



Description des données

Niveau de formation – Géodonnées

Édition 2025

Editeur : Office fédéral de la statistique (OFS)
Renseignement : geostat@bfs.admin.ch
Rédaction : BILD-S, GEO, POP
Domaines : 00 Bases statistiques, 02 Population

Concept de mise en page : Section PUB
Download : www.statistik.ch
Copyright : OFS, Neuchâtel 2025
La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée

Table des matières

1.	Introduction	3
2.	Structure spatiale	3
3.	Méthodologie	3
4.	Qualité des données et protection des données	4
5.	Structure des données	5
6.	Description des variables	6
7.	Applications	7
8.	Contact	7
9.	Références	7

1. Introduction

Le produit « Niveau de formation - Géodonnées » fournit des informations à fine résolution spatiale pour l'ensemble du territoire suisse sur le niveau de formation le plus élevé de la population résidante permanente âgée de 25 à 64 ans. Les données proviennent de plusieurs enquêtes statistiques menées par l'administration fédérale et sont complétées par un modèle d'imputation en deux étapes qui estime les données manquantes (OFS, 2025f).

Le niveau de formation est exprimé dans les géodonnées selon une nomenclature à trois niveaux (OFS, 2025a) :

- Ecole obligatoire ;
- Degré secondaire II (formation professionnelle ou générale) ;
- Degré tertiaire (formation professionnelle supérieure ou hautes écoles).

Les données sont disponibles pour les années de référence 1990 et 2000, puis pour chaque année à partir de 2017. A chaque publication, l'état le plus actuel ainsi qu'une série chronologique couvrant toutes les années disponibles sont mis à disposition. L'année de référence la plus récente est antérieure d'au moins deux ans à l'année de publication. La mise à jour des données est effectuée chaque année.

Les ensembles de données sont fournis sous forme de fichiers CSV avec coordonnées et sous forme de GeoPackage dans le système de référence suisse MN95 (EPSG :2056) et peuvent être directement importés et traités dans des systèmes d'information géographique (SIG).

2. Structure spatiale

2.1 Zone couverte

Le produit couvre l'ensemble du territoire suisse. La résolution à l'hectares permet d'effectuer des analyses du niveau de formation à petite échelle ; parallèlement, des informations agrégées sont disponibles au niveau des régions USPAT1 (« unités spatiales statistiques de base de premier niveau ») et des communes.

2.2 Entités spatiales

Pour la diffusion des données, les agrégations spatiales par hectare, par régions USPAT1 et par communes sont préparées sur la base des données relatives au niveau de formation, dans la résolution spatiale la plus fine – selon les coordonnées métriques des bâtiments (cf. chapitre 3.3 Géocodage) –.

Hectares

Chaque hectare est identifié par les coordonnées de son point d'angle sud-ouest (« E_KOORD » / « N_KOORD » selon MN95). Les données sont publiées sous forme de grille de 100 × 100 mètres. Les valeurs indiquées proviennent de l'agrégation des enregistrements individuels dont les coordonnées sont localisées dans une cellule de la grille correspondant à un hectare.

Régions USPAT1

Pour l'ensemble de la Suisse, 3607 USPAT1 ont été définies (OFS, 2025b). L'agrégation des informations sur le niveau de formation dans ces unités spatiales permet des analyses régionales plus détaillées que celles réalisables au niveau communal.

Communes

L'agrégation des données au niveau communal s'effectue sur la base des limites communales du produit swissBOUNDARIES3D (swisstopo, 2025). Les limites communales pouvant évoluer au fil du temps, ce sont celles en vigueur au 1er janvier de l'année de publication qui sont utilisées.

2.3 Système de coordonnées

Toutes les coordonnées sont indiquées dans le système de coordonnées suisse MN95 (EPSG :2056). Les géométries des régions USPAT1 et des communes utilisées pour l'agrégation sont également disponibles dans ce cadre de référence.

3. Méthodologie

Les présentes géodonnées résultent d'un processus de production en plusieurs étapes. Celui-ci comprend la combinaison de différentes sources de données, le traitement des valeurs manquantes (imputation) et le géocodage à l'aide du registre des bâtiments et des logements.

3.1 Sources

Le niveau de formation de la population résidante permanente est déterminé chaque année à partir de plusieurs sources (OFS, 2025f).

