## Fonctionnement et défis

ď



Séminaire ALIMining 29-30 septembre 2025 Toulouse





Présentation du projet

Usage dans la recherche



3

La collecte des données

**Défis** 





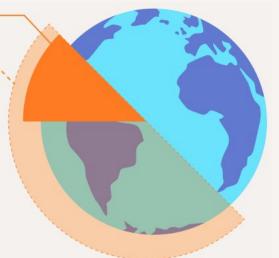
## PRÉSENTATION DU PROJET

## L'alimentation : une priorité pour la santé et l'environnement

Dans le monde, plus d'**un milliard de personnes** sont considérées comme obèses, et dans l'UE, un enfant sur trois l'est, <u>selon l'OMS</u>.

Et la mauvaise nouvelle est que ce phénomène est en hausse : selon la World Obesity Foundation, d'ici 2035, plus de 4 milliards de personnes, soit plus de la moitié de la population mondiale, seront obèses.

De nombreux consommateurs se retrouvent **dépassés** par la quantité d'informations que l'on trouve sur les emballages alimentaires, peinant parfois à lire la longue liste d'ingrédients en petits caractères, à déchiffrer de mystérieux additifs alimentaires ou à différencier les étiquettes des aliments... Pourtant, il existe une réelle volonté de faire de meilleurs choix pour notre santé et celle de notre planète, notamment en matière d'alimentation.



#### Food accounts for...

1/3

of greenhouse gas emissions Source: i4CE - 2019

70%

of drinking water consumption  $\underline{\texttt{Source} : \texttt{OECD}}$ 

#### The majority

of global packaging waste Source: Article by Nemat et all



#### Qui sommes-nous?

Pionniers de la transparence alimentaire depuis **2012**, Open Food Facts est une **communauté** construisant une base de données collaborative et ouverte de produits alimentaires.

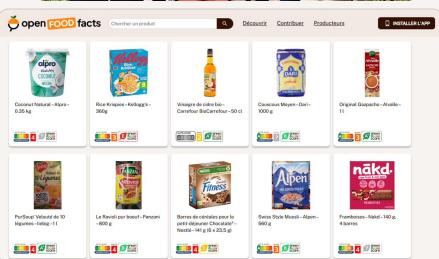
Nous recueillons des données sur l'impact **nutritionnel** et **environnemental** des produits alimentaires.

Association à but non lucratif, indépendante de l'industrie alimentaire, reconnue en tant que <u>bien public numérique des</u> Nations Unies.

Axé sur la communauté : contributeurs, scientifiques, applications mobiles, producteurs et institutions alimentaires.

Q Open source et open data





#### Notre mission

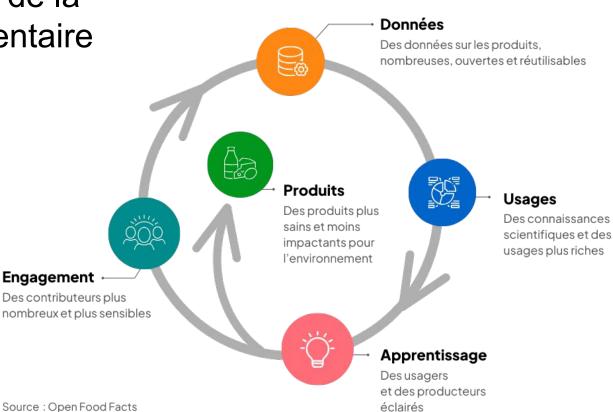
Donner à chacun les moyens d'agir grâce à la donnée : permettre aux consommateurs, aux producteurs, aux chercheurs, aux gouvernements... de mesurer et changer à grande échelle l'impact de l'alimentation sur notre Santé et sur la planète



Le cercle vertueux de la transparence alimentaire

Le pourquoi derrière nos actions





## Notre impact en quelques chiffres





produits du monde entier



producteurs engagés partageant des informations sur les produits



langues disponibles



articles scientifiques citant Open Food Facts



contributeurs dans le monde



visiteurs uniques mensuels

# USAGE DANS LA RECHERCHE

## De quelles données parle-t-on?

#### Données brutes : tout ce que vous pouvez voir sur l'emballage du produit !





ton. Luontaisesti laktoositon. • INGREDIENTES: agua, leche de coco (27%) [crema de coco, agua], agua de coco (27%), piña [7%], azúcar, almidón modificado, correctores de acidez (ácido citrico, citratos de sodio), estabilizante [pectinas], aroma natural, extracto de zanahorias, citrato tricálcico, fermentos (S. thermophilus, L. bulgaricus). Puede contener trazas de soia. Sin lácteos y sin gluten. Naturalmente sin lactosa. • INGREDIENTES: água, leite de coco (27%) [creme de

