



Guy FAGHERAZZI

PhD, HDR. Chercheur en épidémiologie, Responsable Scientifique Etude E4N Centre de Recherche en Epidémiologie et Santé des des Populations Inserm UMR 1018, Université Paris Sud – Paris Saclay.

L'épidémiologie à l'ère du digital.

Utilisation des objets connectés et des réseaux sociaux dans les études de cohorte

janvier 2018



Cliquez ici pour voir l'intégralité des ressources associées à ce document

L'épidémiologie à l'ère du digital

*Utilisation des objets connectés et des réseaux sociaux
dans les études de cohorte*

Guy Fagherazzi, PhD, HDR
Chercheur en épidémiologie

E-Mail : guy.fagherazzi@gmail.com / Twitter : [@GFaghe](https://twitter.com/GFaghe)



Vers un omique digital

Génome
Epigénome
Protéome
Transcriptome
Métabolome
Microbiome
et...

LE "DIGITOSOME"

= l'intégralité des données
digitales générées par un
individu au cours de sa vie
(smartphones, capteurs,
dispositifs médicaux, réseaux
sociaux, géolocalisation,...)





Besoins en épidémiologie

- ▶ Maximiser le recrutement & limiter l'attrition
- ▶ Collecter des données de bonne qualité sur les facteurs et maladies d'intérêt
- ▶ Optimiser les coûts d'analyse et de traitement

**Gold standard =
étude de cohorte**

Etude prospective
sur une longue
période

Avant la e-santé & le Big Data

1940-50 : Premières études de cohorte

- ▷ L'étude Framingham (santé cardiovasculaire) (1948)
- ▷ L'étude de la santé des survivants des bombes atomiques au Japon (1950)

1970 : Premières grandes cohortes

- ▷ L'étude de la Nurses' Health Study (1976) : études des risques liés à la prise de pilule

1990 : Nouvelle vague d'études à grande échelle

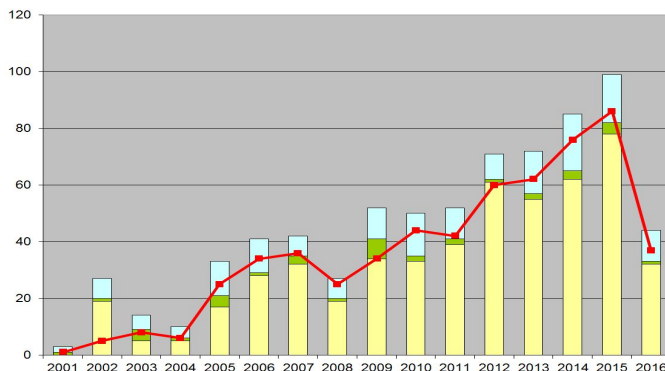
- ▷ The Million Women Study UK (1996) : santé des femmes
- ▷ L'étude E3N (1990) : facteurs de risque de cancer du sein



La plus grande étude sur la santé des femmes en France

- ▷ 98 995 femmes
- ▷ 27 ans de suivi
- ▷ Questionnaires tous les 2-3 ans depuis 1990

797
publications



Cancer



Diabète



Autres maladies
chroniques



Nutrition



Facteurs
hormonaux



Pharmaco-
épidémiologie

Prélèvements sanguins

Prélèvements salivaires

06/1990

01/1992

06/1993

01/1995

04/1997

07/2000

07/2002

07/2005

06/2008

12/2011

12/2014

Q1

Q2

Q3

Q4

Q5

Q6

Q7

Q8

Q9

Q10

Q11

Niveau socio-économique, passé gynécologique, tabac, activité physique, antécédents médicaux personnels / familiaux, prise de médicaments,



Alimentation de l'année précédente, alcool, tabac

Tabac, alcool, activité physique

Antécédents familiaux de cancer

Mesures anthropométriques, tabac

Enfance, tabac, activité physique



Alimentation de l'année précédente, compléments alimentaires, alcool, tabac, activité physique, dépression

Autonomie, dépression, santé bucco-dentaire, capacités visuelles et auditives

Traitements, compléments alimentaires

Mises à jour de l'état de santé et des traitements médicaux dans tous les questionnaires

Passé gynécologique, prise de traitements hormonaux



Mises à jour des cycles menstruels ou du statut ménopausique

Mises à jour des mesures anthropométriques

Données de remboursement de la MGEN

e3n.fr

Pionnière dans le recueil de données dès 1990

- ▷ Scanner optique des questionnaires papiers (600 pages de questionnaires par heure)
- ▷ Système de reconnaissance de caractère et lecture optique
- ▷ Vidéo-codage and vérification des données numérisées