Les différentes sources de données sont vérifiées et combinées afin d'obtenir une image aussi complète que possible des niveaux de formation. Voir OFS 2025f pour plus d'informations.

3.2 Imputation et valeurs manquantes (misings)

Le niveau de formation des personnes pour lesquelles aucune donnée n'est directement disponible est imputé sur la

base de l'enquête structurelle et de variables supplémentaires issues de différents registres. Au total, environ 10 % des données relatives au niveau de formation achevé le plus élevé sont imputées (OFS, 2025f) .

Depuis l'année de référence 2017, le niveau de formation de chaque personne est disponible soit via une source (environ 90 %), soit par imputation (environ 10 %). Pour les années de référence 1990 et 2000, il existe quelques valeurs manquantes ou « missings » qui ne peuvent pas être imputées. Les données de 1990 ou 2000 ne sont publiées par entité spatiale que si le niveau de formation peut être déterminé pour plus de 50 % des personnes.

3.3 Géocodage

Les informations sur le niveau de formation sont géocodées sur la base du registre fédéral des bâtiments et des logements (RegBL). Outre les coordonnées des bâtiments, le RegBL contient l'identifiant fédéral des bâtiments (EGID) (OFS, 2025c) qui figure également dans les registres des personnes. Cela permet de localiser les personnes – et donc aussi leur niveau de formation – dans l'espace.

Les coordonnées initiales du RegBL proviennent des coordonnées de bâtiment relevées de manière exhaustive pour les recensements de la population de 1990 et 2000. Elles sont complétées par les coordonnées des bâtiments relevés dans le cadre des recensements des entreprises des secteurs secondaire et tertiaire entre 1995 et 2008 et des bâtiments des exploitations agricoles recensées entre 1996 et 2008. Depuis 2010, les coordonnées du RegBL sont relevées en continu, complétées et, si nécessaire, révisées et optimisées en collaboration avec la mensuration officielle.

Optimisation grâce aux données actuelles du RegBL

Le géocodage dans le RegBL étant en constante amélioration, les coordonnées des bâtiments de l'année de publication sont utilisées. La condition préalable est que les données de formation puissent être attribuées à un bâtiment (EGID) et que les coordonnées actuelles du RegBL (année de publication) ne s'écartent pas de plus de 200 m des coordonnées valables pour l'année de collecte des données de formation. Si la différence est supérieure à 200 m, les coordonnées initiales de l'année de collecte sont conservées, car dans de tels cas, il peut y avoir non seulement des améliorations du géocodage, mais aussi des corrections.

Pour l'année de collecte 1990, l'attribution des personnes aux bâtiments a été effectuée à l'aide du VZ-CODE (identificateur de bâtiment du recensement de 1990 composé du numéro de commune + circonscription électorale + numéro de bâtiment) et associé à l'EGID actuel à l'aide d'une table de correspondance, car l'EGID n'existait pas encore à l'époque.

Attribution des communes et des USPAT

Les informations sur la commune et l'USPAT proviennent également en grande partie du RegBL (état de l'année de publication). Dans les cas où un lien direct avec les données

actuelles RegBL à l'aide de l'EGID n'est pas possible (en particulier pour les années de référence 1990 et 2000), ces informations sont complétées par un recoupement spatial avec les géométries des communes (selon swissBOUNDARIES3D) et des régions USPAT1 valables pour l'année de publication.

Hectares « collectifs »

Les personnes sans adresse dans la commune ou celles qui ne sont pas rattachées à un bâtiment - et donc à un EGID - ne peuvent être localisées et donc géocodées. Afin de garantir l'exhaustivité du recensement, ces personnes ont été attribuées, par commune, aux coordonnées dites « centrales » de la commune (hectares collectifs). Les données relatives à ces hectares collectifs peuvent avoir des influences indésirables dans les analyses à petite échelle ou fausser les résultats. Elles ont donc été supprimées de la grille d'hectares et des régions USPAT1. Elles sont toutefois incluses dans l'agrégation communale.

4. Qualité des données et protection des données

4.1 Protection des données

Afin de protéger les informations relatives aux individus, les règles suivantes s'appliquent à la publication des données sur le niveau de formation le plus élevé (trois niveaux : école obligatoire, degré secondaire II, degré tertiaire) :

Au moins 10 personnes par entité

Les données ne sont pas publiées si une entité spatiale (hectare, commune ou USPAT1) compte moins de 10 personnes âgées de 25 à 64 ans.

