



HAL
open science

HistorIRIS : table de passage des IRIS de 1999 à 2022

Lucie Adélaïde, Morgane Stempfelet, Coline Babut, Alexandre Bulté,
Bérengère Ehrhart, Nadège Jaccy

► To cite this version:

Lucie Adélaïde, Morgane Stempfelet, Coline Babut, Alexandre Bulté, Bérengère Ehrhart, et al.. HistorIRIS : table de passage des IRIS de 1999 à 2022 : Data paper du Challenge Green Data for Health. Commissariat Général au Développement durable (CGDD). 2023, pp.gd4h.ecologie.gouv.fr/defis. hal-04176155

HAL Id: hal-04176155

<https://hal.science/hal-04176155>

Submitted on 2 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open licence - etalab

Green Data for Health

Data papers du Challenge Green Data for Health

HistorIRIS : table de passage des IRIS de 1999 à 2022

Août 2023

Mots clés

Référentiel, IRIS, Communes, découpage infracommunal, Insee

Auteurs

Lucie Adélaïde¹, Morgane Stempfelet¹, Coline Babut², Alexandre Bulté²,
Bérengère Ehrhart², Nadège Jaccy²

¹ Direction Santé Environnement Travail, Santé publique France

² Bénévole du Data Challenge du GD4H

SOMMAIRE

- Résumé
- Éléments de contexte et motivation
- Description des données
- Intérêt des données
- Collecte et traitement des données
- Plus-value et perspectives
- Bibliographie

Résumé

Les IRIS (Îlots Regroupés pour l'Information Statistique) sont des découpages géographiques infracommunaux utilisés par l'Insee pour la diffusion de données statistiques en France. Leurs contours évoluant au fil des années, une table de passage des IRIS de 1999 à 2022 a été produite pour permettre de combiner différents indicateurs d'intérêt (environnementaux, sanitaires, socio-économiques...) produits à l'échelle des IRIS dans des millésimes différents. Elle facilite le travail préalable de préparation et d'harmonisation des données dans un seul et même référentiel en vue d'analyses.

Éléments de contexte et motivation

Contexte

Les IRIS sont des découpages géographiques utilisés par l'Insee pour la diffusion de données statistiques en France. Ils correspondent à une échelle infra-communale utilisée pour produire des statistiques locales.

L'évolution des contours des IRIS chaque année depuis 1999 nécessite actuellement une conversion des millésimes des différents indicateurs environnementaux, sanitaires et socio-économiques disponibles pour passer d'un référentiel géographique donné à un autre.

Dans ce contexte, une table de passage des IRIS de 1999 à 2022 a été produite dans le cadre du Data Challenge du [Green Data for Health](#) (GD4H). Le GD4H est une offre de service incubée au sein de l'ECOLAB, laboratoire d'innovation pour la transition écologique du Commissariat général au développement durable. L'objectif du Data Challenge était de développer des outils ancrés dans la communauté de la donnée en santé-environnement afin d'adresser des problématiques partagées.

Motivation

De nombreux indicateurs (environnementaux, sanitaires, socio-économiques...) sont produits et étudiés à fine échelle (IRIS) pour rendre compte au mieux de l'exposition de la population et étudier leurs effets sur la santé. Ces analyses nécessitent une étape préalable de conversion des millésimes d'IRIS des indicateurs étudiés dans un unique millésime d'IRIS afin de les combiner.

La table de passage des IRIS d'un millésime à un autre n'existait pas, là où une table de passage des communes existait.

Le Challenge GD4H a offert la possibilité de créer une table de passage des IRIS depuis 1999 qui permet de relier les IRIS d'une période à ceux d'une autre période, en identifiant les fusions et les scissions entre deux périodes. Cette table facilite les croisements de données de périodes différentes en permettant de comparer des données statistiques sur des zones géographiques qui ont évolué au fil du temps.

L'objectif de ce data paper est de décrire ces données et leur intérêt scientifique afin d'en susciter des réutilisations, et de détailler les conditions techniques de cette réutilisation.

Description des données

L'ensemble des ressources issues du défi est présent sur GitLab :

<https://gitlab.com/data-challenge-gd4h/historiris>.

Une table de passage synthétique des IRIS de 1999 à 2022 est disponible au lien suivant : https://gitlab.com/data-challenge-gd4h/historiris/-/blob/main/data/iris/output/table_passage_1999_2022.csv.

Ce fichier est téléchargeable au format csv. Il contient 3 365 lignes et une colonne par millésime avec les codes IRIS correspondant pour chaque année.

Il convient de noter que la table de passage saute directement de 1999 à 2008 (les années de 2000 à 2007 sont absentes), l'Insee n'ayant pas produit de référentiel entre ces deux années.

Il existe également des tables de passages détaillées entre deux millésimes successifs (sauf pour le saut 1999-2008) :

<https://gitlab.com/data-challenge-gd4h/historiris/-/tree/main/data/iris/output>

Intérêt des données

Intérêt scientifique des données

La table de passage des IRIS de 1999 à 2022 permet la combinaison de différents indicateurs d'intérêt (environnementaux, sanitaires, socio-économiques...) produits à l'échelle des IRIS dans des millésimes différents. Elle facilite le travail préalable de préparation et d'harmonisation des données dans un seul et même référentiel en vue d'analyses.

