Data Article

Data for the assessment of vulnerability and resilience in the field of environmental health in the north of France

Delphine Brousmiche[a], Michaël Genin[b], Florent Occelli[a], Lukas Frank[a], Annabelle Deram[a], Damien Cuny[a], Caroline Lanier[a],*

[a] Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France, ULR 4515 – LGCgE, Laboratoire de Génie Civil et géo-Environnement, F-59000 Lille, France
[b] Univ. Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des technologies de santé et des pratiques médicales, F-59000 Lille, France

A R T I C L E   I N F O

A B S T R A C T

The integration of multidimensional data is necessary to improve the understanding of environmental and social inequalities in health. The challenge is to define a dataset that provides the most holistic description possible of the territory. This article presents a relevant dataset to characterize the territorial accumulation of health determinants in the second most densely populated region of metropolitan France (Hauts-de-France Region, in the north of France). The multidimensional dataset combines data related to the economic, social, environment, services, health and policy dimensions at fine scale (i.e., each municipality). Data outlining a negative impact on health inequalities (e.g. anthropogenic pressures, socioeconomics factors related to vulnerability, etc.) are considered to be as important as data outlining a positive impact on health inequalities (e.g. natural resources, diversity and economic drive, etc.). The proposed theoretical framework relies on data reuse. Over one hundred variables covering a time frame from 2008 to 2017 were collected from a dozen public and national database providers. The use of official organizations ensured the quality of the collected data. The Geographic Information System, designed

to map and catalogue ready-to-use data, was used to generate new data or to deal with missing data. Finally, 50 variables, including mostly quantitative but also qualitative data, were selected after application of inclusion and exclusion criteria. The resulting dataset provides a broad characterisation of the 3,817 municipalities in the Hauts-de-France Region. These data will help to discriminate the distribution pattern of vulnerability and resilience levels in this region. This novel approach is described in the paper "How can we analyse environmental health resilience and vulnerability? A joint analysis with composite indices applied to the north of France", which provides a detailed description of the methodology used to develop composite indices. This research could therefore be of use to researchers, policy makers and stakeholders in the field of environmental health seeking to identify the weaknesses but also the strengths of municipalities.

Specifications Table

Subject	Public Health and Health Policy
Specific subject area	Territorial characteristics and environmental and social inequalities in health
Type of data	Table
How data were acquired	Data *reuse*, collected from official national organizations: AGENDA21FRANCE.ORG, BPE (Permanent Facilities Database), CNAF (National Health Insurance Fund), DGCL (General Management of Local Authorities), DGFiP (General Directorate of Public Finance), DGPR (General Directorate for Risk Prevention), DREES (Directorate of Research, Studies, Evaluation and Statistics), Geofabrik GmbH, GEOIDD (National Observation and Statistics Service), IGN (National Institute for Geographic and forest information), INSEE (French National Institute of Statistics and Economic Studies), MEDDE (Ministry of Ecological and Solidarity Transition), ODT (French National Observatory of Territories) and REE (National Directory of Companies and Establishments).
Data format	Raw, calculated
Parameters for data collection	Three inclusion criteria: (i) a plausible association with environmental health in the literature, (ii) comprehensive, spatially contiguous and contemporary data for the entire study area, (iii) production and regular updating by a reliable organization (thereby ensuring acceptable data quality). One exclusion criterion: data associated with individual behavior (*i.e.* smoking, food habits, physical activities)
Description of data collection	36 % of raw data were integrated in the dataset 64 % of data have been calculated (evolution calculation, data converted to per capita, data per 1,000 inhabitants, raw data converted to rate, transdisciplinary adaptation of calculations, area calculations with GIS software: ArcGIS v10.7, Q-GIS v2.14) Data-management operations were made with Excel (2016) Statistical analysis were made with R software (R-3.3.2)
Data source location	"Hauts-de-France" Region, France Primary data sources: Table II-Table IV
Data accessibility	Data are hosted on a public repository. Repository name: Zenodo Data identification number: 10.5281/zenodo.3701557 Direct URL to data: https://zenodo.org/record/4889374#.YLYg7agzY2w
Related research article	Delphine Brousmiche, Michaël Genin, Florent Occelli, Lukas Frank, Annabelle Deram, Damien Cuny, Caroline Lanier How can we analyse environmental health resilience and vulnerability? A joint analysis with composite indices applied to the north of France Science of The Total Environment https://doi.org/10.1016/j.scitotenv.2020.142983

Value of the Data

- The data are useful for environmental health issues as they allow users to map and discriminate the distribution pattern of vulnerability and resilience levels at a municipality scale.
- These data allow a cross-comparison between 3817 municipalities based on 6 dimensions of health determinants. This can result in a more detailed identification of the strengths and weaknesses of municipalities faced with the challenges of social and environmental inequalities in health.
- Researchers, policy makers and other stakeholders may use these data to build new databases on health determinants at a fine scale to compare their findings in different regions of France or in different countries.
- These multidimensional data will help society to consider possible levers for action to improve public health and well-being for all in the short term, and to develop a "Health in All Policies" approach over the long term.

1. Data Description

Validated by the National Institute for Geographic and forest information (IGN) arrested at the date of 12/31/2016, the region had 3,817 municipalities, a municipality being a small administrative entity in France. 50 data were collected and are fully available in the Zenodo repository (accession number: 10.5281/zenodo.3701558, Table 1 "Data accessibility"). The 6 dimensions of health inequalities determinants were investigated: "Economics" dimension in Table 1, "Environment" dimension in Table 2, "Policy" dimension in Table 3, "Health" dimension in Table 4, "Social" dimension in Table 5 and "Services" dimension in Table 6. These tables include:

- the given acronym, the description of the data and specifications,
- the data unit,
- their(s) source(s),
- the reference year of the data,
- the rate of missing data,
- a mention if the variable is "raw" or "calculated",
- a statistical description with (i) the median and the inter-quartile range [Q1-Q3] calculated after the imputation of missing data step for quantitative data and (ii) the modalities of qualitative data and their frequency distribution which is expressed as a percentage.

