A user-driven process for INSPIRE-compliant land use database: example from Wallonia, Belgium

Article *in* Annals of GIS · January 2021

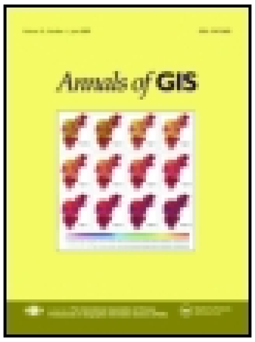
DOI: 10.1080/19475683.2021.1875047

3 authors:

Benjamin Beaumont
Service Public de Wallonie

Tais Grippa
Université Libre de Bruxelles

Moritz Lennert
Changement pour l'égalité



Annals of GIS

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tagi20

A user-driven process for INSPIRE-compliant land use database: example from Wallonia, Belgium

Benjamin Beaumont , Taïs Grippa & Moritz Lennert

ANNALS OF GIS
https://doi.org/10.1080/19475683.2021.1875047

A user-driven process for INSPIRE-compliant land use database: example from Wallonia, Belgium

Benjamin Beaumont [a], Taïs Grippa [b] and Moritz Lennert [b]

[a]Remote Sensing and Geodata Unit, Institut Scientifique De Service Public, Liège, Belgium; [b]Department of Geosciences, Environment & Society, Université Libre De Bruxelles, Belgium

ABSTRACT

Regional land use monitoring at high spatial, temporal, and thematic resolution is an important expectation of Walloon stakeholders. Over the last decade, increased data-processing capacities and the annual acquisition of remotely sensed data have resulted in the production of a large amount of relevant geodata. The INSPIRE directive and its obligations for 2020 serve as a path for the development of a new user-driven and open-source hierarchical land use classification system mapping scheme, as presented in this paper. The process includes intensive user consultation, the development of an entire automatic processing chain, and efforts to address challenges such as big data handling, the variability of input data properties, and reproducibility. The thematically detailed land use map, with its 69 classes, is already widely used by Walloon stakeholders, and new demands for updating have already emerged. Based on a European classification system that is compulsory for all member states, INSPIRE-compliant land use maps will make it possible to carry out cross-border studies and compare spatial planning strategies between states.

ARTICLE HISTORY
Received 15 October 2020
Accepted 6 January 2021

KEYWORDS
Land Use; geo-Information; open-Source; inspire HILUCS; user-Driven Cartography

1. Introduction

Land use (LU) reflects the intensity and location of human activities' impact on the natural environment. Monitoring the state and change in LU and regulating its future evolution constitute key issues for the development of sustainable territorial policies. INSPIRE Directive 2007/2/EC defines existing LU as 'the territory characterised according to its current functional dimension or socio-economic purpose (e.g., residential, industrial, commercial, agricultural, forestry, recreational)' (European Parliament, Council of the European Union 2007, 13). It objectively depicts the use and functions of a territory as it has been and effectively still is in real life. Local and regional stakeholders benefit from LU maps for decision-making as LU data enable them to assess past or draw new spatial planning strategies for promoting sustainable development (Herspergeret al. 2018), protecting and reinforcing environmental assets (Cunha and Magalhães 2019), and improving built, economic, and social environments of communities (Barton 2009). When combined with land cover (LC) data, LU data offer great tools, especially for monitoring land take (Science for Environment Policy 2016), modelling flood risks (Akter et al. 2018; Stephenne et al. 2017), and managing various geospatial data sets (Aubrecht et al., 2009). This paper describes how Wallonia, in southern Belgium, developed a new LU mapping scheme that provides INSPIRE-compliant data built and designed to answer regional and local stakeholders' needs.

EU member states had to provide a compliant LU spatial dataset by October 2020, as LU was listed in the third annexe of INSPIRE's spatial data themes. The Hierarchical INSPIRE Land Use Classification System (HILUCS) is INSPIRE's mandatory LU nomenclature (INSPIRE Thematic Working Group Land Use 2013) in order to ensure the interoperability of spatial data sets and services across all European countries. HILUCS is a multi-level (three levels) nomenclature developed according to land (nature, networks, and built-up) and economic (primary, secondary, and tertiary sectors) perspectives, thereby resulting in six classes at Level 1: (1) primary production, (2) secondary production, (3) tertiary production, (4) transport networks logistics and utilities, (5) residential use, and (6) other use. Level 1 is further subdivided into 27 and 65 classes at Levels 2 and 3, respectively.

