



**Observatoire de la qualité
des environnements intérieurs**

Documentation

Relative au jeu de données
CNL2

Décembre 2025
SAM-IDE-2025-062



Mises à jour du document

Le 19/12/25 par rapport à la version mise à disposition sur le site data.gouv le 14/12/2024

>Principales modifications effectuées :

- ajout des versions du jeu de données ;
- ajout d'un DOI pour citer le jeu de données ;
- homogénéisation des noms de fichiers ;
- reformulations dans la section 3

Préambule

L'Observatoire de la Qualité des Environnements Intérieurs (OQEI), est piloté par l'Agence Nationale de Sécurité Sanitaire de l'alimentation, de l'environnement et du travail (Anses) et par le Centre Scientifique et Technique du Bâtiment (CSTB). Dans le cadre du programme de travail de l'OQEI, l'action du présent rapport a été financée par la DGPR, l'Ademe, la DGS et la DHUP.

Toute utilisation ou modification de ce rapport par des tiers, sous quelque forme que ce soit, est faite sous leur seule et entière responsabilité, sans que celle de l'OQEI ne puisse être recherchée. Toute reproduction ou représentation intégrale ou partielle, par quelque procédé que ce soit, des pages publiées dans le présent document, faite sans l'autorisation de l'Anses et du CSTB est illicite et constitue une contrefaçon. Seules sont autorisées, d'une part, les reproductions strictement réservées à l'usage du copiste et non destinées à une utilisation collective et, d'autre part, les analyses et courtes citations justifiées par le caractère scientifique ou d'information de l'œuvre dans laquelle elles sont incorporées (Loi du 1er juillet 1992 - art. L 122-4 et L 122-5 et Code Pénal art. 425).

L'OQEI décline toute responsabilité dans les conclusions et avis qui pourraient être associés à la réutilisation des données mises à disposition, qui n'engagent que leurs auteurs.

Auteurs

DESVIGNES Virginie – CSTB DSAM/IDE

GREGOIRE Anthony – CSTB DSAM/IDE

Relecteurs

RAMALHO Olivier – CSTB DSAM/IDE

Citation suggérée pour le présent document

DESVIGNES Virginie, GREGOIRE Anthony, RAMALHO Olivier. 2025. Observatoire de la Qualité des Environnements Intérieurs – Documentation relative aux jeux de données CNL2. France : OQEI

Citation suggérée du jeu de données

RAMALHO Olivier, DASSONVILLE Claire, GREGOIRE Anthony, SIVANANTHAM Sutharsini, LAFAURIE Emma, DESVIGNES Virginie, Version V1.0.0 of Dataset published 2025 via OQEI
<https://doi.org/10.82199/oqei-cn2-data-6bdt-pb04>

Mots clés

Enquête nationale, polluant, résidentiel, COV, COSV, particules, PM_{2,5}, formaldéhyde, perturbateurs endocriniens, phtalates, HAP, PCB, PBDE, NO₂, air intérieur, composés organiques volatils, radon, foyer, dioxyde d'azote

National survey, pollutant, dwellings, VOCs, particles, particulate matter, PM_{2,5}, formaldehyde, endocrine disruptors, phthalates, PAHs, PCBs, PBDEs, NO₂, indoor air, volatile organic compounds, radon, household, nitrogen dioxide

Commenté [VD1]: A revoir :
indoor air quality, occupant characteristics,
French dwelling, building, national campaign,
domestic activities, equipment

Table des matières

Abréviations	6
1 Contexte et objectif de la seconde campagne nationale logements	7
2 Liste des fichiers mis à disposition	7
3 Description du jeu de données CNL2	8
3.1 Données relatives aux concentrations de polluants dans l'air	9
3.2 Données relatives aux caractéristiques des logements et des ménages	9
3.3 Données relatives aux caractéristiques des individus	10
3.4 Données relatives aux caractéristiques des pièces	10
4 Utilisation du jeu de données CNL2	11
4.1 Import des fichiers avec le logiciel R	11
4.2 Utilisation des données avec le logiciel R	12
4.3 Utilisation des données avec le logiciel SAS	12
5 Conclusion	12
6 Bibliographie	12

Abréviations

CNL2	Seconde Campagne Nationale Logements
JDD	Jeu de données
LDM	Limite de détection de la méthode
LQM	Limite de quantification de la méthode
OQAI	Observatoire de la qualité de l'air intérieur
OQEI	Observatoire de la qualité des environnements intérieurs
QAI	Qualité de l'air intérieur

1 Contexte et objectif de la seconde campagne nationale logements

L'Observatoire de la qualité de l'air intérieur (OQAI) présente les résultats de la deuxième campagne nationale sur la qualité de l'air dans les logements en France. Elle couvre la période de novembre 2020 à février 2023. La campagne associe des mesures de qualité de l'air au domicile des participants sur une semaine de 7 jours à une enquête sur les caractéristiques des logements, des ménages et des équipements présents, complétée par la fréquence des activités domestiques et la perception des occupants sur leur confort. Le recueil d'un auto-questionnaire relatif à la santé respiratoire et allergique des occupants pour exploitation par Santé publique France clôture l'enquête.