In the context of data *reuse*, the use of official organizations ensured the credibility of the collected data [1]. The good completeness of the data-set is pointed up regarding the low rate of missing data. 62 % of variables has no missing data (31/50 collected data). For the remaining 38 % (19/50), the percentage of missing data ranges from 0.03 % (1/3,817 values) to 2.07 % (79/3,817 values). 36 % of the raw data were integrated directly into the dataset, without computation. The other 64 % were obtained by calculation: some could be described as usual (*e.g.* percentage, data per inhabitant), others inspired by previous works specialised in socio-ecology systems (*e.g.* human to ecosystem service value ratio index) or resulting from a transdisciplinary adaptation (*e.g.* Shannon-Weaver index to calculate the representation of socio-professional categories within each municipality).

Tables 7–10.

2. Experimental Design, Materials and Methods

2.1. Evolution calculations

This section concerns four data related to "Economics" dimension (Table 1: the 5-year survival rate of companies "EC_5YSURVRATE", the Employment evolution for attendance

Table 1

Metadata related to the "Economics" dimension.

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median –[Q1; Q3]or modalities (%)	Link to primary data
EC_5YSURVRATE	The 5-year survival rate of companies	Calculation of a survival rate between 2 non-consecutive years.	%	INSEE, ODT, REE	2009-2014	0.21	Calculated	11.1–[0; 50]	https://www.insee.fr/fr/statistiques/2521005?sommaire=2544975
EC_EMPLOY	Employment rate	Proportion of economically active persons aged 15-64 (in employment) in the total population aged 15-64.	%	INSEE, ODT	2013	0	Raw	65.6–[61.2; 69.1]	https://www.insee.fr/fr/statistiques/2861854
EC_EVEMPATTEND	Employment evolution for attendance activities	Attendance activities are defined as activities that are implemented locally for the production of goods and services designed to satisfy the needs of residents or tourists in the municipality.	%	INSEE	2009-2014	0.29	Calculated	0–[-0.3; 0.4]	https://www.insee.fr/fr/statistiques/1893206
EC_EVLABFOR	Evolution of the labour force	Calculation of a change in the number of people aged 15-64 in the labour force between 2 non-consecutive years.	%	INSEE	2008-2013	0	Calculated	0.43–[-2.1; 2.89]	https://www.insee.fr/fr/statistiques/2518836
EC_INCOME	Average income per consumption unit	The disposable income of a household comprises earned income (net of social contributions), income from assets, transfers from other households and social benefits (including pensions and unemployment benefits), and net of direct taxes. The consumption unit (CU) is an OECD unit of measurement that weights individuals in a household according to their age, allowing comparison between households. CU are generally calculated in the following manner: 1 CU is allocated to the first adult in the household, 0.5 CU to other persons aged 14 or over and 0.3 CU to children under 14.	€/CU	INSEE, ODT	2012	0	Raw	19,546.7 –[17,904.7; 21,412.7]	https://www.insee.fr/fr/statistiques/2388572 http://carto.observatoire-des-territoires.gouv.fr/#l=fr;i=filosofi.med_disp;v=map56

(*continued on next page*)

Table 1 (*continued*)

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median –[Q1; Q3]or modalities (%)	Link to primary data
EC_MUNLIVEWORK	Share of the labour force living and working in the same municipality	The labour force considered is aged 15 and over.	%	INSEE	2014	0	Calculated	15.6–[11.9; 20.5]	https://www.insee.fr/fr/statistiques/2861854
EC_PARTLABFOR	Evolution of the share of the labour force in the total population	Calculation of the evolution of the share of the labour force aged 15 and over within the total population between 2 non-consecutive years.	%	INSEE	2009-2014	0.29	Calculated	-1.2 – [-5.1; 3.3]	https://www.insee.fr/fr/statistiques/2861854 https://www.insee.fr/fr/statistiques/2862200
EC_PRECARIOUS	Precarious employment rate	Share of the labour force aged 15-64 in fixed-term contracts, temporary employment, assisted employment and apprenticeships among the labour force aged 15-64.	%	INSEE	2014	0	Calculated	7.6–[6.1; 9.1]	https://www.insee.fr/fr/statistiques/2863622 https://www.insee.fr/fr/statistiques/2861854

Table 2

Metadata related to the "Economics" dimension (final section).

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median –[Q1; Q3]or modalities (%)	Link to primary data
EC_SHANSPC	Shannon-Weaver index on the representation of socio-professional categories	Use of the Shannon-Weaver index to describe a population in terms of the number of categories (category richness) and their relative abundance within those categories (species equitability). 15 socio-professional categories were considered for the calculation, identified at the place of residence.	Index	INSEE	2014	0.29	Calculated	0.7–[0.5; 0.9]	https://www.insee.fr/fr/statistiques/1893206
EC_UNEMPLOY	Unemployment rate	Share of people aged 15-64 who are unemployed and looking for work.	%	INSEE	2014	0	Raw	8.5–[6.6; 10.9]	https://www.insee.fr/fr/statistiques/2861854

Table 3

Metadata related to the "Environment" dimension.