Formation de classes

Le nombre de personnes par niveau de formation n'est pas fourni sous forme de valeurs absolues, mais en pourcentages du nombre total de personnes par unité géographique. Ces pourcentages sont ensuite regroupés en cinq classes selon la modalité suivante :

- 1 = 0–<20 % ;
- 2 = 20–<40 % ;
- 3 = 40–<60 % ;
- 4 = 60–<80 % ;
- 5 = 80–100 %.

Le nombre de personnes âgées de 25 à 64 ans est également indiqué par classes, avec une résolution par hectare :

- 1 = 10–19 personnes ;
- 2 = 20–29 personnes ;
- 3 = 30–49 personnes ;
- 4 = 50–119 personnes ;

- 5 = plus de 120 personnes.

Pourcentages pour l'USPAT1 et les communes de 100 habitants ou plus

Les pourcentages relatifs aux trois niveaux de formation sont publiés si une région USPAT1 ou une commune compte au moins 100 personnes («BBTOT25_64») et si celles-ci ne possèdent pas toutes le même niveau de formation. Dans le cas contraire, seules les données par classe sont publiées.

4.2 Qualité des données

Les points suivants décrivent les principaux aspects liés à la qualité des données.

- **Stabilité spatiale et temporelle**

L'utilisation de grilles d'un hectare et de régions USPAT1 garantit une identification claire et stable des objets dans le temps.

- **Couverture fine (accent sur les zones urbaines)**

Dans les zones rurales, les données de nombreux hectares ne sont pas publiées en raison de la faible densité de population et des mesures de protection des données ; les publications au niveau des hectares concernent donc principalement les zones urbaines.

- **Données agrégées par hectare**

Une agrégation par commune ou USPAT1 des données publiées à l'hectare ne peut pas être comparées aux données publiées pour les communes ou les USPAT1, car les données des hectares regroupant moins de 10 personnes ne sont pas publiées. Les valeurs USPAT1 agrégées par communes correspondent aux données communales après soustraction des données des hectares collectifs (sauf si d'autres mesures de protection des données s'appliquent).

- **Imputation**

Environ 10 % des valeurs relatives au niveau de formation sont imputées. L'imputation n'est pas répartie de manière aléatoire dans l'espace : certains hectares peuvent contenir exclusivement des valeurs imputées, tandis que les proportions au niveau communal et USPAT1 peuvent atteindre environ 50 %.

- **Données manquantes (missings)**

À partir de 2017, une donnée sur le niveau de formation le plus élevé est disponible pour chaque personne, soit à partir d'une source, soit imputée. Pour les années de recensement 1990 et 2000, il existe des données manquantes («BBTOT25_64_MISS»). Seules les valeurs pour lesquelles au moins la moitié des données de la population résidente («BBTOT25_64») de l'entité sont disponibles sont publiées. En plus des mesures de protection des données, cela entraîne la suppression de quelques hectares supplémentaires. Pour l'année de référence 2000, aucune donnée sur le niveau

de formation n'est ainsi disponible pour la commune d'Héremence (numéro OFS 6084) et pour certaines régions USPAT1.

5. Structure des données

5.1 Fichiers et noms de fichiers

Pour chaque niveau d'agrégation spatiale – hectares, communes et régions USPAT1 –, les géodonnées actuelles ainsi que des fichiers de séries chronologiques avec toutes les années de référence disponible sont fournis.

Formats

- **CSV (séparateur « ; »)**

Contient les variables avec les coordonnées en hectares (maillage de 100 m) dans le MN95. La paire de coordonnées « E_KOORD » / « N_KOORD » représente le point d'angle sud-ouest d'une cellule dans la résolution en hectares. Pour chaque cellule, les valeurs indiquées se rapportent à la totalité de la surface de 100 × 100 m.