Les paramètres mesurés ont été choisis sur la base d'une hiérarchisation sanitaire mise à jour d'après leur dangerosité et de leur occurrence, mais également au regard de leur impact sur le confort. Plus de 170 polluants et paramètres sont ainsi mesurés dans l'air des logements pour cette campagne nationale : composés organiques volatils (dont aldéhydes et certaines substances considérées émergentes), composés organiques semi-volatils (parmi lesquels hydrocarbures aromatiques polycycliques, phtalates, retardateurs de flamme bromés, polychlorobiphényles, alkylphénols, muscs), pesticides, particules fines, dioxyde d'azote, dioxyde de carbone, radon, température, humidité relative, débit d'air ou pression aux bouches.

A la fin du terrain de l'enquête, 571 logements ont été enquêtés. Un total de 1516 individus a répondu aux questions de l'enquête. Les logements enquêtés sont répartis dans 321 communes et 84 départements. L'élaboration d'un jeu de pondération permet d'exprimer les résultats observés à l'échelle du parc de 29,7 millions de résidences principales en France métropolitaine continentale.

2 Liste des fichiers mis à disposition

Huit fichiers sont mis à disposition en dehors de la présente note :

- le fichier « **cnl2_data_polluant_v1.0.0_2024-12-02.csv** » relatif aux mesures des polluants dans l'air ;
- le fichier « **cnl2_dico_polluant_v1.0.0_2025-12-17.xlsx** » correspondant au dictionnaire des variables du fichier précédent ;

- le fichier « **cnl2_data_logement-menage_v1.0.0_2025-12-17.csv** » relatif aux caractéristiques des logements et des ménages ;
- le fichier « **cnl2_dico_logement-menage_v1.0.0_2025-12-17.xlsx** » correspondant au dictionnaire des variables du fichier précédent ;

- le fichier « **cnl2_data_individu_v1.0.0_2024-12-02.csv** » relatif à la description des individus du logement ;
- le fichier « **cnl2_dico_individu_v1.0.0_2024-12-02.xlsx** » correspondant au dictionnaire des variables du fichier précédent ;

- le fichier « **cnl2_data_piece_v1.0.0_2024-12-02.csv** » relatif à la description des pièces du logement ;
- le fichier « **cnl2_dico_piece_v1.0.0_2024-12-02.xlsx** » correspondant au dictionnaire des variables du fichier précédent ;

Les dictionnaires décrivent les variables présentes dans le jeu de données et fournissent des informations sur le code, le type, le libellé et les modalités ou l'intervalle numérique possible ainsi que l'enchaînement logique amenant à la question/variable quand celle-ci est issue d'un questionnaire.

3 Description du jeu de données CNL2

Les données couvrent un échantillon de 571 logements répartis dans le parc de résidences principales en France métropolitaine continentale.

Les données ont été collectées dans le cadre de deux visites au domicile des occupants, espacées d'une semaine. Les enquêtes ont été réalisées tout au long de l'année sans favoriser une saison plutôt qu'une autre. Une seule enquête (incluant la campagne de mesures) a été menée dans chaque logement.

Selon les variables, les données sont exprimées à différentes échelles : le logement, une pièce du logement, un individu du logement.

Chaque logement ou foyer est rattaché géographiquement au niveau départemental. Le niveau infra-départemental n'est pas disponible.

Chaque foyer a exprimé librement son consentement écrit et signé à participer à l'enquête après présentation détaillée du processus d'enquête.

3.1 Données relatives aux concentrations de polluants dans l'air

Le fichier « **cnl2_data_polluant_v1.0.0_2024-12-02.csv** » contient la table des données relatives aux concentrations de polluants dans l'air.

Cette table est au format *tidy long* et se compose de 66 808 lignes et 20 colonnes.

Après la ligne d'en-tête indiquant le nom des champs, chaque ligne représente le résultat d'analyse d'un polluant pour un logement. Chaque colonne correspond à une variable décrite dans le dictionnaire de variables « **cnl2_dico_polluant_v1.0.0_2025-12-17.xlsx** ».