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3] or modalities (%)	Link to primary data
EN_ACCGSPACE	Share of population with access to green space within 400 meters of their homes	Green spaces of less than 0.5ha have been excluded. The distance of 400 meters has been used as it is more widely accepted in the literature than the 300-meter distance recommended by the WHO. This criterion included parks, forests and playground. It excludes areas that are inaccessible to the general population (e.g. campsites, cemeteries, golf courses, grassy areas, stadiums and zoos). The population data are those of the INSEE's 200-meter grid.	%	INSEE, Geofabrik	2011-2017	0	Calculated	97.2–[82.0; 100]	http://download.geofabrik.de/europe/france.html https://www.insee.fr/fr/statistiques/2520034
EN_GHGEMIS	Greenhouse gas (GHG) emissions	The Global Warming Potential (GWP) expresses, in thousands of tonnes of CO_2 equivalent, the cumulative effect of the substances emitted into the air that contribute to the increase in the greenhouse effect. It is calculated on the basis of the warming power of each gas over the next 100 years compared to that of CO_2. The following direct greenhouse gases are taken into account: CO_2 (non-biotic), CH_4, N_2O, HFCs, PFCs and SF_6. GWP expressed here excludes carbon sinks. These sinks mainly correspond to the absorption of CO_2 for vegetation growth and methane absorption by forest soils. Similarly, emissions from biotic sources (non-managed forests, natural grasslands, wetlands and rivers, forest fires) are not included. Solely the emissions caused by human activities or managed by humans are taken into account. These exclude emissions from international shipping and aviation.	Thousands of TEQ CO_2	ODT	2012	0.58	Raw	4 – [5; 8]	http://www.observatoire-des-territoires.gouv.fr/observatoire-des-territoires/fr/emissions-de-gaz-%C3%A0-effet-de-serre-hors-puits

(*continued on next page*)

Table 3 (continued)

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3] or modalities (%)	Link to primary data
EN_HESV	Monetisation of ecosystem services per capita	Use of the Corine Land Cover database to determine 4 land-use categories (grassland, urban, cultivated and/or arable land, forests) and The Economics of Ecosystems and Biodiversity valuation database to assign a monetary value to the ecosystem services provided by these 4 categories [5,6].	USD/ha/year	INSEE, MEDDE	2012-2014	0	Calculated	5.8–[2.6; 11.6]	https://www.statistiques.developpement-durable.gouv.fr/corine-land-cover-0 https://www.insee.fr/fr/statistiques/2862200
EN_M2NATSPACE	Area (m²) of natural spaces per inhabitant	Use of open street map data (land use, points of interest): areas that are inaccessible to the general public (parks, playgrounds, campsites, cemeteries, grassy areas, golf courses, stadiums, zoo) were also included. Water surface (lakes, ponds, rivers, etc.) data were added.	m²/inhab	INSEE, Geofabrik	2014-2017	0	Calculated	1,872.4 – [648.5; 4,619.0]	http://download.geofabrik.de/europe/france.html https://www.insee.fr/fr/statistiques/2520034
EN_PSSNB	Number of contaminated sites (polluted sites and soils)	Polluted sites and soils are sites that present an actual or potential perennial risk to human health or the environment due to pollution caused by current or former human activity. The database registers all the sites that require or are the subject of public, preventive or curative action. These are therefore situations that are clearly identified, dealt with, being dealt with or about to be dealt with.	Count	GEOIDD, MEDDE, DGPR	2015	0	Raw	0 – [0; 0]	http://geoidd.developpement-durable.gouv.fr/geoclip_stats_o3/#l=fr;i=sol_pollution.sitpol_nb;v=map1

Table 4

Metadata related to the "Policy" dimension.

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3] or modalities (%)	Link to primary data
PO_AGENDA21	Municipality covered by an Agenda 21	Agenda 21 are developed at the local authority or sub-departmental level as a measure of sustainable development. This action plan reflects the voluntary commitment of communities to sustainable development. Qualitative data recoded in a binary variable.	-	GEOIDD, MEDDE, AGENDA21FRANCE.ORG	2014	0	Raw	0: not covered by an Agenda 21 (81.3 %) 1: covered by an Agenda 21, local and/or sub-regional (18.7 %)	http://geoidd. developpement-durable. gouv.fr/geoclip_stats_o3/ #s=2014;l=fr;i=agenda21. p_a21_pop_infradep;v= map1
PO_EXPTOTINVEST	Share of expenditure allocated to municipal facilities and services compared to the total investment	The total investment corresponds to the municipality's investments (equipment, urban development, *etc.*) and loan repayments.	%	DGFiP	2014	0.68	Calculated	81.4–[60.5; 93.4]	https://www.data.gouv.fr/ fr/datasets/ donnees-comptables-et -fiscales-des-collectivit es-locales/
PO_FINANPOT	Financial potential per inhabitant	The financial potential measures the wealth of a municipality by calculating all the stable resources of a local authority in relation to the reference population. These resources include all tax resources and allocations paid by the State on an automatic basis or recurring basis, and which are essential elements for the balancing of local authority budgets.	€/inhab	ODT, DGCL, INSEE	2013	0	Raw	567–[486.4; 723.2]	http://www. observatoire-des-ter ritoires.gouv.fr/ observatoire-des-terri toires/fr/ potentiel-financier- par-habitant

(*continued on next page*)

Table 4 (*continued*)