Capable de gérer une masse de données importante



Pionnière

- ▷ Scarce
- (600)
- ▷ Système
- lecture
- ▷ Vidéo-
- numérique

Capable de g

**La production de connaissance en épidémiologie
est actuellement un long processus !**

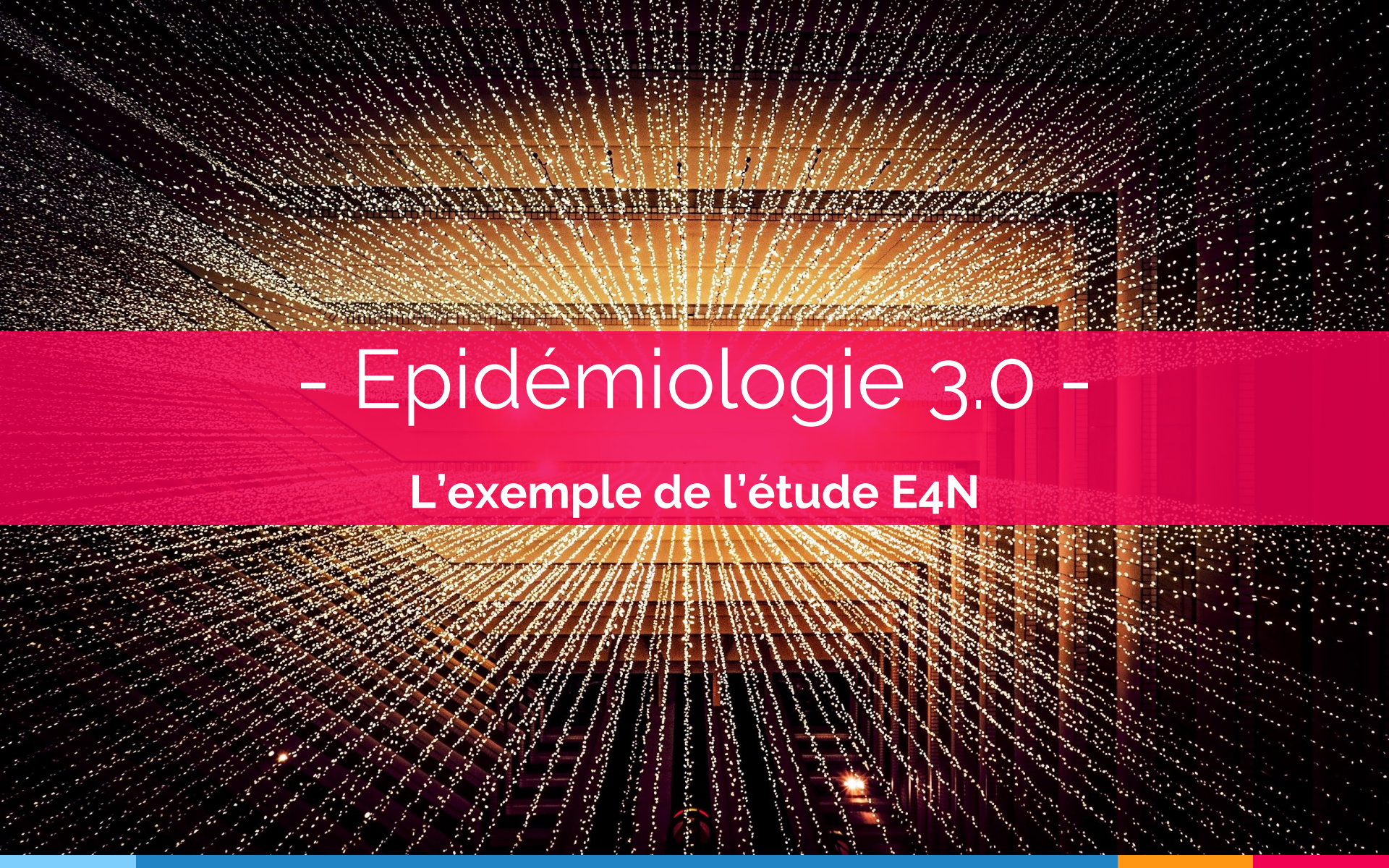
Cela nécessite :

- Du temps (≈ 3 ans pour mettre à disposition des données aux chercheurs)
- Nombreuses ressources humaines
- Budget ++



Comment faire mieux et plus vite ?





- Epidémiologie 3.0 -

L'exemple de l'étude E4N



L'étude prospective E4N

Une cohorte familiale
unique en son genre !

Sélectionnée comme
"Investissement d'Avenir" par
l'Agence Nationale de
Recherche (ANR)



+ de 150 000 participants

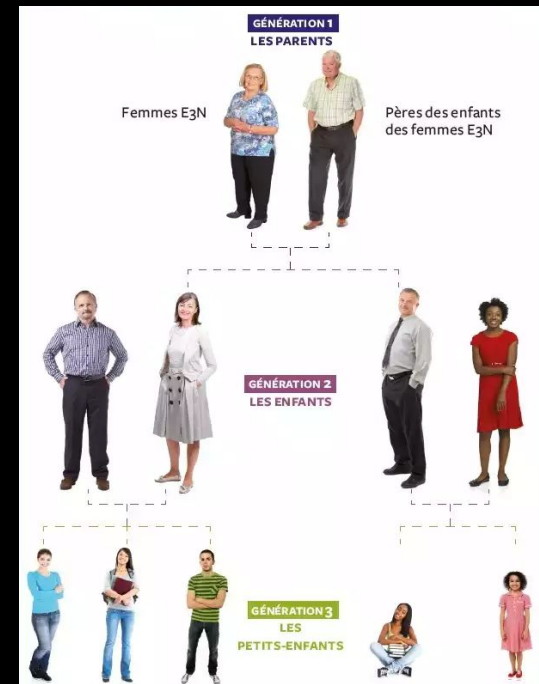
100 000 femmes E3N

20 000 pères biologiques des enfants des femmes E3N

50 000 enfants (en 2018)

20 000 petits-enfants (en 2019)

3 générations



1^{re} génération : questionnaires papiers

2^{ème} & 3^{ème} générations : e-cohorte en ligne

Échantillon de salive pour tous les participants



Trans-générationnel

- ▷ Hérité et transmission des déterminants de santé
- ▷ Génétique et épigénétique des pathologies chroniques

Expertise sur les expositions

- ▷ Epigénétique et mode de vie (tabac, AP...)
- ▷ Microbiote et alimentation
- ▷ Mobilité socioéconomique à travers les générations et impact sur la santé

e-épidémiologie

- ▷ Intégrer les nouvelles technologies pour le recueil d'informations en épidémiologie