- **GeoPackage (GPKG)**

Contient, en plus des variables, les géométries complètes : polygones pour les cellules d'un hectare, les régions USPAT1 et les communes. Les géométries généralisées K4 (OFS, 2025d) sont utilisées pour les communes. Les variables dans le GPKG du fichier de série chronologique sont réparties par année de recensement (par exemple 1990, 2000, 2017+). Chaque variable se présente donc sous la forme NOM_DE_LA_VARIABLE_ANNÉE, par exemple COMPULSORY_2017, SECONDARYII_2019 ou IMPUTATION_2000.

Noms de fichiers

Les noms de fichiers sont conçus pour être aussi explicites que possible.

Exemples pour l'année de référence actuelle :

- *bildungsstand_YYYY_AGGR_2056.csv*
- *bildungsstand_YYYY_2056.gpkg* (contient des couches pour les hectares, les régions USPAT1 et les communes)

où :

- « YYYY » : année de référence « ERHJAHR » ;
- « AGGR » : niveau d'agrégation spatiale – *ha* pour les hectares, *gmde* pour les communes et *uspat* pour les régions USPAT1 ;
- « 2056 » : code EPSG du système de coordonnées (MN95).

Exemple de fichier de série chronologique :

- *bildungsstand_AGGR_2056.csv*
- *bildungsstand_2056.gpkg* (contient des couches pour les hectares, les régions USPAT1 et les communes)

où :

- « **AGGR** » : niveau d'agrégation spatiale – *ha* pour les hectares, *gmde* pour les communes et *uspat* pour les régions USPAT1 ;
- « **2056** » : code EPSG du système de coordonnées (MN95).

6. Description des variables

La liste des variables complète cette description. Elle contient notamment les types de données des variables.

6.1 Variables temporelles

- **ERHJAHR : année d'enquête**
La date de référence correspond au 31 décembre de l'année de collecte des données, également appelée année de référence.
- **PUBJAHR : année de publication**
Année de publication de l'ensemble de données. Le statut des communes et des régions USPAT1 utilisé pour l'agrégation est celui du 1er janvier de l'année de publication.

6.2 Variables territoriales

Ces variables identifient la zone pour laquelle les données sur le niveau de formation sont publiées. Les agrégations par régions USPAT1 et communes sont basées sur l'état territorial en vigueur au 1er janvier de l'année de publication.

- **RELI : clé primaire d'un hectare**
Clé primaire d'un hectare. Elle se compose des chiffres 2 à 5 des coordonnées E et N (les quatre premiers chiffres de « E_KOORD » et les quatre derniers chiffres de « N_KOORD »).
- **E_KOORD : coordonnée Est**
- **N_KOORD : coordonnée Nord**
Paire de coordonnées du point d'angle sud-ouest d'un hectare dans le système de coordonnées MN95.
- **U1_ID : code USPAT1**
Le code USPAT1 se compose du chiffre 1 suivi du numéro à un chiffre de la grande région, du numéro à deux chiffres du canton et du numéro USPAT à trois chiffres (OFS, 2025b) .

- **U1_NAME : nom USPAT1**

Les noms des USPAT1 sont repris des noms de communes ou de quartiers ou des noms figurant dans la base de données des localités swissNAMES3D de swisstopo. Afin d'éviter toute confusion avec les noms de communes, le préfixe « USPAT1 » est ajouté devant le nom.

- **GDENR : numéro de commune OFS**

Le numéro de commune OFS est unique pour chaque état communal donné depuis 1960 (date de référence). Le même numéro peut toutefois être réutilisé pour un autre état communal établi à une autre date (OFS, application des communes suisses, 2025e) .

- **GDEHISTID : numéro d'historisation de la commune**

Le numéro d'historisation de la commune est un identifiant non significatif qui permet d'identifier de manière univoque les entrées qui ont été supprimées ou ajoutées dans la liste officielle des communes.

6.3 Variables thématiques

- **BBTOT25_64**

Population résidante permanente âgée de 25 à 64 ans selon la statistique de la population et des ménages (STATPOP) pour les années de recensement à partir de 2017. Pour les années de recensement 1990 et 2000, selon le recensement fédéral de la population (RFP).

Dans la résolution par hectare, les données sont présentées par classes :

- 1 = 10–19 personnes ;
- 2 = 20–29 personnes ;
- 3 = 30–49 personnes ;
- 4 = 50–119 personnes ;
- 5 = plus de 120 personnes.

Dans les régions USPAT1 et les communes, il existe, pour des raisons de protection des données, la classe « 0–9 personnes ». Celle-ci est indiquée par la valeur 9.