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3]or modalities (%)	Link to primary data
PO_HOUSINGTAX	Housing tax proceeds compared to the number of primary and secondary residences	The housing tax is a local tax which varies according to the characteristics of a home, its location and the personal situation of its occupants (income, composition of the household, *etc.*) on 1st January every year. Owners, tenants and free occupants of their main dwelling must pay this tax, and a housing tax must also be paid for a secondary residence. The total sum is then divided by the number of primary and secondary residences counted within the municipality.	€/number of primary and secondary residences	DGFiP, INSEE	2015	0	Raw	472.4 – [392.8; 567.9]	http://carto. observatoire-des-terri toires.gouv.fr/#s=2015;z= 10411,6631335,650975, 477012;l=fr;i=fisc_loc. th_prod_locaux;v=map1
PO_MUNDEBT	Proportion of medium- and long-term debt of the municipality, per inhabitant	It measures the medium and long-term debt burden of the municipality. The annuity of the debt is obtained by summing the interest on the loans of the municipality and the amount of the principal repayment. The addition of these two amounts permits the measurement of the exact weight of the long and medium-term debt for municipalities.	€/inhab	DGFiP, INSEE	2014	0	Calculated	44.7–[18.8; 77.1]	https://www.data.gouv.fr/ fr/datasets/ donnees-comptables-et -fiscales-des-collectiv ites-locales/ https://www.insee.fr/fr/ statistiques/2862200

Table 5

Metadata related to the "Policy" dimension (cont.).

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3]or modalities (%)	Link to primary data
PO_MUNEXPEQUIP	Expenditure for municipal equipment, per inhabitant	This expenditure corresponds to the expenditure of the municipality on tangible and intangible fixed assets or fixed assets under construction (construction of a school, a tramway line or a sports facility), in relation to the reference population.	€/inhab	DGFiP, INSEE	2014	0	Calculated	139.8–[59.2; 297.6]	https://www.data.gouv.fr/fr/datasets/donnees-comptables-et-fiscales-des-collectivites-locales/ https://www.insee.fr/fr/statistiques/2862200
PO_MUNSUBS	Amount of subsidies paid by municipal-ities, per inhabitant	The subsidies are paid for sports and cultural associations, communal centre for social welfare, local mixed-economy companies, structures such as tourist boards, *etc*, in relation to the reference population. The subsidies are part of the municipality's operating costs (costs of personal among others), which take priority over investment costs.	€/inhab	DGFiP, INSEE	2014	0	Calculated	13.1–[6.6; 23.9]	https://www.data.gouv.fr/fr/datasets/donnees-comptables-et-fiscales-des-collectivites-locales/ https://www.insee.fr/fr/statistiques/2862200
PO_MUNRESULT	Overall financial result of the municipality, per inhabitant	This result corresponds to the difference between the accounting result (difference between total operating income and operating expenses) and the financing requirement or capacity of the investment section, in relation to the reference population.	€/inhab	DGFiP, INSEE	2014	0	Calculated	108.5–[9.7; 249.5]	https://www.data.gouv.fr/fr/datasets/donnees-comptables-et-fiscales-des-collectivites-locales/ https://www.insee.fr/fr/statistiques/2862200

(*continued on next page*)

Table 5 (*continued*)

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3]or modalities (%)	Link to primary data
PO_RAZ	Eligible areas for regional aid zoning	Companies based in any municipalities located within an AFR ('Aide à Finalité Régionale') zone can be allocated aids for investment and job creation by the State and local authorities. Qualitative data (recoded in a binary variable).	-	ODT	2015	0	Raw	0: non-eligible (82.0%) 1: partially or completely eligible (18.0%)	http://www. observatoire-des-territo ires.gouv.fr/ observatoire-des-territo ires/fr/zones-daide-%C3% A0-finalit%C3%A9-r%C3% A9gionale
PO_SUBSOPEXP	Share of subsidies paid within operating expenses	Operating expenses mainly include personnel costs, compulsory contributions, in particular to the departmental fire and rescue service, subsidies paid and interest on loans.	%	DGFiP	2014	0	Calculated	2.6–[1.4; 4.4]	https://www.data.gouv.fr/ fr/datasets/ donnees-comptables-et -fiscales-des-collecti vites-locales/

Table 6

Metadata related to the "Policy" dimension (final section).

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3]or modalities (%)	Link to primary data
PO_TAXHOUSEHOLDS	Share of taxable tax households	The share of tax households taxed is the percentage of households that are liable to pay income tax. The tax liability of a tax household is the sum of the taxes payable by each person in the household.	%	INSEE, ODT	2013	2.07	Raw	52.1–[48.3; 56.4]	http://carto. observatoire-des-territ oires.gouv.fr/#v=map8;i= filosofi.men_impos;l=fr; z=25439,6620976, 570632,367308
PO_TAXPOT	Tax potential per inhabitant	Tax potential is an indicator that can be used to compare the potential tax wealth of communities. A theoretical tax product is calculated, corresponding to the amount of taxes that each community would collect if it applied average rates or tariffs to its net tax bases. This tax potential is related to the number of inhabitants in order to allow analyses and comparisons that are not possible using gross amounts alone.	€/inhab	ODT, DGCL	2017	0.21	Raw	476.7–[387.4; 622.4]	http://carto. observatoire-des-territo ires.gouv.fr/#l=fr;i= potentiel_fin. pot_fisc_hab;v=map43

Table 7

Metadata related to the "Health" dimension.

