

HEALTH DATA HUB x GRAND DEFI

Appel à Manifestation d'intérêt - Data Challenge en santé

Mars 2021

Qu'est ce qu'un Data Challenge?

Une Problématique

Nous accompagnons dans l'organisation de Data Challenge qui visent à répondre à des **problématiques médicales précises** à l'aide de **l'analyse des données** mises à disposition

Une base de données

Une **base de données de santé anonymisées et homogène** doit être constituée pour servir de source d'information à l'analyse. Une étape d'annotation des données peut être nécessaire pour constituer une vérité de terrain à fournir aux algorithmes

Des participants à la compétition

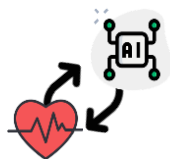
Des data scientists du monde entier accèdent à ces données et développent des **algorithmes d'apprentissage automatique** pour répondre à la question médicale posée.

Une récompense

Les meilleurs équipes sont récompensées à la fin de la compétition par un prix.

Les objectifs de l'appel à manifestation d'intérêt

Le présent appel à manifestation d'intérêt vise à sélectionner des projets de Data Challenge qui bénéficieront d'un accompagnement d'un an de bout en bout par le Health Data Hub et d'un financement du Grand Défi. Cette démarche poursuit les intérêts et objectifs suivants:



Favoriser les interactions
entre le monde de la
Santé et le monde de la
Science des données



Favoriser le
développement d'outils
précis d'assistance aux
professionnels de santé



Promotion de
l'innovation



Occasion pour les
compétiteurs de se faire
repérer par des acteurs
majeurs du domaine et
dénicher des talents



Démontrer par des
résultats rapides
le bénéfice concret de
l'utilisation de la donnée



Enrichir le catalogue de
données et d'algorithmes
partagés à toute la
communauté

Les projets Data Challenges ciblés

Les projets sélectionnés lors de cet appel à manifestation d'intérêt devront répondre aux critères ci-dessous.



Intérêt de la question clinique

La question médicale posée doit présenter un intérêt clinique et être originale.



Pertinence de l'approche prédictive pour y répondre dans le délai imparti

La question médicale posée doit pouvoir être traitée par de l'IA, le score de performance des algorithmes doit être mesurable et les données d'une qualité suffisante pour que les résultats soient fiables.



Faisabilité de la collecte des données et de leur anonymisation

Les données doivent pouvoir être collectées en nombre suffisant dans un temps imparti à un ou des formats numériques homogènes. Les données doivent pouvoir être totalement anonymisées dans le respect du RGPD.



Partage des données

Les données collectées et anonymisées dans le cadre du projet seront partagées en Open Data sur le catalogue des données du Health Data Hub à l'issue du Data Challenge. Des modalités de citation des bases sont toutefois prévues. Les équipes gagnantes devront partager leur algorithme publiquement pour obtenir la récompense prévue.



Forte implication et faisabilité du projet

Une forte implication de votre part est requise, sur toute la durée de l'organisation du data challenge. Plusieurs personnes de votre organisation devront s'impliquer dans le Conseil Scientifique et ou dans le Comité d'Organisation. Les grands jalons proposés pour le projet doivent le rendre réalisable en environ un an.



Combinaison de plusieurs spécialités médicales (facultatif)

Le projet peut faire intervenir plusieurs spécialités médicales et combiner des données de natures différentes mais il devra s'assurer que cela ne remet pas en cause leur caractère anonyme.

L'accompagnement proposé par le HDH

Le [Health Data Hub](#) ou “plateforme des données de santé” est un groupement d'intérêt public mis en place fin 2019 après la publication de la loi “Organisation et Transformation du Système de Santé”. Sa mission principale est de permettre à des porteurs de projets de santé habilités d'accéder à des données de santé non nominatives via une plateforme technologique sécurisée tout en respectant les droits des citoyens. Ces projets doivent obligatoirement servir un intérêt public, c'est-à-dire, viser à améliorer la qualité des soins et accompagner les patients. Parallèlement, il admet un rôle d'animation de l'écosystème dans le but de promouvoir les initiatives innovantes et de favoriser le partage de connaissance entre les acteurs. C'est dans ce cadre, que le HDH est amené à organiser des Data Challenge. Le HDH accompagnera les projets sélectionnés à toutes les étapes de leur réalisation en apportant des moyens techniques, humains, et financiers adaptés.