- **BBTOT25_64_MISS**

Population résidante permanente âgée de 25 à 64 ans « BBTOT25_64 » sans indication du niveau de formation. Cette variable n'existe que pour les années de recensement 1990 et 2000 ; à partir de 2017, il n'y a plus de valeurs manquantes (missings).

- **COMPULSORY_C** (école obligatoire, valeurs classées)
- **SECONDARYII_C** (degré secondaire II, valeurs classées)
- **TERTIARY_C** (degré tertiaire, valeurs classées)

Proportions classées de la nomenclature à trois niveaux du niveau de formation. On distingue cinq classes :

- 1 = 0–<20 % ;
- 2 = 20–<40 % ;
- 3 = 40–<60 % ;
- 4 = 60–<80 % ;
- 5 = 80–100 %.

- **COMPULSORY** (école obligatoire, pourcentages)
- **SECONDARYII** (degré secondaire II, pourcentages)
- **TERTIARY** (degré tertiaire, pourcentages)

Pourcentages correspondant au niveau de formation le plus élevé. Les valeurs sont arrondies de sorte que leur somme soit toujours égale à 100 %. Les deux valeurs présentant les plus grands restes d'arrondi sont arrondies normalement ; la troisième valeur est déterminée comme étant la différence par rapport à 100 %.

- **IMPUTATION**

Pourcentage des valeurs imputées (OFS, 2025f) du niveau de formation le plus élevé par unité territoriale.

7. Applications

Les données relatives au niveau de formation de la population résidente permanente constituent un indicateur central du capital humain spécifique à la formation en Suisse. Compte tenu de l'évolution démographique, du vieillissement de la population et des besoins croissants en main-d'œuvre qualifiée, son analyse revêt une importance particulière.

Les géodonnées agrégées à l'hectare permettent des analyses et des visualisations avec une très fine résolution spatiale. Elles mettent en évidence les différences entre les quartiers des grandes villes et révèlent des schémas spatiaux qui restent cachés au niveau cantonal ou communal (p. ex. la proximité d'un axe de circulation ou d'une haute école). Elles permettent en outre d'étudier de manière très précise la relation avec d'autres variables telles que les caractéristiques du marché du travail ou démographiques et facilitent l'analyse des évolutions et des tendances dans le temps.

8. Contact

Nous espérons que ce produit répondra à vos besoins et sommes ouverts à toute remarque, suggestion d'amélioration ou signalement d'erreur, que vous pouvez adresser à geostat@bfs.admin.ch. Pour toute question thématique concernant le niveau de formation vous pouvez vous adresser à eduperspectives@bfs.admin.ch.

9. Références

- BFS. (2025a). *Höchste abgeschlossene Ausbildung STATPOP - Produktionsprozess, Kurzbeschreibung (do-d-15.10-EXP-01)*. Neuchâtel: Bundesamt für Statistik (BFS).
- BFS. (2025b, August). *Räumliche Gliederungen: USPAT – Raumbezogene statistische Grundeinheiten*. Récupéré sur <https://www.bfs.admin.ch/bfs/de/home/statistiken/quarterschnittsthemem/raeumliche-analysen/raeumliche-gliederungen/uspat.html>
- BFS. (2025c, August). *Eid. Gebäude- und Wohnungsregister (GWR)*. Récupéré sur <https://www.housing-stat.ch/de/index.html>
- BFS. (2025d, August). *Generalisierte Gemeindegrenzen*. Récupéré sur <https://www.bfs.admin.ch/bfs/de/home/dienstleistungen/geostat/geodaten-bundesstatistik/raumgliederungen/generalisierte-gemeindegrenzen.html>
- BFS. (2025e, September). *Applikation der Schweizer Gemeinden*. Récupéré sur <https://www.agvchapp.bfs.admin.ch>
- OFS. (2025f). *Plus haute formation achevée dans la statistique STATPOP+ – Processus et méthode de production (DO-D-15.10-EXP-01)*. Neuchâtel: Office fédéral de la statistique (OFS).
- swisstopo. (2025, August). *swissBOUNDARIES3D*. Récupéré sur <https://www.swisstopo.admin.ch/de/landschaftsmodell-swissboundaries3d>