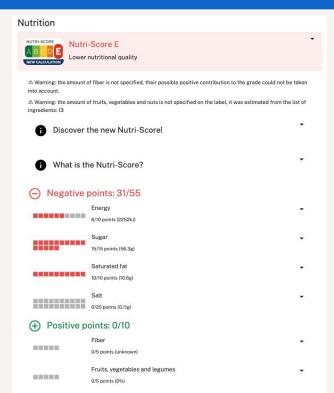


- Code-barres (EAN-13, EAN-8, UPC, ...)
- Nom du produit
- Quantité
- Catégorie
- Poids
- Marque
- Labels et certifications
- Pays et magasins où il est vendu
- Prix!
- Valeurs nutritionnelles
- Listes d'ingrédients
- Portion
- Informations sur l'emballage
- Origine des ingrédients
- ... et + de 150 autres champs



## De quelles données parle-t-on?

#### Mais aussi des informations calculées!

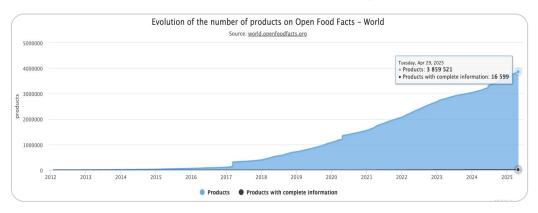


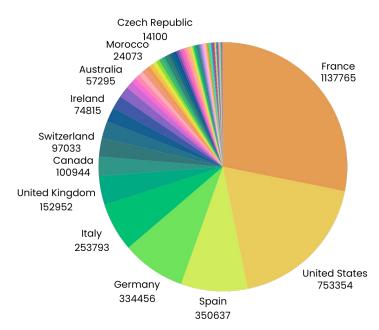
- Nutri-Score: qualité nutritionnelle (officiellement adopté dans 7 pays)
- **Green-Score**: impact environnemental des aliments
- Groupe NOVA: niveau d'ultra-transformation des aliments
- Analyse de la liste des ingrédients :
  - Extraction des additifs
  - Végétarien
  - Huile de palme
- Estimation du % d'ingrédients
- Risques de surexposition à certains additifs (rapports de l'EFSA)
- Messages éducatifs en contexte
- Normalisation de plusieurs champs (à comparer), tout en conservant les données brutes



## De quelle quantité de données parlons-nous ?

- 3855901 produits
- ±1500 nouveaux produits chaque jour
- Scores:
  - 1,3 M produits avec un Nutri-Score
  - 925 000 produits avec un Green-Score
  - o 981000 produits avec un NOVA
- https://world.openfoodfacts.org/data





Nombre de produits par pays

...et plus de 8 millions de photos pour alimenter l'IA (article)

# Open Food Facts Images on AWS Open Dataset: The Ultimate Food Image Dataset

Open Food Facts Images on AWS Open Dataset: The Ultimate Food Image Dataset



## Faire progresser la science grâce à l'open data

Influence d'Open Food Facts dans l' étude	# d' études
Essentiel	17
Une source importante	21
L'une des sources	13
Une source ou une utilisation mineure	9
Mentionné	5
Mentionné mais son utilisation est encore inconnue (non analysé ou article derrière un paywall)	400+



Dréano-Trécant, L., Egnell, M., Hercberg, S., Galan, P., Soudon, J., Fialon, M., ... Julia, C. (2020). Performance of the front-of-pack nutrition label nutri-score to discriminate the nutritional quality of foods products: a comparative study across 8 european countries. *Nutrients*, 12(5), 1303. https://doi.org/10.3390/nu12051303

... A scientific report has also suggested a good discriminating performance of the Nutri-Score for the food offered in Belgium, Italy, Netherlands, Spain, Sweden, Switzerland and United Kingdom, using the data from the Open Food Facts database, a collaborative online project gathering food composition data on manufactured foods from many countries worldwide [36]. ...