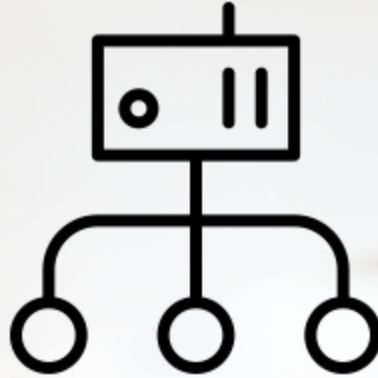
Quelques thématiques





La plateforme web

Data Hub



Questionnaires



SMS



Smartphones



Données
biologiques



Données médico-
administratives

La Plateforme E4N

- ▶ Questionnaires courts en ligne
- ▶ Possibilité de répondre d'un ordinateur, d'une tablette ou d'un smartphone
- ▶ Questionnaires envoyés par SMS
- ▶ Synchronisation avec des objets connectés Nokia et Viomed (et d'autres plus tard)

epiconcept

smart health





Les projets e-santé en cours



*Wearable devices are facilitators, not drivers,
of health behavior change*

Etude observationnelle sur l'utilisation des objets connectés en vie réelle



Financé par



- ▷ Equipement de 700 participants E4N avec le Nokia Go (démarrage mi-2018)
- ▷ Recueil d'information passif & en continue sur 1 an
- ▷ Synchronisation automatique des données avec la plateforme Web E4N
- ▷ Identification des freins & leviers à l'utilisation
- ▷ Collaborations avec l'ICAN et l'EHESP



Etude des déterminants d'un mauvais sommeil



Journal of Medical Internet Research ISSN 1438-8871
The leading peer-reviewed journal for health and healthcare in the Internet age.
Open Access • Top Cited (Impact Factor: 4.5) • Rapid Peer-Review • Medline Indexed (+20 other indices)

Determinants of a poor sleep based on data from 15 000 users of connected devices

Guy Fagherazzi, PhD^{1,2}; Douae El Fatouhi, MSc^{1,2}; Alice Bellicha^{3,4}, MSc; Amin El-Gareh^{1,2}, MSc; Aurélie Affret, MSc^{1,2}; Courtney Dow, MPH^{1,2}; PhD; Lidia Delrieu, MSc^{5,6}; Matthieu Vegreville, MSc⁷; Alexis Normand, MSc⁷; Jean-Michel Oppert, MD, PhD³; Gianluca Severi, PhD^{1,2}

Etude préliminaire, pour se familiariser avec le type de données

- ▷ Accès aux données des clients Withings (Nokia) entre le 01/01/2013 et 01/04/2016
- ▷ **15 839 individus (13 658 hommes & 2 181 femmes)**
- ▷ Modélisation du risque d'avoir un mauvais sommeil

Données disponibles

- ▷ Sommeil
 - Total
 - Profond
 - Ratio profond/total
- ▷ Age, Sexe, Poids, Nb de pas, Fréquence cardiaque, Tension

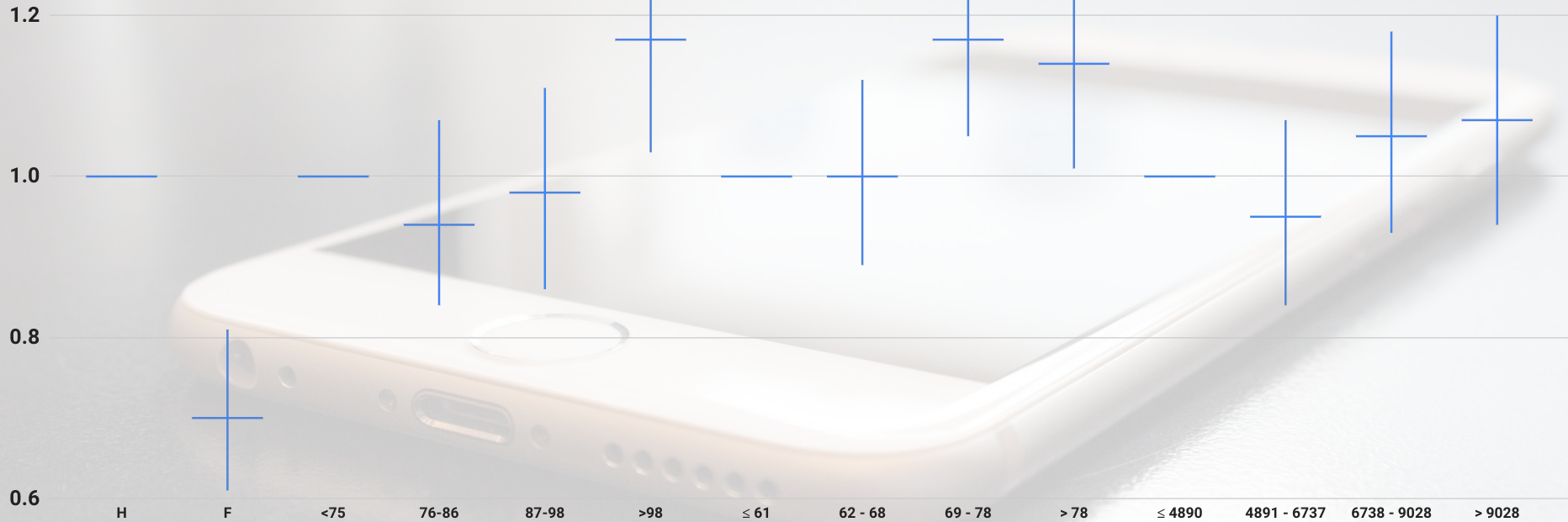
Déterminants d'un mauvais sommeil total (= durée <6h)

Sexe

**Poids
(kg)**

**Fréq cardiaque
(bpm)**

**Pas
(n/jr)**



De l'observation à l'intervention

Essai d'intervention randomisé niché dans l'étude E4N

Evaluation de l'impact de l'utilisation
d'objets et dispositifs médicaux
connectés en prévention primaire et
tertiaire du diabète de type 2



Projet financé par la FRM
+ 2 thèses CIFRE en collaboration avec Visiomed à
partir de 2018



Visiomed

De l'observation à l'intervention

Population générale (n=400)

Groupe intervention (n=200)

Recommandations type
consultation de prévention +
objets connectés (traqueur
d'AP/sommeil, balance)

Groupe contrôle (n=200)

Recommandations seules

Patients diabétiques (n=100)

Groupe intervention (n=50)

Recommandations type
consultation de suivi du diabète +
objets connectés (traqueur
d'AP/sommeil, balance,
glucomètre, tensiomètre)

Groupe contrôle (n=50)

Recommandations seules



Deux visites - 1 an d'intervalle

Questionnaires en ligne sur la plateforme
E4N tous les 3 mois sur QoL, stress, mode
de vie, santé perçue, test sur les
connaissances et l'adhésion aux
recommandations



Visiomed

A hand holding a smartphone. The screen shows a folder titled 'Social Media' containing icons for GroupMe, Facebook, Instagram, Reddit, YouTube, Twitter, and LinkedIn. The text 'Le potentiel des réseaux sociaux' is overlaid in white on a red rectangular background.