Définition et cadrage du Data Challenge

- ✓ Aide au cadrage du projet dans son intégralité
- ✓ Aide juridique notamment relativement à la réglementation applicable au traitement des données de santé
- ✓ Coordination avec des prestataires spécialisés dans l'hébergement de Data Challenges



Aide à la collecte et à l'anonymisation des données

- ✓ Accompagnement dans la collecte des données
- ✓ Accompagnement dans l'anonymisation des données
- ✓ Accompagnement technique dans la phase de labellisation des données
- ✓ Partage des bonnes pratiques



Gestion de l'infrastructure d'hébergement des données

- ✓ Accompagnement par des data engineers & data managers pour constituer la base de données du Data Challenge
- ✓ Serveurs sécurisés



Mettre en valeur les résultats

- ✓ Communication (vidéos, mise en Open Data, interview des gagnants, ...)
- ✓ Accompagnement à la publication d'articles scientifiques concernant le Data Challenge et ses résultats
- ✓ Accompagnement à l'ouverture des données
- ✓ Mise en relation des acteurs de l'écosystème

Possibilité de financement par le Grand Défi en IA et santé

En complément de l'accompagnement par le HDH, les projets sélectionnés pourront bénéficier d'un financement du Grand Défi : « **Améliorer les diagnostics médicaux à l'aide de l'intelligence artificielle** »

<https://www.gouvernement.fr/grand-defi-sante-ameliorer-les-diagnostics-medicaux-a-l-aide-de-l-intelligence-artificielle>

Le financement en question pourra atteindre 200 K€.

Le financement sera proposé sur le régime d'aide de-minimis.

Ce régime d'aide peut financer les projets à hauteur de 200K€ maximum sur une période glissante de 3 ans.

Les projets financés seront invités à déclarer les aides reçues au titre de-minimis au cours des 3 dernières années.

The logo for bpi france, with 'bpi' in dark grey and 'france' in yellow.

L'organisation du Data Challenge nécessitera une forte mobilisation de votre équipe.

Un Data Challenge réussi passe aussi par une forte implication de l'ensemble des partenaires !



Définition et cadrage du Data Challenge

- ✓ Implication de plusieurs personnes dans le Comité d'Organisation et le Conseil Scientifique du Data Challenge sur toute la durée de l'organisation
- ✓ L'organisme organisateur du Data Challenge sera le responsable juridique du projet



Collecte et anonymisation des données

- ✓ Motivation et implication des organismes à partir desquelles les données sont collectées
- ✓ Mise en place de la stratégie de la collecte des données (numérisation, anonymisation...)



Gestion de l'infrastructure d'hébergement des données

- ✓ Suivi de la gestion des données



Mettre en valeur les résultats

- ✓ Rédaction d'articles dans des revues scientifiques
- ✓ Collaboration avec le HDH dans les travaux de communication
- ✓ Collaboration dans la mise en oeuvre de la politique de partage des données

Exemple de réalisation : Détection des lésions du col de l'utérus en partenariat avec la SFP



Premiers échanges entre la SFP et le HDH
Constitution d'un **Comité d'Organisation**
et d'un **Conseil Scientifique**



Collecte des données
auprès des Centres
inclueurs – **20 centres,**
5000 lames collectées



Annotation de **2500**
lames



Data Challenge



Description de l'organisation

Comité de travail	Composition	Rôle	Fréquence de réunion
Equipe Projet	-Chef de Projet HDH avec l'appui de juristes et de Data Engineers du HDH -Président de la SFP -Secrétaire Générale de la SFP	-Suivi quotidien du projet (suivi juridique, organisationnel et technique)	-suivi quotidien
Comité d'organisation	-Président de la SFP -Secrétaire Générale de la SFP -DPO de la SFP	-Coordination auprès des centres inclueurs -Gestion des prestataires	-Réunions mensuelles avec les centres inclueurs -Réunions mensuelles de coordination
Conseil Scientifique	-3 médecins membres de la SFP -Directeur scientifique du HDH -Professeur associé à l'université de Paris.	-Définition de la question médicale -Mise en place d'une métrique pour évaluer la performance des algorithmes	-Réunions mensuelles
Comité d'Annotation	-Secrétaire Générale de la SFP -5 médecins membres de la SFP	-Mise en place d'une stratégie d'annotation des données -Annotation diagnostique des données	-Réunions fréquentes lors de la phase d'annotation des données qui a duré 2 mois