... These findings are complementary to studies conducted in other food databases such as Open Food Facts providing data on multiple industrial foods of different brands. ...





Monteiro, C. A., Cannon, G., Levy, R. B., Moubarac, J.-C., Louzada, M. L., Rauber, F., ... Jaime, P. C. (2019). Ultra-processed foods: what they are and how to identify them. *Public Health Nutrition*, 22(5), 936–941. https://doi.org/10.1017/S1368980018003762

... A cell phone app created by the non-profit organization Open Food Facts, based in France, already enables consumers to identify among more than 145 000 packaged products the more than 75 000 that are ultra-processed. ...



## Autres projets de recherche notables utilisant OFF





EREN Nutri-Net Santé, 2009-maintenant

Étude des habitudes de consommation et de la santé d'une cohorte de plus de 170 000 participants sur plus de 10 ans. En croisant les données de consommation détaillées avec les informations d'Open Food Facts sur les ingrédients, les additifs et le contenu nutritionnel, les chercheurs peuvent évaluer plus précisément les impacts à long terme sur la santé de divers composants alimentaires, y compris les additifs et les aliments ultra-transformés.

Lire aussi: Consumption of certain emulsifying food additives is associated with an increased risk of cancer (Feb 2024)

Menichetti Lab, 2023

#### Machine learning prediction of the degree of food processing

La professeure Giulia Menichetti a utilisé OFF pour former un modèle d'apprentissage automatique capable de prédire le degré de transformation d'un aliment, en fonction de son contenu nutritionnel et de ses additifs. Ce modèle, appelé FoodProX, attribue à chaque produit une note allant de 0 (non transformé) à 1 (ultra-transformé). OFF a permis de tester le modèle sur plus de 230 000 produits.







Iris Van Dam & Stefanie Vandevijvere (2022)

Benchmarking the nutrition-related commitments and practices of major French food companies

https://link.springer.com/article/10.1186/s12889-022-13780-y

... For packaged food and non-alcoholic beverage manufacturers and supermarkets (own-brand products), the healthiness of the complete product portfolios was analysed using Open Food Facts data for France in 2018...

Au 16 mai 2025, Google Scholar retourne:

866 résultats en cherchant « Open Food Facts » :

https://scholar.google.com/scholar?g=%22open+food+facts%22

liday, Anna Leibinger, Oliver Huizinga, Carmen Klinger, Elochukwu Okanmelu, ert, Eva Rehfuess, Peter von Philipsborn (2023) plication of the WHO Nutrient Profile Model to products on the German market:

plications for proposed new food marketing legislation in Germany

os://www.medrxiv.org/content/10.1101/2023.04.24.23288785v1

Methods. We applied the WHO NPM to a random sample of 660 food and verage products across 22 product categories on the German market drawn from en Food Facts, a publicly available product database. ... Lino Galiana, Milena Suarez Castillo (2022)

Fuzzy matching on big-data: An illustration with scanner data and crowdsourced nutritional data

https://www.linogaliana.fr/pdf/JMS2022/S28\_2\_ACTE\_GALIANA\_JMS2022.pdf

... In this paper, we enrich a large retailer dataset with nutritional information extracted from Open Food Facts, completed with the ANSES Ciqual dataset...

Gero Laurenz Höhn, Martijn Huysmans Healthy Food Traditions? Nutritic Geographical Indications https://lirias.kuleuven.be/retrieve

... Thus, we base our quantitative database of OpenFood Facts(ht to thousands of observations income 10 countries...

tprint of food products from packaging data. joss.03329.pdf

n interfaced to the Open Food Facts

ng packaging information...

s contributors for creating the database this

Rodríguez-Martín, N.M.; Córdoba, P.; Sarriá, B.; Verardo, V.; Pedroche, J.; Alcalá-Santiago, Á.; García-Villanova, B.; Molina-Montes, E. (2023)

Characterizing Meat- and Milk/Dairy-like Vegetarian Foods and Their Counterparts Based on Nutrient Profiling and Food Labels. *Foods* **2023**, *12*, 1151. https://doi.org/10.3390/foods12061151

... Nutritional data of the food products were obtained from websites of supermarket and food companies, and from Open Food Facts. The latter was the main information source used to retrieve the food products. ...