**Le potentiel des
réseaux sociaux**

**World Diabetes
Distress Study**

Estimated number of people with diabetes worldwide and per region in 2015 and 2040
(20-79 years)

North America and
Caribbean

2015 **44.3 million**
2040 **60.5 million**

Europe

2015 **59.8 million**
2040 **71.1 million**

Middle East and
North Africa

2015 **35.4 million**
2040 **72.1 million**

Western Pacific

2015 **153.2 million**
2040 **214.8 million**

South East
Asia

2015 **78.3 million**
2040 **140.2 million**

South and
Central America

2015 **29.6 million**
2040 **48.8 million**

Africa

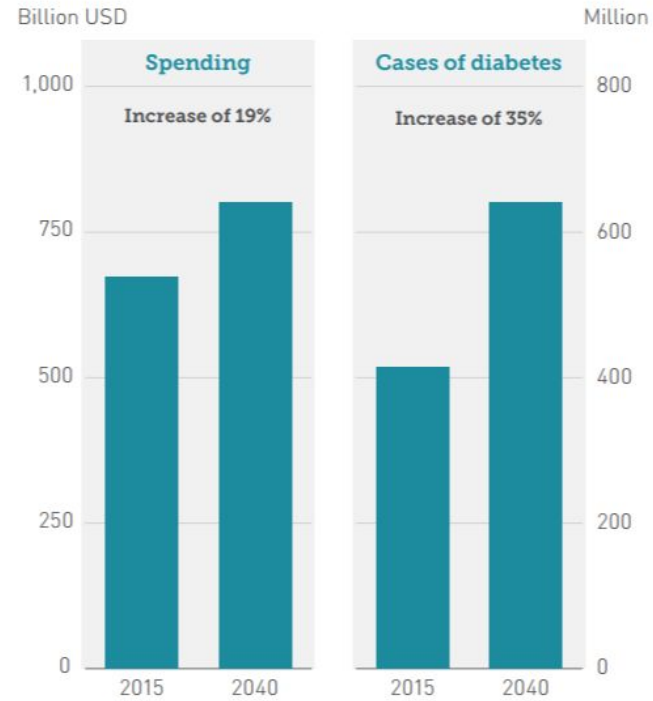
2015 **14.2 million**
2040 **34.2 million**

World

2015 **415 million**
2040 **642 million**



Global health spending to treat diabetes



Every **6 seconds** a person dies from diabetes (**5.0 million deaths**)



Poorly managed diabetes
leads to **serious
complications**
and early death

Watch now on Twitter

With good self-management and
**health professional
support**, people with diabetes can
live a long, **healthy life**

Détresse liée au diabète (DD)

DD = facteurs psychologiques (fardeau émotionnel, stress, anxiété) liés à la gestion du diabète au quotidien

- ▷ Alimentation
- ▷ Activité physique
- ▷ Traitements
- ▷ Crainte des hypoglycémies
- ▷ Peur des complications...

Associée à

- ▷ Mauvaise qualité de vie
- ▷ Risque de dépression
- ▷ Moins bon suivi médical et observance
- ▷ Déséquilibre glycémique
- ▷ HbA1c élevée
- ▷ Risque de complications

Détresse liée au diabète (DD)

DD = concept qui a émergé dans les années 1990

- ▷ **25% des adultes diabétiques anglais ont, à un moment donné dans la vie, ressenti une détresse liée à leur diabète**
- ▷ **Taux similaires ailleurs en Europe, Australie, USA**

La définition et l'évaluation de la DD sont hétérogènes (6 à 10 échelles existantes aujourd'hui)

- ▷ **DD souvent évaluée chez les adultes avec l'échelle PAID (Problem Areas in Diabetes)**
- ▷ **Plus récemment par l'échelle DDS (Diabetes Distress Scale)**

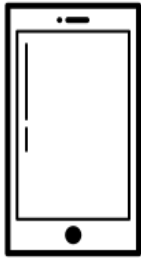
	Not a problem	Minor problem	Moderate problem	Somewhat serious problem	Serious problem
1. Not having clear and concrete goals for your diabetes care?	0	1	2	3	4
2. Feeling discouraged with your diabetes treatment plan?	0	1	2	3	4
3. Feeling scared when you think about living with diabetes?	0	1	2	3	4
4. Uncomfortable social situations related to your diabetes care (e.g., people telling you what to eat)?	0	1	2	3	4
5. Feelings of deprivation regarding food and meals?	0	1	2	3	4
6. Feeling depressed when you think about living with diabetes?	0	1	2	3	4
7. Not knowing if your mood or feelings are related to your diabetes?	0	1	2	3	4
8. Feeling overwhelmed by your diabetes?	0	1	2	3	4
9. Worrying about low blood sugar reactions?	0	1	2	3	4
10. Feeling angry when you think about living with diabetes?	0	1	2	3	4
11. Feeling constantly concerned about food and eating?	0	1	2	3	4
12. Worrying about the future and the possibility of serious complications?	0	1	2	3	4
13. Feelings of guilt or anxiety when you get off track with your diabetes management?	0	1	2	3	4
14. Not "accepting" your diabetes?	0	1	2	3	4
15. Feeling unsatisfied with your diabetes physician?	0	1	2	3	4
16. Feeling that diabetes is taking up too much of your mental and physical energy every day?	0	1	2	3	4
17. Feeling alone with your diabetes?	0	1	2	3	4
18. Feeling that your friends and family are not supportive of your diabetes management efforts?	0	1	2	3	4
19. Coping with complications of diabetes?	0	1	2	3	4
20. Feeling "burned out" by the constant effort needed to manage diabetes?	0	1	2	3	4

L'échelle PAID

Limites de l'échelle

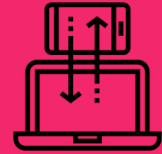
- ▷ Fermée
- ▷ Auto-rapportée
- ▷ Des traits de la DD sont manquants (problème au travail, coûts des traitements, relations avec les professionnels de santé)
- ▷ Interprétée avec le clinicien
 - Biais et effet "Make my doctor happy"
- ▷ Risque de déni

Le changement de paradigme



Utiliser le potentiel du **digitosome** des individus avec un diabète de type 1 ou type 2 pour étudier les déterminants de la détresse liée au diabète

Partout dans le monde !