Les variables sont regroupées comme suit :

- Identification du logement et de la pièce de mesure : CODELOG, CODEINTEXT, CODEPIECE ;
- Identification de la substance mesurée : FAMILLEPARAM, NOMPARAM , NUMCAS, CODEPARAM ;
- Prélèvements des échantillons d'air dans les logements : DUREEPRELEV, DEBITPRELEV, UNITEDUREE, UNITEDEBIT ;
- Méthodes et résultats analytiques : LDM, LQM , INFLDM, INFLQM, CONCENTRATION_NUM, CONCENTRATION_EXPR, UNITELDMLQM, UNITECONC, EXPLOITABLE.

La censure a été traitée dans la variable CONCENTRATION_NUM. Lorsque les concentrations sont inférieures à la LDM, la valeur de concentration indiquée est LDM/2. Lorsque les concentrations sont comprises entre LDM et LQM, les valeurs brutes fournies par le laboratoire ont été conservées.

3.2 Données relatives aux caractéristiques des logements et des ménages

Le fichier « **cnl2_data_logement-menage_v1.0.0_2025-12-17.csv** » contient la table des données relatives aux caractéristiques des logements et des ménages.

Cette table est au format *tidy wide* et se compose de 572 lignes et 227 colonnes.

Après la ligne d'en-tête indiquant le nom des champs, chaque ligne représente une observation par logement. Chaque colonne correspond à une variable décrite dans le dictionnaire de variables « **cnl2_dico_logement-menage_v1.0.0_2025-12-17.xlsx** ». Les variables sont regroupées comme suit :

- Identification du logement : CODELOG ;
- Plan de sondage : POIDS_SONDAGE, NUM_UP, CODEEGION, QPV ;
- Environnement du logement : REG, Libelle.region, DEP, Libelle.departement, XPB10, XPB11, XPB111, XPB112, XPB113, XPB114, XPB115, XPB11b, XPB12, XPB131, XPB132, XPB133, XPB134, XPB135, XPB136, XPB137, XPB138, XPB139, XPB13a ;
- Enquête : DVISIT1, DVISIT2, KODSJ, KODCH ;

- Caractéristiques du logement : PCTL, PCPP, PCPS, PCNE, PCEL, PCOL, PCAS, PCJA, IAC, IAC1, IAC2, MATER01, MATER02, MATER03, MATER04, MATER05, MATER06, MATER07, MATER1, REAB, REAB1, REAB2, REAB3, TRAV, TRAVI01, TRAVI02, TRAVI03, TRAVI04, TRAVI05, TRAVI06, TRAVI07, TRAVI08, TRAVI09, TRAVI10, TRAVI11, TRAVI12, TRAV2, KCC0, KCC1, KCC1a, KCC201, KCC202, KCC203, KCC204, KCC205, KCC206, KCC207, KCC208, KCC209, KCC210, KCC2a, KCCM11, KCCM12, KCCM13, KCCM14, KCCM15, KCCM16, KCCM17, KCCM18, KCCM19, KCCM2, CHEM1, CHEM2, KCL1, KCL1b, KCU1, KCU1a, KECS, KEIC, KEIC1, CEM01, CEM02, CEM03, CEM04, LSTU2, LSTU, LSTUC, LNPL1, LNPL2, LNPL3, LNPL4, LNPL5, LNPL6, LNPL7, LNPL8, LNPL9, LNPL0, HCU1, HCU3, HSRF, HPLB, HPLB1, HPLB2, HPLH, HPLH1, DCA1, DCA2, DGAR0, DGAR1, DGG3, FNSJ11, FNSJ12, FNSJ13, FNSJ14, FNSJ15, FNSJ16, FNSJ1a, FNSJ1b, FNCHI2, FNCHI3, FNCHI4, FNCHI5, FNCHI6, FNCHIa, SURFCHB, SURFSEJ, HAUTSEJ ;
- Description et aux habitudes du ménage : MAIANO, MAIAMOIS, MAIAYEAR, ACTP1, ACTP2, ACTP3, QPX1, QPX2, QPX3, QOD1, QOD2, QOD3, QBR1, QBR2, QBR3, QBR4, QBR5, QBR6, CHAUS, FUMEUR1, FUMEUR2, ANI1, ANI211, ANI212, ANI213, ANI214, ANI215, ANI22, QATRA1, QOM31, QOM32, QOM33, QOM34, QOM35, HGTM, HGLP, MENAG, KID15, STRMEN, REVI, FC10B, SOMSEJNO, SOMSEJADUADO, SOMSEJENFANT, SOMCHBNO, SOMCHBADUADO, SOMCHBENFANT, CHAUF, CHEM, CHAPP, AFUM, AVAP, CUI1, CUI4, CUI5, PRESS, MBLS1, MBLS2, MBLS3, MBLS4, MBLS4a, Q1, Q2, Q2INOW, Q2IAVANT, Q3, Q4NETTOI, Q4AERE, Q4CUI, Q4LOISIR, Q4TRAV, Q4MOB, Q5CHAUF, Q5NETTOI, Q5DESINF, Q5DESO, Q5INSEC, Q5TABAC.