Rakhmawati, N. A., Firmansyah, A. A., Effendi, P. M., Abdillah, R., Cahyonc Auto halal detection products based on <u>euclidian</u> distance and cosine simile Journal on Advanced Science, Engineering and Information Technology, 8( <a href="https://doi.org/10.18517/ijaseit.8.4-2.7083">https://doi.org/10.18517/ijaseit.8.4-2.7083</a>

In this study, we build a system that can compare products that have certified halal with halal certified products based on its ingredients. T are collected from Open Food Facts, Institute For Foods, Drugs, An Indonesian Council Of Ulama (LPPOM MUI) and our halal system...

... Open Food Facts (https://world.openfoodfacts.org/), a crowdsour database system provides some of the products that are labelled as

### Un intérêt croissant dans le monde entier





# LA COLLECTE DES DONNÉES

#### Comment sont collectées les données?



#### Utilisateurs du site

https://world.openfoodfacts.org

#### Base de données publique

- Recherches produits
- Contributions de données et d'images: modifications, actualisations, nouveaux ajouts



#### Utilisateurs de l'appli

Disponible pour iOS et Android

#### L'appli mobile de scan d'Open **Food Facts**

- Le scan pour des recherches et ajouts rapides
- Personnalisable. personnalisée et plus flexible



#### Open Food Facts

https://world.openfoodfacts.org

#### Base de données publique

- +3,8 millions de produits
- +200 réutilisateurs
  - Autres applis
  - Groupes de recherche



#### Bases de données externes

ex. USDA

Bases de données alimentaires publiques



#### Plateforme Pro

https://world.pro.openfoodfacts.org

#### Espace dédié pour chaque producteur

- Import de données et images
- Gestion Produit
- Analyses de données et rapports

#### **Autre applis**

ex. Elcoco, Smantry, ethic advizor, food less, waistline ...

#### De nombreuses applis réutilisent les données d'Open Food Facts

Partage de nouvelles données ajoutées selon les règles d'Open Database License (ODbL)

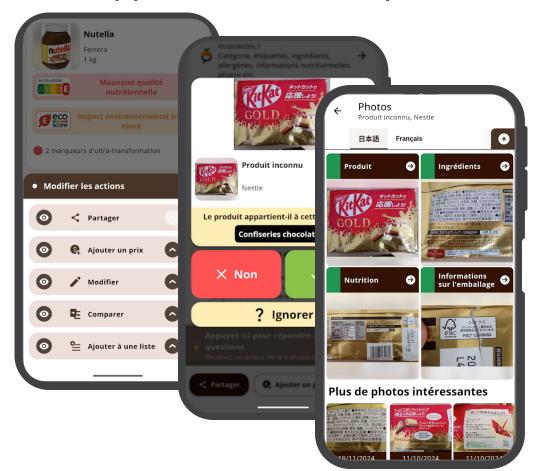


## Une application mobile pour engager le grand public



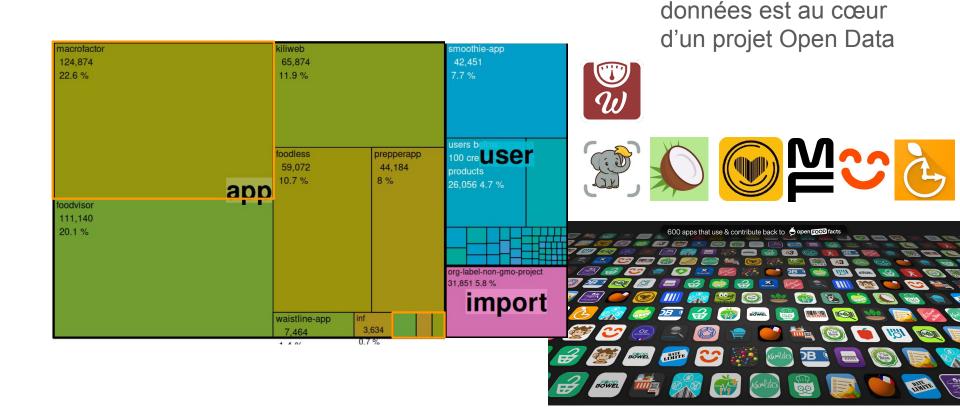
- Un effort sur l'interface
- Nutri-Score, Nova et Green-Score
  - Score personnalisé (respectant la vie privée)