Exemple de réalisation : Détection des lésions du col de l'utérus en partenariat avec la SFP

Déroulé de la compétition

- **La phase d'apprentissage** : les compétiteurs entraînent leur algorithme à partir des données annotées.
- **La phase de test intermédiaire** : les compétiteurs testent leur algorithme sur un jeu de données intermédiaire.
- **La phase de validation finale** : soumission finale des algorithmes. L'ensemble de la compétition a duré 2 mois.

574

participants
du monde
entier

36

équipes
constituées

Des scores
très
prometteurs
(Score de 94%)

Interprétation
clinique des
**résultats en
cours**



Data Challenge en pathologie du col de l'utérus - Société Française de Pathologie - Health Data Hub

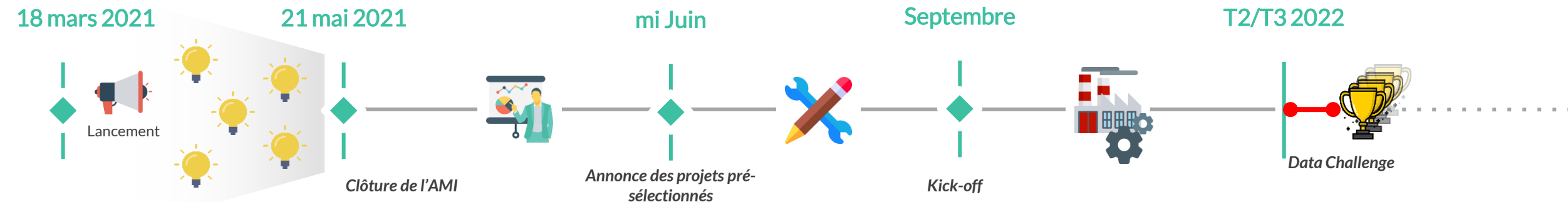
<https://www.youtube.com/watch?v=Ue0Lt1RKaAE&t=502s>

Exemple : répartition type des frais pour une société savante

Répartition type des frais pour une société savante

	Coût en 2020	Porteur du coût
Plateforme Data Challenge (2 mois)	env. 70 000 €	BPI via GD
Récompense gagnants Data Challenge	25 000 €	Société savante
Prestation DPO	env. 6 000 €	BPI via GD
Rétribution inclusion données (attention ce coût ne peut être pris en charge ni par la BPI ni par le HDH)	env. 25 000 €	Société savante
Serveur d'hébergement HDH, disques durs, etc.	env. 1 000 €	HDH
Envoi courrier d'information aux patients	env. 800 €	HDH
Outil d'annotation	env. 6 000 €	Société savante /HDH
TOTAL	Env. 130 000 €	

Modalités de sélection et calendrier de l'appel à manifestation d'intérêt



Dépôt des candidatures

Les porteurs de projets sont invités à déposer leur candidature sur la plateforme TPS d'ici le **21/05/2021**.

Auditions & pré-sélection

Les candidatures seront analysées par **un jury** pour établir une pré-sélection sur la base des critères présentés en page 4

Sélection des projets

Approfondissement des échanges avec les projets pré-sélectionnés pour affiner leur cadrage et s'assurer de leur réelle faisabilité.

Sélection des projets

Lancement des Data Challenge

Lancement des projets sélectionnés. Selon la complexité de la collecte des données, la date de lancement et la durée du Data Challenge varieront.

Accompagnement Post Challenge

Valorisation des résultats à travers des outils de communication (vidéos, publications scientifiques, interview des gagnants...)

Mise en Open Data des données & des algorithmes

Modalités de réponse



Formulaire de
Candidature

- <https://www.demarches-simplifiees.fr/commencer/ami-data-challenge>



Date limite
de candidature

- 21 Mai 2021



Plateforme
de remise

- [démarches-simplifiées.fr](https://www.demarches-simplifiees.fr)



Contact

- data.challenge@health-data-hub.fr