Digitosome

= intégralité des données générées en ligne par un individu au cours de sa vie (smartphones, capteurs, dispositifs médicaux, réseaux sociaux, géolocalisation,...)

- “Utilisons Twitter !”
- “Mais... pourquoi Twitter ?”
- “Ben...parce que Twitter!”



Quelques infos

- ▷ 313 millions/mois d'utilisateurs actifs
- ▷ 15 % des utilisateurs envoient 85 % des Tweets
- ▷ > 4,7 millions de personnes tweetent
- ▷ Partout dans le monde
- ▷ Données publiques

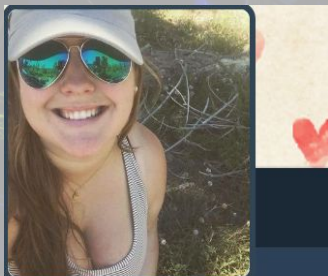
Les langues sur Twitter

- ▷ Anglais (34%)
- ▷ Japonais (16%)
- ▷ Espagnol (12%)
- ▷ Malais (8%)
- ▷ Portugais (6%)
- ▷ Arabe (6%)
- ▷ Français (2%)
- ▷ Turc (2%)



Le diabète sur Twitter

Données non disponibles dans
un contexte clinique



Kate

@aDoseOfDiabetes

26 years old. type 1 diabetic for 25 years. I like pretty things and nice people.

📍 California, USA

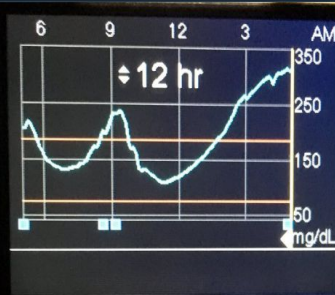


Kate

@aDoseOfDiabetes

Went to bed at 10....not sure why this keeps happening but I hate waking up at 4 feeling so sick

🌐 À l'origine en anglais



MeaT @TimeaTheTranny

Il y a 13 mins

Everytime i feel a minor ache in my foot i assume it's because of diabetes.



Type1bri @type1bri · 18 févr.

My latest flourey with the libre to check basal and ratios has been a success, no changes needed atm until next time #gbdoc #doc #ourd

🌐 À l'origine en anglais

🔄 JllYllyFish Retweeté



EnvyDaTropic™ @envydatropic

il y a 3 jours

If it doesn't clog arteries or cause diabetes I'm not eating it



152



203



Tasha @tasha_louise

Il y a 15 mins

Just because you can't SEE my conditions, does not mean they're not there #diabetes #mentalhealth #longstandingillness



Splash_Bro_3 Retweeté



Bo @bohunter15

Il y a 40 mins

Happy birthday bro! @Splash_Bro_3 🎉 have a good one and don't let your diabetes get in the way 🙏





Le diabète sur Twitter

Tirer le plein bénéfice de
l'effet "communauté"
DT1 et DT2 en ligne

**1^{ère} phase : étude des tweets -> profils de
détresse liée au diabète**

- ▷ Anglais
- ▷ Français
- ▷ Espagnol

Extension à d'autres langues ensuite

2^{ème} phase : intégration d'un chatbot

- ▷ Moteur de recrutement de participants via les réseaux sociaux
- ▷ E-cohorte mondiale
- ▷ Collecte de données contextuelles sur la maladie

3^{ème} phase : plateforme Web

- ▷ Appariement à des objets et DM connectés
- ▷ Appariement avec bases de données existantes via la géolocalisation des Tweets