## Une application mobile qui facilite la collecte



- Priorité à la collecte des images
- Encouragements à la participation (scores)
- Validation des données IA
- Un focus sur la qualité

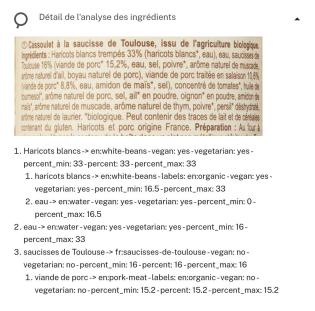
#### Des ré-utilisateurs très actifs



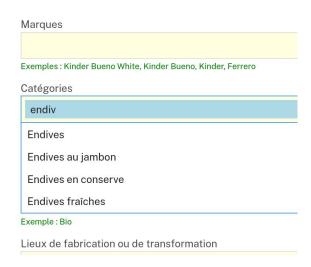
La mutualisation des

#### Flexibilité sur la donnée utilisateur

Un principe: on laisse le contributeur saisir ce qu'il veut, on structure à postériori. On guide toutefois l'utilisateur



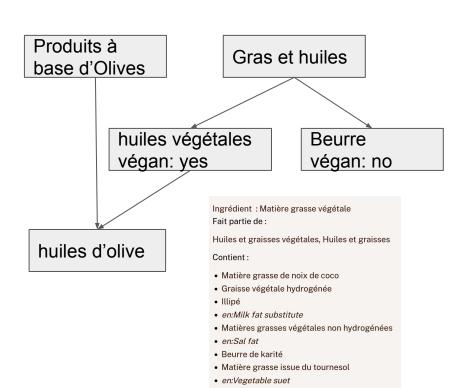
Effort conséquent pour analyser la donnée brute



Guider l'utilisateur

## Taxonomy - Structurer à posteriori

Les taxonomies donnent une vision structurée de la connaissance et permettent de raffiner la donnée brute.





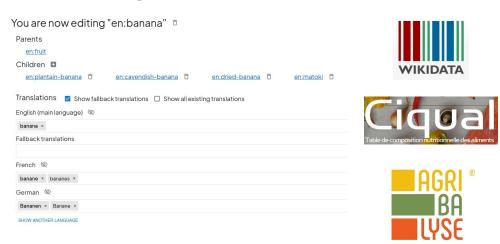
### Taxonomy - une construction collaborative

Les taxonomies sont construites par la communauté.

Elles se fondent sur une observation des usages et une recherche spécifique.







traduction, synonymes, caractéristiques, etc.

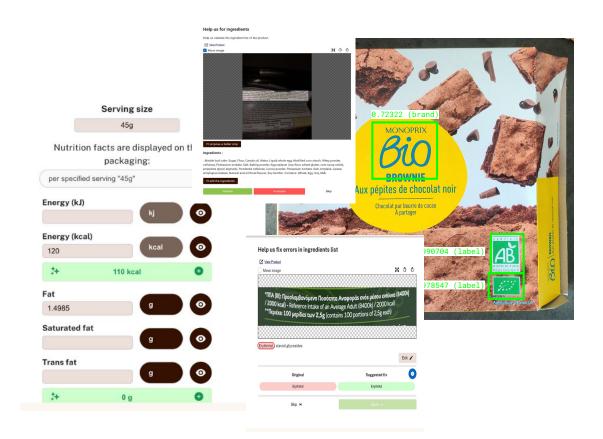
Liens à d'autres bases de données

## Utilisation l'intelligence artificielle pour amplifier la collecte

Robotoff utilise des heuristiques et de l'apprentissage automatique pour trouver de nouvelles données, au moindre doute il y a validation humaine



## Des modèles avancés, tous Open Source



- Nutrisight: détection de la nutrition
- Détection des ingrédients
- Corrections des ingrédients
- Détection des logos et classifications (labels, marques, scores, etc)
- Données, modèles, etc. disponibles

#### **Modération**



#### Signaler un problème



Informations incomplètes ou incorrectes?