À ce titre, cette table est utile et utilisable très largement dans le monde de la recherche et plus largement dans tous les travaux menés à l'échelle des IRIS et inscrits dans le temps.

Collecte et traitement des données

Collecte et processus de traitement mis en œuvre

- *Données utilisées :*

Les données mobilisées pour la création de la table de passage sont :

- Les tables de référence et les tables de passage biannuelle IRIS de l'Insee (disponible sur : <https://www.insee.fr/fr/information/2017499>)
- La table de passage annuelle communale 2023 de l'Insee (disponible sur : <https://www.insee.fr/fr/information/2028028>)
- Les données du recensement 1999 (Insee non diffusé) afin de construire la liste des IRIS 1999 avec les codes communes associés

- *Méthode :*

À partir des données de l'Insee, les modifications d'un millésime d'IRIS à l'autre ont été analysées dans des notebooks Python/Jupyter. L'objectif était de proposer des prototypes de table de passage biannuelle en décrivant et en expliquant les différences entre deux tables de référence des IRIS d'années successives.

Afin de pallier au manque de cohérence sur plusieurs millésimes de la documentation de l'INSEE et de faciliter son exploitation, deux notebooks de consolidation de cette documentation sur les périodes 1999-2012 et 2012-2022 ont été créés.

Les différents notebooks sont disponibles ici :

<https://gitlab.com/data-challenge-gd4h/historiris/-/tree/main/notebooks>.

Ils ont permis de créer un script automatisé permettant la construction de la table de passage. Ce script a été soumis à plusieurs tests unitaires, d'intégration et de cohérence.

Les scripts des tests sont disponibles ici :

<https://gitlab.com/data-challenge-gd4h/historiris/-/tree/main/tests>

Conditions d'accès et de réutilisation

L'ensemble des ressources est en accès libre. Le code est publié sous licence [MIT](#). Les données sont publiées sous [Licence Ouverte 2.0](#).

Cependant, lors de l'utilisation de la table de passage et il est important d'avoir en tête les éléments limitant suivants :

- L'Insee n'ayant pas produit de référentiel entre 1999 et 2008, la table de passage saute directement de 1999 à 2008 (les années de 2000 à 2007 sont absentes).
- A l'heure actuelle, seuls les IRIS ayant fait l'objet de modifications sont présents dans ce fichier. Les IRIS inchangés peuvent être récupérés et ajoutés à partir du référentiel (liste des IRIS) de l'Insee d'intérêt (<https://www.insee.fr/fr/information/2017499>).
- Il s'agit d'une première version. Des erreurs ou incohérences persistent (par exemple, si les modifications ne sont pas documentées par l'Insee). Certaines ont déjà été identifiées et remontées dans la rubrique « Issues » de GitLab (<https://gitlab.com/data-challenge-gd4h/historiris/-/issues>) afin de pouvoir les traiter. Lors de l'utilisation de la table, il est fortement recommandé de vérifier que la conversion s'est bien effectuée et de faire remonter à la communauté tout problème rencontré dans la rubrique « Issues » du git.

L'objectif est de faire évoluer la table de passage et le git associé de manière collaborative. Votre contribution est indispensable à la consolidation de la table de passage.

Plus-value et perspectives

La table de passage produite et les ressources associées sont des outils collaboratifs amenés à évoluer (correction d'erreurs ou d'incohérences) et à être mis à jour avec les nouveaux millésimes de l'Insee.

Bibliographie

- Institut national de la statistique et des études économiques | Insee [En ligne]. Découpage infracommunal ; [cité le 29 juin 2023]. Disponible : <https://www.insee.fr/fr/information/2017499>
- Institut national de la statistique et des études économiques | Insee [En ligne]. Découpage communal ; [cité le 29 juin 2023]. Disponible : <https://www.insee.fr/fr/information/2028028>



Green Data for Health

Data papers du Challenge Green Data for Health

Le Green Data for Health

Le Green Data for Health (GD4H) est une offre de service incubée au sein de l'ECOLAB, laboratoire d'innovation pour la transition écologique du Commissariat général au développement durable (Ministère de la Transition écologique et de la Cohésion des territoires et Ministère de la Transition énergétique).

Action phare du 4^o Plan National Santé Environnement, le GD4H vise à permettre une meilleure mobilisation et valorisation, par les acteurs de la recherche et de l'expertise, des données environnementales au service de la santé-environnement.

Pour en savoir plus : gd4h.ecologie.gouv.fr

Le Challenge Green Data for Health

Le Challenge Green Data for Health est une initiative portée par l'Ecolab du Commissariat général au développement durable (Ministères Ecologie Energie Territoires). Dans ce cadre, 15 équipes ont été constituées pour mener 15 projets visant à mieux mobiliser la donnée environnementale au service de la santé-environnement.

Pour en savoir plus : gd4h.ecologie.gouv.fr/defis