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3]or modalities (%)	Link to primary data
HE_ACCGP	Potential spatial accessibility to general practitioners	This indicator is based on three variables: demand, supply, distance. It is calculated at the municipal level, taking into account the supply and demand of neighbouring municipalities. This accessibility is considered 'potential' because it is based on the geographical location of demand (population) and supply (general practitioners). It measures the activity level of general practitioners, based on the number of consultations and visits recorded over one year. This number is converted into Full-Time Equivalents (FTE).	Full-time equiva-lent/inhab.	DREES	2013	0	Raw	47.7–[31.9; 67.4]	http://www.data.drees.sante.gouv.fr/ReportFolders/reportFolders.aspx
HE_ACCPHARM	Potential spatial accessibility to pharmacist	This indicator is based on three variables: demand, supply, distance. It is calculated at the municipal level, taking into account the supply and demand of neighbouring municipalities. This accessibility is considered 'potential' because it is based on the geographical location of demand (population) and supply (pharmacist). It measures the activity level of pharmacists, based on the number of consultations and visits recorded during a year. These acts are converted into full-time equivalents (FTE).	Full-time equiva-lent/inhab.	DREES	2013	0	Raw	23.2–[15.6; 33.0]	http://www.data.drees.sante.gouv.fr/ReportFolders/reportFolders.aspx

(*continued on next page*)

Table 7 (*continued*)

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3]or modalities (%)	Link to primary data
HE_BIRTH	Proportion of births	Locally registered births correspond to births counted from civil status reports, and recorded at the mother's home address.	Number per 1,000 inhab.	INSEE	2014	0	Raw	1.1–[0.7; 1.5]	https://www.insee.fr/fr/statistiques/2521169
HE_DEATH	Proportion of deaths	The place of death is the last home address of the deceased and not the place of death.	Number per 1,000 inhab.	INSEE	2014	0.47	Raw	0.7–[0.4; 1.1]	https://www.insee.fr/fr/statistiques/2521169
HE_HEALTHPROF	Number of health professionals, per 1,000 inhabitants	Number of health professions, all specialities combined, counted at the zip-code of the place of activity, per 1,000 inhabitants.	Number per 1,000 inhab.	INSEE	2016	0.21	Calculated	0–[0; 3.2]	https://www.insee.fr/fr/statistiques/3568614?sommaire=3568656
HE_SHANMED	Shannon-Weaver index on the representation of medical and paramedical professions	Use of the Shannon-Weaver index to describe a population in terms of the number of categories (category richness) and their relative abundance within those categories (category equitability). 25 medical and paramedical professions were considered for the calculation.	Index	INSEE	2016	0.03	Calculated	0–[0; 0.3]	https://www.insee.fr/fr/statistiques/3568614?sommaire=3568656

Table 8

Metadata related to the "Social" dimension.

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3]or modalities (%)	Link to primary data
SO_DENSITY	Population density	Number of inhabitants per ha.	inhab/ha	INSEE	2014	0.21	Calculated	59.1–[33.3; 131.0]	https://www.insee.fr/fr/statistiques/2862200
SO_EMPEDUC	Employment in education and training, per 1,000 inhabitants	Employment in education and training corresponds to jobs in school and university education (primary, secondary and higher education) and vocational training, including the organisation of such education.	Number per 1,000 inhab	INSEE	2014	0	Calculated	0–[0; 9.0]	https://www.insee.fr/fr/statistiques/1893206
SO_GRIDDENS	Communal density grid	In order to consider the municipal population and its distribution in space, the communal density grid is based on the distribution of the population within the municipality by dividing the territory into one-kilometre squares. The size of these agglomerated areas within the municipalities will make it possible to characterize them (rather than the usual communal density). Qualitative data.	-	INSEE, ODT	2015	0.24	Raw	1: densely populated area (2.1 %) 2: intermediate densely populated area (13.6 %) 3: sparsely populated area (56.1 %) 4: very sparsely populated area (28.2 %)	http://www.observatoire-des-territoires.gouv.fr/observatoire-des-territoires/fr/grille-communale-de-densit

Table 8 (*continued*)

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3]or modalities (%)	Link to primary data
SO_IMMIG	Share of immigrants in the total population	An immigrant is a person born abroad and residing in France. A distinction is made between foreign and immigrant populations: an immigrant is not necessarily a foreigner and conversely, some foreigners were born in France. Immigrant status is permanent: an individual continues to belong to the immigrant population even if he/she becomes French by acquisition.	%	INSEE	2015	0	Raw	1.5–[0.8; 2.8]	https://www.insee.fr/fr/ statistiques/3564100? sommaire=3561107 https://www.insee.fr/fr/ statistiques/3561125? sommaire=3561139
SO_LESSQUALIF	Share of working-age population with few or no qualifications	Ratio between the number of out-of-school persons aged 15 years or over with no diploma (or at most a BEPC which is a French certificate of general education, awarded by schools at the end of the first four years of general secondary education) and the total number of out-of-school persons aged 15 or over.	%	INSEE	2014	0.29	Calculated	34–[28; 40]	https://www.insee.fr/fr/ statistiques/2862015
SO_POP65	Share of the population aged over 65 years	Percentage of the population aged over 65 years in the total population.	%	INSEE	2015	0	Calculated	16.3–[13.7; 19.3]	https://www.insee.fr/fr/ statistiques/3564100? sommaire=3561107

Table 9

Metadata related to the "Social" dimension (final section).

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3]or modalities (%)	Link to primary data
SO_PUBSCHOOL	Number of state schools, per 1,000 inhabitants	This correspond to the sum of the number of state nursery schools, state elementary schools and state elementary schools with a pre-elementary class.	Number per 1,000 inhab	INSEE, BPE, GEOIDD	2014	0	Calculated	1.4–[0; 2.9]	http://geoidd. developpement-durable. gouv.fr/geoclip_stats_o3/ #l=fr;v=map1
SO_SINGLEFAM	Share of single-parent families	A single-parent family includes a lone parent and one or more unmarried (childless) children.	%	ODT	2013	0.13	Raw	9.6–[5.1; 14.3]	http://carto. observatoire-des-terri toires.gouv.fr/#s=2015;l= fr;i=insee_rp_hist_xxxx. part_menfammono;v= map56
SO_SOCBENEF	Share of persons covered by a social benefit	Percentage of persons covered by a benefit from the family branch, such as family benefits, early childhood benefits, family supplement, education allowance for a disabled child, family support allowance, back-to-school allowance, family housing allowance, social housing allowance, *etc.*	%	CNAF	2014	2.07	Raw	44.6–[39.6; 49.6]	http://data.caf.fr/dataset/ population-des- foyers-allocataires- par-commune
SO_STUDY	Study continuation rate for 18-24 years old	Share of the population aged 18 to 24 pursuing higher education.	%	INSEE	2013	0.65	Calculated	39.1–[30.0; 49.2]	https://www.insee.fr/fr/ statistiques/2044692

Table 10
Metadata related to the "Services" dimension.