Catégorie, étiquettes, ingrédients, allergènes, informatic



ℚ Zoom

Code-barres:

3175680011480

Signaler cette image à nos modérateurs

Nutripatrol: Outil collaboratif de Sésan traitement des signalements (images, données)

Si les informations ne correspondent pas à celles figurant sur l'emballage, vous pouvez les compléter ou les corriger. Merci! Open Food Facts est une base de données collaborative, et chaque contribution est utile pour tous.

Compléter ou corriger les informations du produit

Si vous souhaite vous ne pouvez

Signaler ce pro



2025-09-25 / 08:45

human

Robotoff detection: face



### Amplifier et améliorer les données?

Pour améliorer continuellement la qualité des données, nous nous appuyons sur 4 piliers :

#### Communauté

 Ajoute constamment des données et photos
 une véritable
 patrouille de réviseurs

#### "Règles du métier"

mises en place pour identifier les erreurs sur les fiches produits. *Par exemple*, si la somme des poids des ingrédients est supérieure au poids total du produit, il y a une anomalie.

#### **Machine Learning**

limite le risque d'erreurs de saisie. Les votes sont éventuellement consolidés pour les utilisateurs anonymes

#### Plateforme Pro

cette plateforme permet aux producteurs de télécharger leurs données produites en masse et ainsi de corriger/compléter les contributions de la communauté.

## Data quality team

pwe\_pct

2023

April

 A continuous community effort to fix quality errors (3 errors / day)

2024

April

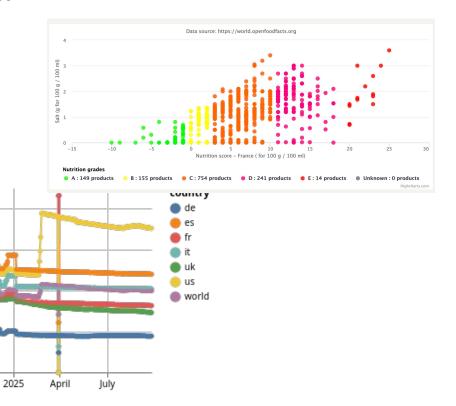
date

July

October

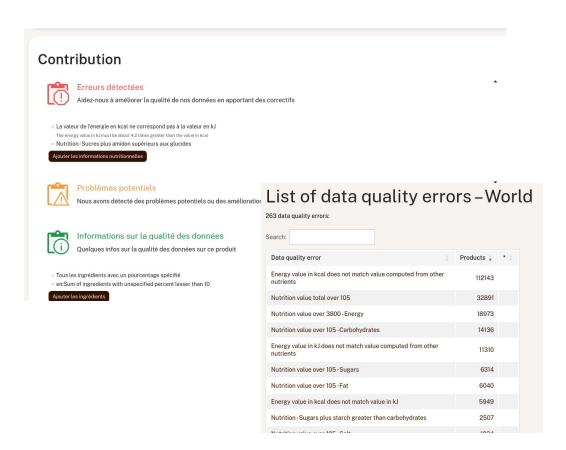
October

Specific analysis by volunteers



## Data quality indicators

- A lot of indicators on data quality
  - visible on the website
  - easy to filter
- Very useful for academic research (filter out data with errors)



## 4 DÉFIS

## La compréhension des données sur l'emballage

Certains éléments restent difficile à collecter ou analyser

- reflets, cylindres, etc.
- parsing des ingrédients: une tâche ardue et très dépendante de la langue
- limites réglementaires (ex: "flavoring" est permis, peu d'information, pas d'origines ou de proportion des ingrédients, etc.)
- poids des emballages
- Certaines données peuvent éventuellement être déduites (exemple: notre projet "recipe estimator")
- Les LLMs ne constituent pas aujourd'hui une solution du fait de leur niveau de prix, mais surtout de la vérifiabilité et de la fiabilité des résultats.

La solution idéale: quelques photos de l'emballage pour avoir toutes les données produit... un rêve qui se réalise un peu plus chaque jour !

Aidez nous! Il y a matière à des projets de master ou de thèse

## Étendre les usages (quelques exemples)

**Au delà du code barre**: Historiquement le code barre offre un mapping très simple entre le produit dans le monde réel et ses données.

Mais il ne fonctionne pas pour tout: produits en vrac, produits bruts, produits non alimentaires, etc.