**Données brutes
Twitter**

**Nettoyage,
préparation, filtrage
et analyses des
données (machine
learning, data/text
mining)**

**Information à
valeur ajoutée,
utilisable à des fins
de recherche**

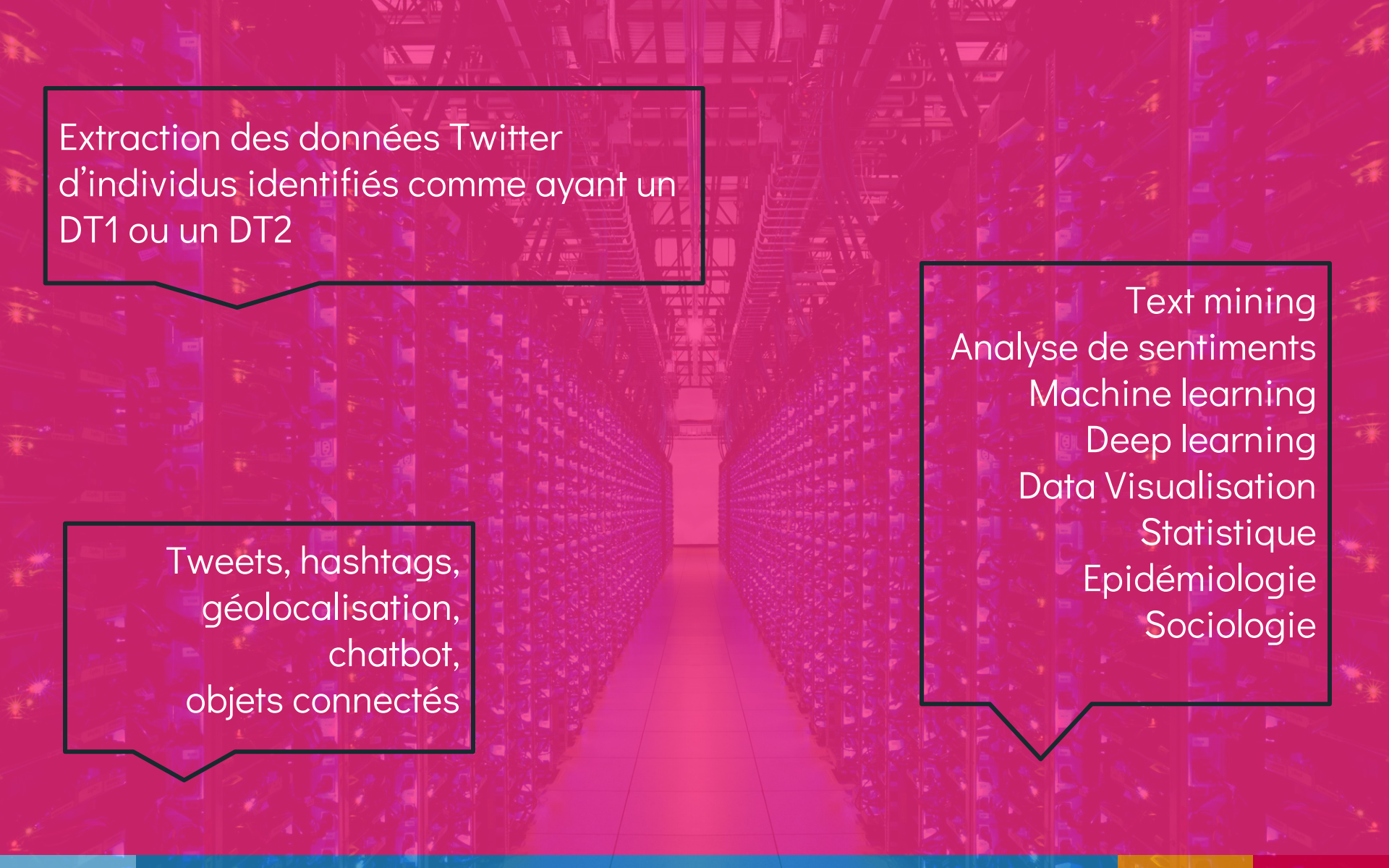
Processus





BIG DATA

Health Crowd Sourcing
au service de la
recherche sur le diabète



Extraction des données Twitter
d'individus identifiés comme ayant un
DT1 ou un DT2

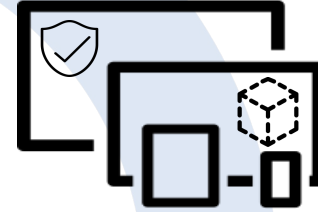
Tweets, hashtags,
géolocalisation,
chatbot,
objets connectés

Text mining
Analyse de sentiments
Machine learning
Deep learning
Data Visualisation
Statistique
Epidémiologie
Sociologie

Aller au-delà des données Twitter

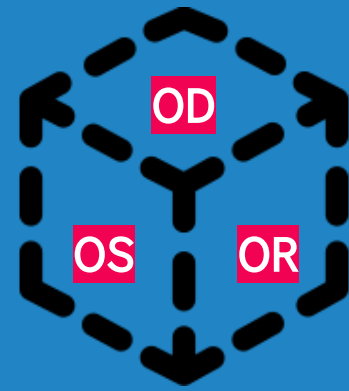


Echange avec “Diabot”,
un chatbot qui va
interagir avec les
volontaires DT1 ou
DT2 pour collecter de
l’information sur la
pathologie



Participants volontaires
vont synchroniser leur
objets et DM connectés
sur une plateforme web
sécurisée (activité,
sommeil, tension,
CGM...) -> **Open Data**

Une plateforme “Open” pour la recherche sur le diabète



Open Data

Bases de données

Open Source

Moteur d'extraction de
Tweets sur une
pathologie

Chatbot de suivi de
patients

Algorithmes text-mining
et machine learning

Méthodologie

Open Research

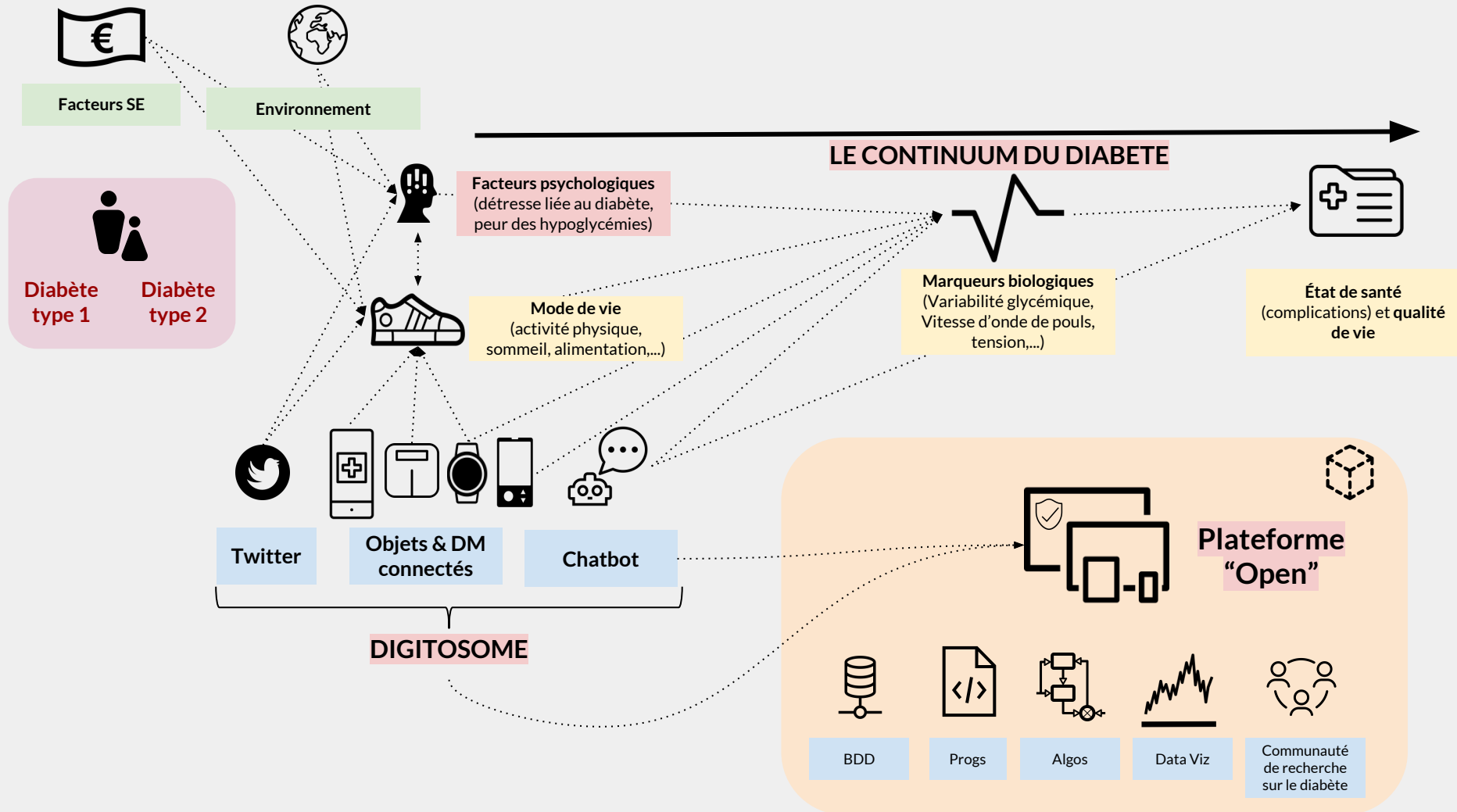
Création d'une
communauté de
recherche mondiale
autour du diabète

