

## Communiqué de presse

Paris, le 19 mai 2022

### Le Health Data Hub et Octopize s'associent sur la génération de données synthétiques

Le Health Data Hub et Octopize-Mimethik Data lancent ce mois-ci une collaboration pour la création d'un tutoriel qui permettra d'explorer de manière interactive et pédagogique différentes approches à la génération de données synthétiques en santé. Prenant la forme d'un notebook Python, ce tutoriel présentera les grands enjeux associés à cette problématique, ainsi que les avantages et inconvénients de différentes méthodes permettant de générer de telles données. Le notebook sera publié en open source une fois finalisé, au troisième trimestre 2022.

#### Une collaboration pour mieux appréhender les données synthétiques

Cette collaboration permet à deux approches existantes de la génération de données synthétiques représentatives des données source de se rencontrer pour créer un tutoriel commun et pédagogique sur le sujet. Le Health Data Hub a développé et publié en open source un générateur de données fictives respectant la structure formelle d'une base de données de santé, en se basant uniquement sur le schéma de la base source. La société Octopize-Mimethik data a, quant à elle, développé une solution dite d'Avatarisation permettant de générer, à partir des données sources, des profils fictifs (les Avatars) qui "*conservent la qualité et la structure des données originales*". La collaboration ambitionne de diffuser la connaissance sur les nouvelles approches de données synthétiques et accélérer leurs usages au profit de la recherche médicale.

#### L'utilisation des données synthétiques

Le respect de la confidentialité est un enjeu majeur du traitement des données de santé, du fait de leur sensibilité. Mesures de sécurité techniques et organisationnelles, pseudonymisation, anonymisation des résultats font généralement partie de l'arsenal déployé pour le traitement sécurisé de ces données source à des fins de recherche.

Ce tutoriel permettra d'explorer une toute autre approche pour garantir le respect de la confidentialité des données sensibles : utiliser des données synthétiques. Générées de manière à ne présenter aucun risque pour la vie privée - puisqu'elles ne sont associées à aucune personne réelle - tout en maintenant la valeur informative des données originales sur certains aspects cruciaux, ces données synthétiques permettent de découpler les usages des données de santé.

Grâce à leur collaboration, Octopize et le Health Data Hub donneront ainsi à la communauté les clés de compréhension de la génération de données de synthèse et de leur potentiel de réutilisation. Formation, estimations de faisabilité, appréhension de jeux de données complexes, voire réalisation d'études complètes avec des résultats statistiquement équivalents : les données de synthèse constituent une voie prometteuse pour accélérer le développement des projets de recherche par la valorisation des données de santé.

### **A PROPOS DU HEALTH DATA HUB**

Le Health Data Hub est un groupement d'intérêt public créé par la Loi du 24 juillet 2019 relative à l'organisation et la transformation du système de santé. Il associe 56 parties prenantes, en grande majorité issues de la puissance publique (CNAM, CNRS, France Assos Santé...) et met en œuvre les grandes orientations stratégiques relatives au Système National des Données de Santé fixées par l'État et notamment le ministère des Solidarités et de la Santé. C'est un service à destination de l'écosystème de santé, des acteurs à l'origine de la collecte de données, des porteurs de projets d'intérêt général et de la société civile. En ce sens, il promeut l'innovation en santé et l'accessibilité des données et des connaissances par le biais, entre autres, d'événements fédérateurs comme l'organisation de data challenge et d'appels à projets.

Abonnez-vous à l'infolettre sur le [site internet](#)

[Foire Aux Questions](#) du HDH

Retrouvez [nos engagements citoyens](#)

Suivez le Health Data Hub sur [LinkedIn](#) et [Twitter](#)

#### **Contact presse :**

[presse@health-data-hub.fr](mailto:presse@health-data-hub.fr)

### **A PROPOS D'OCTOPIZE**

Octopize – Mimethik Data est une startup Deeptech nantaise qui ambitionne de devenir le leader européen de l'anonymisation. Elle a développé et breveté une méthode unique d'anonymisation des données personnelles, dont la conformité est attestée par la CNIL en juin 2020 : les avatars. La méthode est commercialisée sous forme de logiciel ou de service permettant de nouveaux usages de façon éthique. Elle est déjà reconnue dans le secteur de la santé et dans d'autres verticales. La startup compte une dizaine de personnes. En septembre 2021, Octopize a réalisé une levée de fonds de 1,5 million d'euros auprès de plusieurs investisseurs en capital-risque, Bpifrance et des Business Angels. Lauréate 2022 du concours i-Nov, sur décision du Premier ministre, elle s'ouvre à d'autres secteurs économiques et confirme son ambition. .

Pour en savoir plus : <https://octopize-md.com/>

Fondateur : Olivier BREILLACQ – [linkedin.com/in/olivier-breillacq](https://www.linkedin.com/in/olivier-breillacq)

#### **Contact presse :**

[contact@octopize-md.com](mailto:contact@octopize-md.com)