L'identification par l'utilisateur dans une nomenclature n'est pas une tâche simple. L'identification par l'IA peut-

être: peu précise, manquer de données (ex:

provenance), pas toujours fiable.

Les prix: le projet Open Prices est le projet fou de notre communauté pour créer une base de données open data sur les prix qui soit significative pour la recherche et certains usages.





115,000 prices 3,700 locations

40,000 proofs 1,800 contributors

## L'évolution des produits au cours du temps

Open Food Facts maintient l'ensemble des versions des produits dans sa base.

Ces données pourraient servir de base à l'observation de l'évolution des produits dans le temps (impact de mesures publique comme le Nutri-Score sur les reformulations, shrinkflation, usage des additifs, etc.)

#### Ceci demanderait:

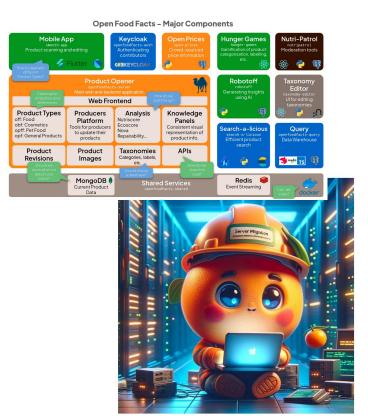
- des capacité de traitement plus importantes (puissance de calcul, indexes)
- des algorithmes pour différencier mises à jour des données (corrections, complétion), des reformulations/transformations des produits
- des algorithmes d'identification de certaines transformations

#### Maintien du commun

La maintenance du code et des nombreux composants demande de la compétence et du temps

L'infrastructure est également mise à l'épreuve par le succès du projet et l'intensification du crawling

Le maintien de la complétude, la fraîcheur et la qualité des données demande une communauté vivante





Open Food Facts Temporarily Unavailable

## Être plus pertinent pour la recherche

Nos taxonomies sont faites pour se lier à d'autres référentiels, mais ce lien reste parfois à faire, par exemple les liens avec FoodEx2 ou FoodOn.

Il y a toujours des nouvelles pistes de recherches qui nécessitent plus de données, par exemple quel élément de l'emballage a un contact alimentaire (contamination par l'environnement).

Les exports de données et l'API sont à améliorer pour être plus accessibles. Un projet est en cours pour simplifier les données des images et de la nutrition.

Notre communauté est toujours motivée pour aider les scientifiques.

## Expansion à l'international

L'expansion à l'international passe aujourd'hui beaucoup par l'application, les applications tierces, etc.

Les volontaires du projet viennent de nombreux pays.

Des entités locales seraient nécessaires pour rendre le projet plus robuste, divers et mieux adapté culturellement (médiation culturelle). Nous avons initié un programme d'ambassadeurs (actuellement: Canada, Maroc, Belgique, Allemagne, Danemark)

Il peut être un fabuleux support d'une information publique de qualité et d'aide au changement

Aidez nous! Avez vous des contacts intéressés dans un pays donné?

#### Les financements

































Open Food Facts est un projet qui veut rester gratuit. Mais il a besoin d'un noyau de salariés pour assurer la maintenance technique, l'animation de la communauté, les partenariats, etc.

Nous nous appuyons sur plusieurs sources de financements:

- des fondation privées
- des entitées étatiques
- des dons directs
- des projets académiques
- des projets européens (ex: Foodture avec l'ADEME et l'INRAE)

Aidez nous! Soyons partenaires sur des projets de recherche

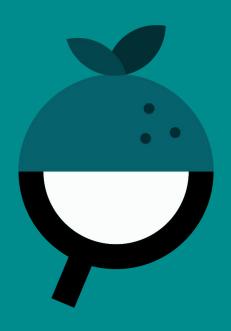
## Fédérer les usages

Open Food Facts est un projet communautaire, basé sur le partage de connaissances.

Nous aimerions que scientifiques et communauté travaillent main dans la main pour partager des projets, des bonnes pratiques, des outils, des résultats de leurs recherche.

Exemple: Mobilisation de la communauté pour peser les emballages, défis du mois sur Open Prices qui visent certaines catégories de produits, volontaires travaillant sur l'estimation des recettes, etc.

Aidez nous! Venez découvrir le projet "Open Food Facts" for science lors de l'atelier.





contact@openfoodfacts.org