To meet the INSPIRE deadline, in 2017 Walloon authorities funded the WALOUS project aimed at the user-driven elaboration of new INSPIRE-compliant LC and LU datasets for the region. The previous LC and LU dataset, created in 2007 (SPW-ARNE 2007), relied on a mixed LC and LU nomenclature inspired by Corine Land

CONTACT Benjamin Beaumont b.beaumont@issep.be Remote Sensing and Geodata Unit, Institut Scientifique De Service Public, Liège, Belgium
This article has been republished with minor changes. These changes do not impact the academic content of the article.

Cover that no longer meets today's needs. Given users' high expectations, the WALOUS approach had to meet the end-users' needs to the greatest extent possible (Beaumont et al. 2019b), be reproducible, and rely on open-source solutions in order to reduce costs and improve the timeliness of future production.

The LU and LC mapping did not need to meet the same prerequisites. Regarding the LC mapping, Wallonia had already disposed of three successful studies. Grippa et al. (2017) developed an open-source semi-automated processing chain for urban object-based LC mapping that Beaumont et al. (2017) further applied to the urban areas of the city of Liège. Radoux et al. (2019) developed a pixel-based LC mapping approach for ecosystem monitoring purposes in the Lifewatch project framework covering the entire region. These papers highlighted the quality of the input data available for Wallonia. Indeed, each year the authorities acquire coverage from aerial orthophotos (and a derived digital surface model) with sub metre spatial resolution. Lennert et al. (2019) described the first steps towards a new reproducible and open-source approach for building the new 2018 Walloon LC database (i.e., Bassine et al. 2020). Beaumont et al. (2019a) discussed a supervised LU proof-of-concept mapping scheme for the city of Liège. The approach uses spatial indicators derived from the LC data set along with geocoded data of residential and company units as inputs to provide a LU map with a legend similar to the European Urban Atlas 2012 (EEA 2012). However, users' needs were not appropriately addressed, and INSPIRE compliance was not pursued.

Although remote sensing is often recognized as a viable source of data from which generalist LU maps can be efficiently created and updated (Rozenstein and Karnieli 2011), with recent examples integrating neural networks and object-based image analysis (OBIA; Huang, Zhao, and Song 2018; Liu et al. 2019), it is of limited help when addressing a nomenclature as precise as HILUCS. Highly specific or rich thematically geographical information system (GIS) data sets are needed to support HILUCS mapping, including statistical information such as census and business activities (Diaz-Pacheco. and Garcia-Palomares 2014).

A successful pilot of an INSPIRE-compliant LU map is the Open Land Use (OLU) map application (SDI4Apps 2017), which provides harmonized HILUCS open data using the data model defined by INSPIRE and integrating various pilot regions' data from various level of details (i.e., Urban Atlas, Corine Land Cover, Land Parcel Identification System [LPIS], and Digital cadastre data). Still, only a few regions were part of the prototypes, and Wallonia was not one of them.

Nevertheless, Wallonia, as well as Belgian federal institutions, manage a diversified and rich catalogue of geodata potentially useful for increasing the output quality of HILUCS. However, the relevant information is found in a diverse collection of data sets with varying thematic, geometric, and temporal completeness and different update frequencies. Semantics are not always directly translatable into HILUCS, and a lot of redundancy occurs among the different data sets.

This paper presents how we addressed the identified challenges to develop a new LU mapping scheme driven by users while also being compliant with INSPIRE and using a transparent and reproducible approach. Section 2 describes the overall methodology. Section 3 details the classification results. Section 4 discusses the major lessons learned and future work.

2. Materials and methods

2.1. Study area

Wallonia, in southern Belgium, is a 16.901 km2 EU-NUTS 1 statistical region sharing borders with Flanders, the Netherlands, Germany, Luxembourg, and France (Figure 1). Its mean population density is 215.7 inhabitants/km2 (IWEPS 2020a). The northern and western parts of Wallonia are mostly occupied by arable lands. Forests and grasslands cover the southern and eastern parts while the Sambre and Meuse River areas (central-cluster) cross Wallonia's three most populated municipalities of Charleroi, Liège, and Namur (SPW-DGO2 2018).