Hackathon

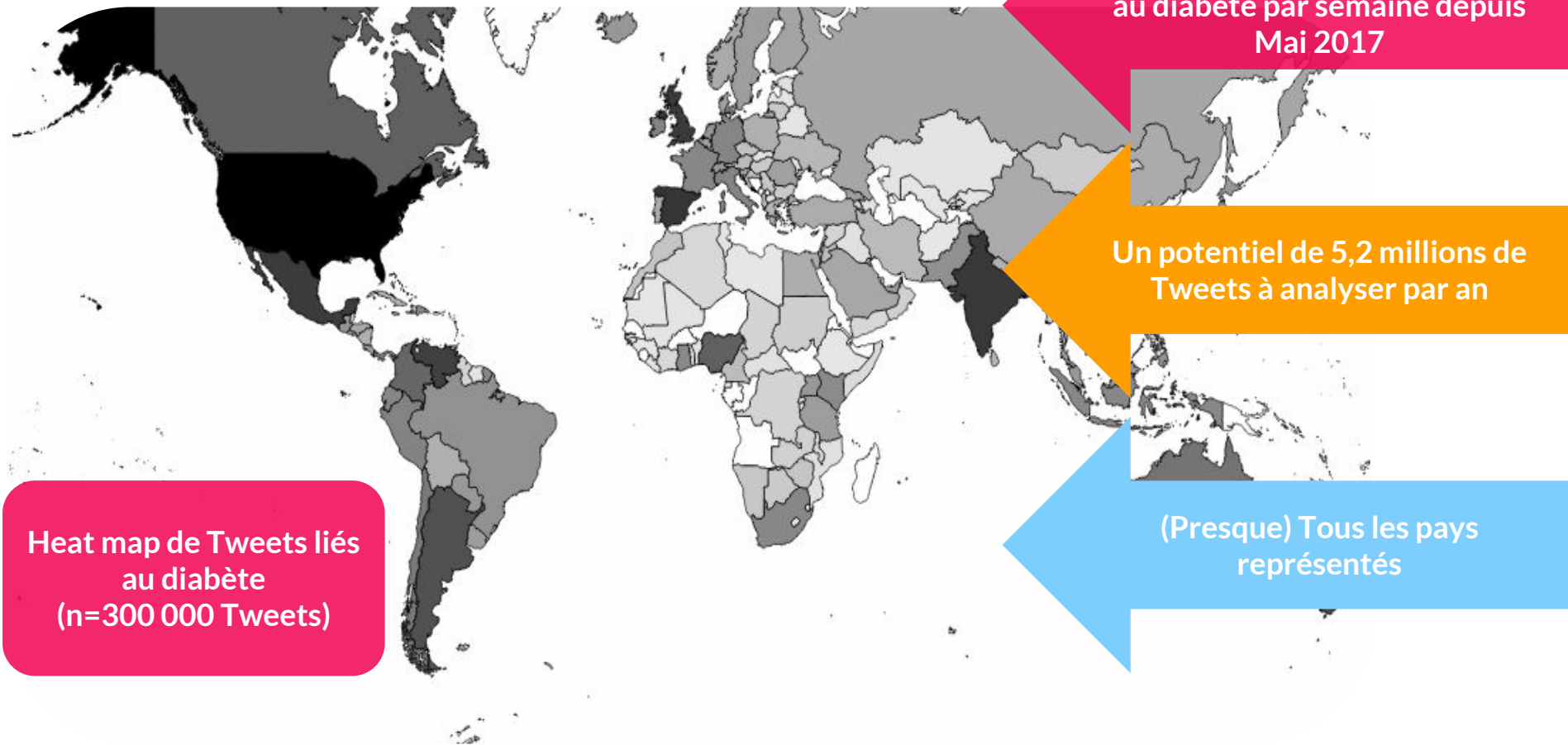
Transparence &
recherche reproductible

Publications en Open
access

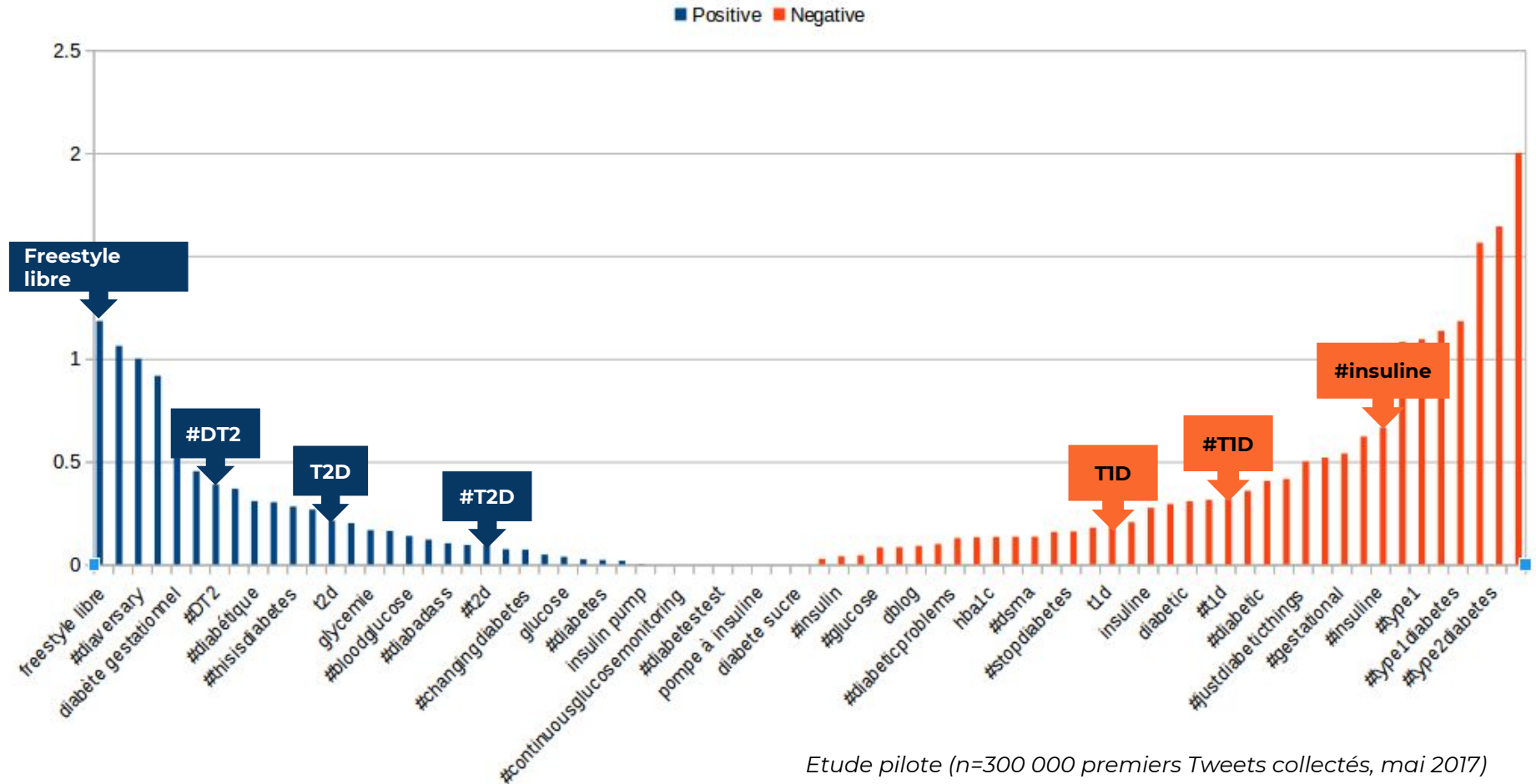
Schéma conceptuel



Etude pilote



Analyse de sentiments sur les hashtags fréquents





**Mieux
comprendre la
détresse liée au
diabète**

**Améliorer le suivi
de la maladie et la
qualité de vie**

**Détecter des
signaux faibles à
partir de Big Data**

**Générer des
nouvelles
hypothèses de
recherche**

En bref

1^{ère} étude
mondiale sur la
santé
psychologique
des patients
diabétiques

Une source
d'information
sans précédent

Chatbot &
méthodologie
transférables à
d'autres
pathologies
chroniques

Développement
de la recherche
médicale de
demain



Le mot de la fin

Vers un “omique digital”

Génome
Epigénome
Protéome
Transcriptome
Métabolome
Microbiome
et...

LE “DIGITOSOME”

= l'intégralité des données digitales générées par un individu au cours de sa vie (smartphones, capteurs, dispositifs médicaux, réseaux sociaux, géolocalisation,...)

Mieux caractériser (i.e. de manière objective et continue) :

- ▷ Le mode de vie (facteurs modifiables : alimentation, activité physique, sommeil, poids) & facteurs psychologiques