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3]or modalities (%)	Link to primary data
SE_4GCOV	Part of area with 4G cover provided by the most efficient operator	This indicator represents the share of areas covered for 4G by the operator offering the best coverage rate in the municipality. It reflects the availability, outside buildings, of access to a service, as displayed by the operators on their coverage maps.	%	ODT	2016	0	Raw	92 – [57; 99]	http://www. observatoire-des-territoires. gouv.fr/ observatoire-des-territoires/ fr/node/2289
SE_EQUIP15MIN	Share of population with average access to 12- mid-range municipal facilities and services in 15 minutes or less	The equipment in the intermediate range are in common use, without being close by. These may include supermarkets, bookshops and stationery shops, clothing stores, college, opticians, MOT centres, *etc;*	%	ODT	2011	0	Raw	94 – [85.5; 97.7]	http://www. observatoire-des-territoires. gouv.fr/ observatoire-des-territoires/ de/volution-de-la -part-de-la-population- ayant-acc-s-en-moyenne- aux-12-quipements-de-la -gamme-interm-dia? rech=1
SE_LRH	Low-Rent Housing (LRH) per 1,000 inhabitants	Low-rent housing is housing units managed by a public or private low-rent housing organization, which benefits from partial, direct (subsidy) or indirect public funding (various advantages including credits, tax exemptions, *etc.*).	Number per 1,000 inhab	INSEE	2013	0	Raw	0 – [0; 18.9]	https://www.insee.fr/fr/ statistiques/2044711

(*continued on next page*)

Table 10 (*continued*)

Acronym	Data	Specifications	Unit	Source(s)	Time period	Missing data (%)	Raw / Calculated	Median – [Q1; Q3]or modalities (%)	Link to primary data
SE_OWNER	Proportion of people owning their main residence	Ownership status applies to owner, co-owner and first-time homeowner households.	%	INSEE	2014	0.21	Calculated	81.5 – [73.9; 86.5]	https://www.insee.fr/fr/statistiques/2862034
SE_PRIMROAD	Length of primary roads per capita	The primary road is defined as a portion of a roadway intended for pedestrian or cycles and includes only sections of major road networks.	m/inhab	IGN, INSEE	2016	0.08	Calculated	0 – [0; 4.1]	http://professionnels.ign.fr/bdtopo https://www.insee.fr/fr/statistiques/3677785?sommaire=3677855
SE_TENANT	Share of population who are tenants of their main residence	The rental stock comprises the public and the private sectors.	%	INSEE	2014	0.29	Calculated	16.7 – [11.7; 24]	https://www.insee.fr/fr/statistiques/2862034
SE_TENANTLRH	Share of tenants who rent their main residence in Low-Rent Housing	LRH dwellings offer rents well below those of the private rental sector.	%	INSEE	2014	0.29	Calculated	0 – [0; 4.8]	https://www.insee.fr/fr/statistiques/2862034

activities "EC_EVEMPATTEND", the evolution of the labour force "EC_EVLABFOR" and the evolution of the share of labour force in total population "EC_PARTLABFOR"). Only data corresponding to year n and $(n-5)$ were collected for the evolution calculation. Considering a year n (year 2013: "EC_EVLABFOR"; year 2014: "EC_5YSURVRATE", "EC_EVEMPATTEND", "EC_PARTLABFOR"), the evolution was computed as follows :

$$\text{Evolution} = \frac{\text{Raw data (year (n))} - \text{Raw data (year (n-5))}}{\text{Raw data (year (n-5))}} * 100$$

2.2. Raw data converted to per capita

This section concerns one data related to "Environment" dimension (Table 2: the area (m²) of natural spaces per inhabitant "EN_M2NATSPACE"), four data related to "Policy" dimension (Table 3: the expenditure for the municipality equipment per inhabitant "PO_MUNEXPEQUIP", the amount of subsidies paid by municipalities per inhabitant "PO_MUNSUBS", the overall financial result of the municipality per inhabitant "PO_MUNRESULT", the proportion of medium- and long-term debt of the municipality per inhabitant "PO_MUNDEBT") and one data related to "Services" dimension (Table 6: the length of primary roads per capita "SE_PRIMROAD"). Considering a year n, the per capita variable was computed as follows:

$$\text{Per capita variable} = \frac{\text{Raw data (year (n))}}{\text{Population reference (year (n))}}$$

2.3. Raw data per 1000 inhabitants

This section concerns one data related to "Health" dimension (Table 4: the number of health professionals, per 1,000 inhabitants "HE_HEALTHPROF") and two data related to "Social" dimension (Table 5: the employment in education and training per 1,000 inhabitants "SO_EMPEDUC" and the number of state schools per 1000 inhabitants "SO_PUBSCHOOL"). Considering a year n, the raw data per 1,000 inhabitants was computed as follows:

$$\text{Variable per } 1,000 \text{ inhabitants} = \frac{\text{Raw data (year (n))}}{\text{Population reference (year (n))}} * 1,000$$