3.3 Données relatives aux caractéristiques des individus

Le fichier « **cnl2_data_individu_v1.0.0_2024-12-02.csv** » contient la table des données relatives aux caractéristiques des individus.

Cette table est au format *tidy wide* et se compose de 1535 lignes et 11 colonnes.

Après la ligne d'en-tête indiquant le nom des champs, chaque ligne représente une observation par individu. Chaque colonne correspond à une variable décrite dans le dictionnaire de variables « **cnl2_dico_individu_v1.0.0_2024-12-02.xlsx** ». Les variables sont regroupées comme suit :

- Identification du logement et de l'individu : CODELOG, NUM_OCCUPANT ;
- Caractéristiques de l'individu : LIEN, SEXE, AGENO, AGEINF1, AGEANS, AGEMOIS, AGECAT, DIPLOM, NOCCUA.

3.4 Données relatives aux caractéristiques des pièces

Le fichier « **cnl2_data_piece_v1.0.0_2024-12-02.csv** » contient la table des données relatives aux caractéristiques des pièces.

Cette table est au format *tidy wide* et se compose de 5489 lignes et 21 colonnes.

Après la ligne d'en-tête indiquant le nom des champs, chaque ligne représente une observation par pièce. Chaque colonne correspond à une variable décrite dans le dictionnaire de variables « **cnl2_dico_piece_v1.0.0_2024-12-02.xlsx** ». Les variables sont regroupées comme suit :

- identification du logement et de la pièce : CODELOG, IDN_PIECE, CODEPIECE ;
- Situation de la pièce : Q11 ;
- Type d'appareil contenu dans la pièce : SQ41, Q41, Q42, EP01, EP02, Q91, Q121 ;
- Caractéristiques des ouvrants et menuiserie : Q7A51, Q7A52, Q7A53, Q7A54, Q7A55, Q7A56, Q7A5b, Q7A6, Q7A7 ;
- Caractéristiques de la pièce : Q57.

4 Utilisation du jeu de données CNL2

La campagne de mesures a été élaborée pour être représentative de la qualité de l'air intérieur (QAI) au niveau national, c'est-à-dire représentative de la QAI dans le parc de résidences principales en France métropolitaine continentale. Un jeu de pondération a été calculé pour exprimer les résultats à ce niveau de représentativité. **Les données ne peuvent en aucun cas être utilisées pour des comparaisons géographiques départementales ou régionales car elles ne sont pas représentatives de ces échelles géographiques.**

4.1 Import des fichiers avec le logiciel R

```
read.csv(CNL2_DATA_POLLUANT_V1.0.0_2024-12-02.csv ",
  sep = ";", fileEncoding = "UTF-16LE")
read.csv(CNL2_DATA_MENAGES_V1.0.0_2024-12-02.csv ",
  sep = ";", fileEncoding = "UTF-16LE")
read.csv(CNL2_DATA_INDIVIDUS_V1.0.0_2024-12-02.csv ",
  sep = ";", fileEncoding = "UTF-16LE")
read.csv("CNL2_DATA_PIECE_V1.0.0_2024-12-02.csv ", sep = ";", fileEncoding = "UTF-16LE")
```

4.2 Utilisation des données avec le logiciel R

L'utilisation correcte des données dans le logiciel R (R Core Team 2021) est possible grâce au package « Survey » (Lumley 2020). La déclaration du plan de sondage se fait ainsi :

```
plan <- svydesign(ids = ~ NUM_UP, data = data_cnl2 , strata = ~ `CODE_REGION` + QPV,  
weights = POIDS_SONDAGE)
```

Les résultats doivent être produits à l'aide des fonctions du package « Survey ».

4.3 Utilisation des données avec le logiciel SAS

L'utilisation correcte des données avec le logiciel SAS (SAS software. Copyright, SAS Institute Inc., Cary, NC, USA) est possible en indiquant dans les procédures SURVEY les lignes suivantes :

```
CLUSTER NUM_UP;  
STRATA 'CODE_REGION'n QPV;  
WEIGHT POIDS_SONDAGE;
```

5 Conclusion

Les données mises en ligne ont été arrêtées au 17 décembre 2025. Le jeu de données CNL2 suit un schéma de gestion des versions sémantique (SemVer). La version disponible ici est la V1.0.0. La mise à disposition est un processus continu qui ne manquera pas d'évoluer avec de nouvelles versions des données mises à jour au fur et à mesure de l'avancement des exploitations statistiques que l'OQEI mène.

6 Bibliographie

Lumley, T. 2020. 'Survey: Analysis of Complex Survey Samples. R Package'.

R Core Team. 2021. 'R: A Language and Environment for Statistical Computing'.