- ▷ “Contextualiser l'individu” (géolocalisation, données environnementales...)
- ▷ Des biomarqueurs d'intérêt (surveillance, prédiction...)

Les technologies de demain en recherche médicale



REALITE VIRTUELLE/AUGMENTEE

SERIOUS GAMES

CHATBOTS

- Collecte d'information ludique/passive, éducation thérapeutique, littératie, prévention

TEXT-MINING

- Extraire de l'information pertinente des réseaux sociaux, comptes-rendus médicaux...

INTELLIGENCE ARTIFICIELLE

- Machine/deep learning (prédiction de risque)
- Natural Language Processing (NLP)

Favoriser l'Open (Research) Data !

Si la génétique a fait autant de progrès ces dernières années, c'est en grande partie grâce au partage de données.



Inciter au financement prioritaire de projets de recherche
Open Data / Open Source

Algorithmes (classification de pathologies sur les données du SNIIR-AM,
analyse statistique de données d'objets connectés, prédiction de risque,
text mining)

Exemples : REDSIAM, Epidemium

Favoriser la reproductibilité des résultats de recherche



Merci ! Questions ?
Sur Twitter : [@GFaghe](https://twitter.com/GFaghe)
ou guy.fagherazzi@gmail.com

e3n.fr

e4n.fr