2.4. Raw data converted to rate

This section concerns two data related to "Economics" dimension (Table 1: the share of the labour force living and working in the same municipality "EC_MUNLIVEWORK", the precarious employment rate "EC_PRECARIOUS"), two data related to "Policy" dimension (Table 3: the share of expenditure allocated to municipal facilities and services to the total investment "PO_EXPTOTINVEST", the share of subsidies paid within the operating expenses "PO_SUBSOPEXP"), three data related to "Social" dimension (Table 5: the share of working age population with few or no qualifications "SO_LESSQUALIF", the share of the population aged over 65 years "SO_POP65", the study continuation rate for 18–24 years old "SO_STUDY") and three data related to "Services" dimension (Table 6: the proportion of people owning their main residence "SE_OWNER", the share of the population who are tenants of their main residence "SE_TENANT", the share of tenants who rent their main residence in low-rent housing "SE_TENANTLRH"). Considering a year n, the raw data converted to rate was computed as follows:

$$\text{Percentage} = \frac{\text{Raw data (year (n))}}{\text{Global amount (year (n))}} * 100$$

2.5. Calculation of the share of population with access to green space within 400 meters of their homes (EN_ACCGSPACE)

This data was based on WHO recommendations [2], which proposes in 2016, an indicator related to the health benefits of green spaces: the share of the population living within 300 m of an area of at least 0.5 ha. In our study, the distance of 400 m was chosen for the calculation because this distance is currently consensual in the scientific literature (Table 2). In the review of Labibe et al. (2020), this distance of 400 m is considered to reflect a walking distance of five to 10 min and is the median distance for some health studies [3].

We extracted data from Open Street Map (Geofabrik GmbH) to determine the green space surface area. Parks, forests and playgrounds greater than 0.5 ha have been included whereas areas not accessible to the general population (camping sites, cemeteries, golf courses, grassy areas, stadiums and zoos) have been excluded. We also used the 200*200 m gridded population data from the INSEE. Geographic Information System (ESRI ArcGIS v10.7) were used to generate a 400-meters buffer zone around each green space. The point centroid of each population grid was then intersected with the buffers in order to calculate the proportion of population living in the buffer in a municipality.

2.6. Calculation of the monetisation of ecosystem services per capita (EN_HESV)

This section concerns the monetisation of ecosystem services per capita "EN_HESV", which was inspired by the work of Estoque and Murayama (2014) [4] (Table 2). We downloaded from MEDDE the Coordination of Information on the Environment (CORINE) Land Cover (CLC) database, a pan-European geographic database, which describes the biophysical land use. Among the 5 major classes (*i.e.* artificial territories, agricultural territories, forests and semi-natural environments, wetlands, water surfaces), 44 land cover types are identified by a three-digit notation. In our study, this database was used to create 4 specific land use categories:

- "urban", including codes 111 (continuous urban fabric), 112 (continuous urban fabric), 121 (industrial and commercial areas), 122 (road and rail networks and associated areas), 123 (port areas) and 124 (airports),
- "cultivated and/or arable land", including codes 211 (non-irrigated arable land), 212 (permanently irrigated perimeter), 213 (rice paddies), 221 (vineyards), 222 (orchards and berries), 231 (pastures), 241 (annual crops associated with permanent crops), 242 (complex cropping and parcel systems) and 243 (mainly agricultural areas, interrupted by large natural areas),
- "grasslands", including codes 321 (natural lawns and pastures), 322 (heath and scrub) and 323 (sclerophyllous vegetation),
- "forests", including codes 311 (deciduous forests), 312 (coniferous forests), 313 (mixed forests) and 324 (changing forest and shrub vegetation).

The area in hectares for each of these 4 categories was calculated for each municipality, using Geographic Information System tools. Thereafter, the Economics of Ecosystems and Biodiversity (TEEB) Valuation Database from works by de Groot *et al.* (2012) and van der Ploeg et al. (2010) [5,6] was studied to establish a monetary value to the ecosystem services provided by these four categories. Within TEEB, priority has been given to publications and results from the European Union in order to remain as close as possible to the environmental context of the Hauts-de-France Region. An exception has been made for "grasslands" since there has been only one study for Europe, located in Spain. Due to the difference in the definition of "grasslands" between Spain and France, the global value has been preferentially considered.

The monetary values that have been retained were as follows:

- Urban: 6111 USD/ha/year,
- Cultivated and/or arable lands: 2140 USD/ha/year,
- Grasslands: 2789 USD/ha/year,

• Forests: 3789 USD/ha/year.

The human-to-ecosystem service value ratio index (H-ESV) was computed as follows:

$$\text{H} - \text{ESV} \ = \ \frac{\sum_{i=1}^{n} \text{area}_i \times \text{Monetary Value}_i}{\text{Total population}}$$

where i, $1 \leq i \leq 4$, corresponding to the 4 land cover categories, area is expressed in hectare, the monetary value in USD/ha/year and the total population data comes from the 2014 population census.

2.7. Shannon-Weaver index

This section concerns the calculation of the Shannon-Weaver index [7], applied to (i) the representation of socio-professional categories (Table 1: the Shannon-Weaver index on the representation of Socio-Professional Categories "EC_SHANSPC") and (ii) the representation of medical and paramedical professions (Table 4: the Shannon-Weaver index on the representation of medical and paramedical professions "HE_SHANMED"). The Shannon-Weaver index was computed as follows:

$$\text{H}' = \sum_{i=1}^{S} p_i \times \log(p_i), \ \text{with } p_i = \ \frac{n}{N}$$

Where S is the total number of categories, p_i is the proportional abundance or the percentage of importance of the category, n is the number of individuals in a category and N is the total number of individuals in all categories.

