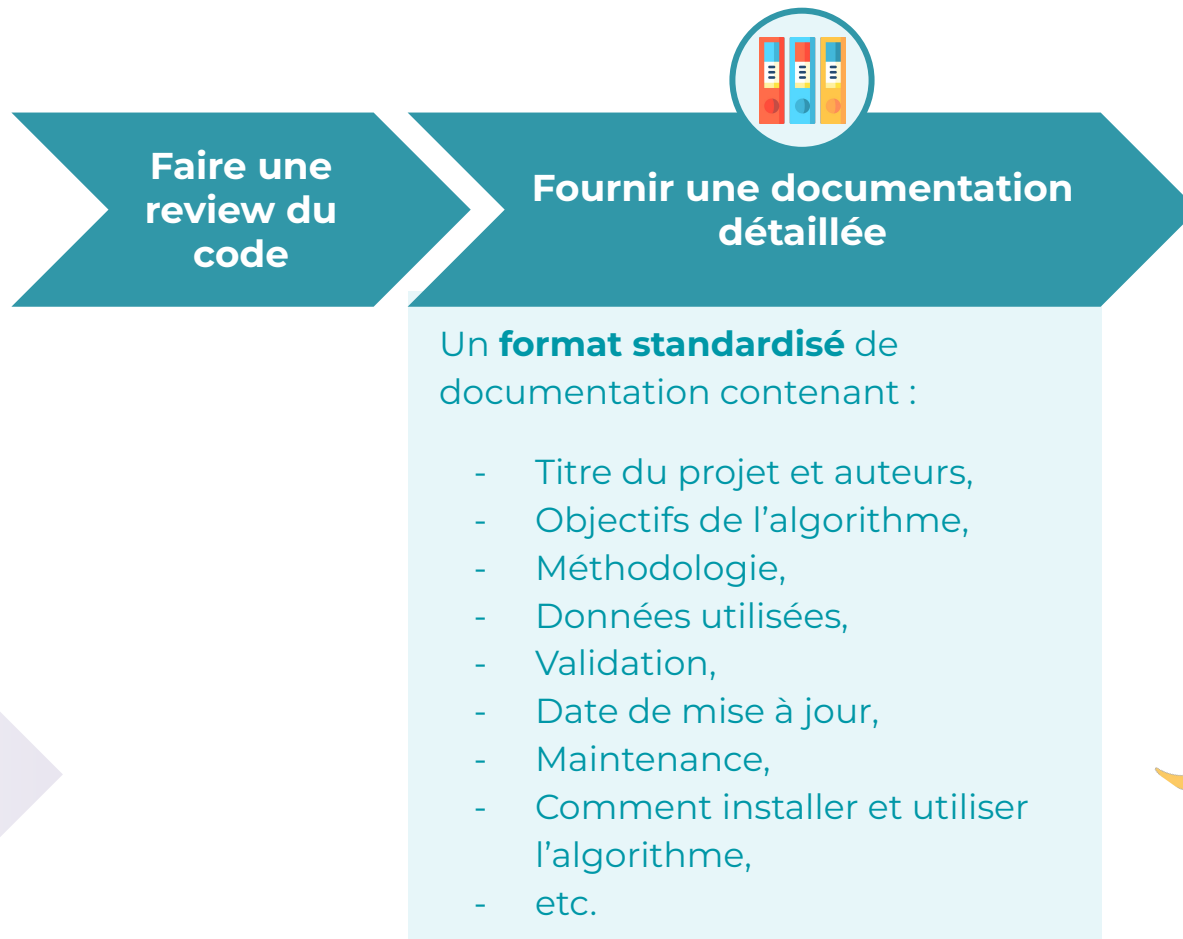
La **relecture du code à ouvrir par un pair** permet notamment de :

- 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



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 fabrika de DataCite sur
laquelle le HDH 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**

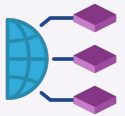
Ces licences sont **permissives**, autorisées aux **administrations** et permettent une **réutilisation libre, gratuite et sans restriction** dans le **respect de la citation des auteurs**



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



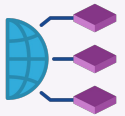
Etape 1 : Création de la page de description de l'algorithme sur le dépôt du porteur via le template du readme.




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



Algorithme moteur et organique

main

handicap-moteur-et-organique

Historique

Rechercher un fichier

Modifier

Code

Update README.md

Laurie Alla rédigé à l'instant

bd463c33

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

README.md

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

Informations sur le projet

12 validations

1 branche

0 étiquette

3,1 Mo stockage de projet

README

Ajouter une LICENCE

Ajouter un CHANGELOG

Ajouter un CONTRIBUTING

Activer Auto DevOps

Ajouter une grappe de serveurs Kubernetes

Configuration CI/CD

Ajouter une page wiki

Configurer les intégrations

Date de création

March 26, 2024



Avez-vous des questions ?



HEALTH
DATA HUB