Shannon Weaver's Index allows a stand to be described by taking into account the number of categories (categories richness) and their relative importance within those categories (categories equitability). The value of the index ranges from 0 (*i.e.* a single category, or one category dominating all others) to log S (*i.e.* all categories are equally abundant).

2.7.1. Calculation of Shannon-Weaver index on the representation of Socio-Professional Categories (EC_SHANSPC)

The data used for this calculation came from an official inventory of the Socio-Professional Categories (SPC) located in each municipality (INSEE). This functional job analysis was provided at workers' home address and each class is presented below (in EC_SHANSPC, S = 15):

• Public administration: jobs related to the sovereign and administrative activities of the State and local authorities, excluding health and education services. In particular, this category includes all the professions of public security and justice,
• Agriculture and fisheries: all occupations directly involved in agricultural production, fishing or forestry,
• Building and public works: all occupations directly involved in the construction of buildings and public works,
• Business-to-business trade: occupations directly involved in wholesale trade and trade between businesses, whether for purchase or sale,
• Design/research: in industry, these professions are involved in the phases prior to manufacturing. This category differs from the function of intellectual services by the innovation dimension included in the work of the trades concerned,
• Culture/leisure: cultural and leisure occupations, including but not limited to sportsmen and women;
• Distribution: all professions involved in sales to private individuals, including commercial craftsmanship,
• Education/training: occupations in school and university education (primary, secondary and higher education) and vocational training, including the organisation of such education. This

function does not include sports or leisure activity leaders, who are included in the culture/leisure category,
- Maintenance/repair: professions primarily oriented towards upkeep and maintenance (excluding building and public works), as well as waste treatment (and thus by extension, the environment),
- Manufacturing: all professions consisting of implementing technical equipment or processes, excluding agriculture and fishing and construction and public works. For the most part, these are occupations that contribute directly to the various stages of the production of material goods and energy,
- Management: professions in business management, banking and insurance,
- Transport/logistics: professions in the transport of people and the flow of goods,
- Intellectual services: professions providing specific knowledge for consultancy, analysis, expertise, *etc.*,
- Health/social work: health and social work professionals, including pharmacists,
- Community services: professions providing everyday services (excluding distribution, transport, education and health).

2.7.2. Calculation of Shannon-Weaver index on the representation of medical and paramedical professions (HE_SHANMED)

The data used for this calculation came from an official inventory of the 24 medical and paramedical professions located in each municipality (INSEE). The information for all medical and paramedical specialities was provided at the professional activity location. The following medical and paramedical professions were considered (in HE_SHANMED, S = 24):

- the general practitioners,
- the cardiology specialists,
- the specialists in dermatology venereology,
- the specialists in gynaecology,
- the specialists in gastroenterology hepatology,
- the specialists in psychiatry,
- the specialists in ophthalmology,
- the specialists in otorhinolaryngology,
- the specialists in paediatrics,
- the specialists in pulmonology,
- the specialists in x-ray diagnosis,
- the specialists in stomatology,
- the dental surgeons,
- the midwifes,
- the nurses,
- the masseur-physiotherapists,
- the speech therapists,
- the orthoptists,
- the chiropodists,
- the audioprosthetists,
- the occupational therapists,
- the psychometricians,
- the dieticians,
- the psychologists.

CRediT Author Statement

Delphine Brousmiche: Software, Formal analysis, Investigation, Writing – original draft; **Michaël Genin:** Methodology, Software, Writing – review & editing; **Florent Occelli:** Conceptu-

alization, Methodology, Writing – review & editing; **Lukas Frank:** Methodology, Software, Investigation; **Annabelle Deram:** Conceptualization, Validation, Writing – review & editing; **Damien Cuny:** Conceptualization, Validation, Writing – review & editing; **Caroline Lanier:** Conceptualization, Validation, Writing – original draft, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgments

References

[1] M. Dekkers, N. Loutas, M. De Keyzer, S. Goedertier, La qualité des données et métadonnées ouvertes. Module de formation 2.2, 2013 https://joinup.ec.europa.eu/sites/default/files/document/2015-05/d2.1.2_training_module_2.2_open_data_quality_v1.00_fr.pdf.

[2] WHO, Urban Green Spaces and Health - a Review of Evidence, World Health Organization Regional Office for Europe, Copenhagen, 2016 http://www.euro.who.int/__data/assets/pdf_file/0005/321971/Urban-green-spaces-and-health-review-evidence.pdf?ua=1.

[3] S.M. Labib, S. Lindley, J.J. Huck, Spatial dimensions of the influence of urban green-blue spaces on human health: a systematic review, Environ. Res. 180 (2020) 108869, doi:10.1016/j.envres.2019.108869.

[4] R.C. Estoque, Y. Murayama, Social–ecological status index: A preliminary study of its structural composition and application, Ecol. Indic. 43 (2014) 183–194, doi:10.1016/j.ecolind.2014.02.031.

[5] R. de Groot, L. Brander, S. van der Ploeg, R. Costanza, F. Bernard, L. Braat, M. Christie, N. Crossman, A. Ghermandi, L. Hein, S. Hussain, P. Kumar, A. McVittie, R. Portela, L.C. Rodriguez, P. ten Brink, P. van Beukering, Global estimates of the value of ecosystems and their services in monetary units, Ecosyst. Serv. 1 (2012) 50–61, doi:10.1016/j.ecoser.2012.07.005.

[6] S. van der Ploeg, D. de Groot, Y. Wang, The TEEB Valuation Database: Overview of Structure, Data and Results - Final report, Foundation for Sustainable Development, Wageningen, the Netherlands, 2010 https://www.es-partnership.org/wp-content/uploads/2016/06/ESVD.-TEEB_Database_Report.pdf.

[7] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423, doi:10.1002/j.1538-7305.1948.tb01338.x.