2.2. User-driven building process

Detailed and accurate knowledge of LU is crucial for many scientific and operational applications. Beaumont et al. (2019b) showed the importance of precisely defining the technical specifications of the database within a strong co-building process with the expected users. The authors focused on the initial stage of the consultation

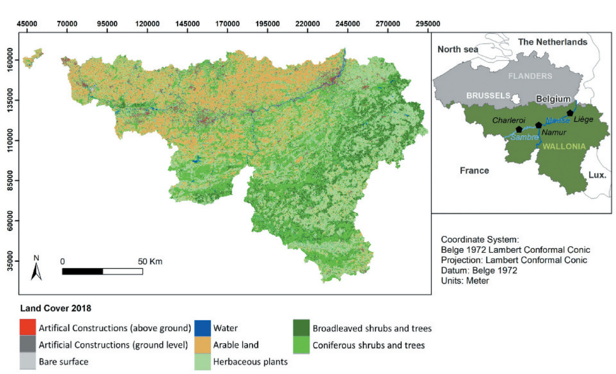


Figure 1. Wallonia, a 16 901 km2 study area illustrated using the new LC 2018 dataset.

process, which consisted of interviews asking open- and closed-ended questions followed by a plenary meeting covering three specific goals: (1) introducing users to the INSPIRE LU data specifications requirements, (2) amending these requirements according to their particular needs, and (3) performing a census of available data, including gathering their experience using these in their work fields. In total, 46 members of 23 institutions participated in this consultation, which lasted 3 months. The institutions represented included public administrations at the national, regional, and city levels as well as research centres, universities, and non-profit organizations active in a wide variety of fields, such as spatial planning, environmental monitoring, agriculture, climate modelling, hydrology, digital transformation, transports, and logistics. Unfortunately, private companies, such as spatial planning and environmental impact consulting firms, declined the invitation to participate.

During the interviews, in order to identify acceptable trade-offs (i.e., co-define methodological priorities), users were asked to allocate a limited budget of 10 points to the following three objectives: (1) minimum mapping unit (MMU), (2) overall accuracy (OA), and (3) update frequency (UF). For example, a user who asks for yearly and very accurate (OA > 90%) LU data has to accept a compromise on MMU. The ternary graph in Figure 2 presents the priority needs users expressed for the specification of the LU map. This graph shows six distinct combinations of compromises, with certain general tendencies. Users expressed a preference for fine-scale LU representation (i.e., cadastral parcels at a minimum), mostly at the expense of OA. Necessary update frequency for data production ranges from 3 to 5 years for most users (85%). Globally, if the new map meets these requirements, end-users would be interested in integrating these new data as a sustainable input into their decision-making processes.

INSPIRE HILUCS is mandatory at the most appropriate level according to the geographic data sets available

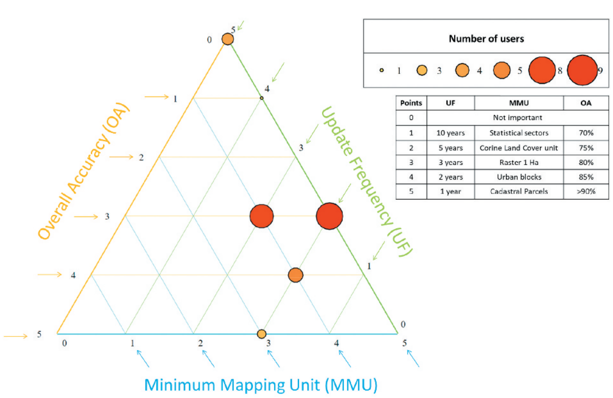


Figure 2. Ternary graph illustrating the compromises among LU objectives as preferred by users.

(INSPIRE Thematic Working Group Land Use 2013). Still, HILUCS is mainly driven by the economic perspective, which leaves out other needs and uses. Consultations with potential users resulted in four amendments to HILUCS with limited impacts on interoperability: (1) is the subdivision of the Level 2 agricultural class into grasslands, arable lands, and Christmas tree farms; (2) the subdivision of Level 2 residential classes according to four population density levels; (3) the elimination of Level 3 for several classes, such as industry and commercial activities, as they were of no use to users; and (4) increased thematic visibility of natural areas in the first level of the legend, resulting in the decision to move the '6_3_NaturalAreas' class to a new Level 1 '7_NaturalAreas' class. Appendix 1 presents the final legend.

A consolidation process and two plenary meetings followed this initial consultation stage. The consolidation process consisted of providing a 1-kilometre buffer of preliminary LU classification results around familiar neighbourhoods for 16 experts (Figure 3). Their task was to validate the LU according to their expert knowledge. Their feedback proved essential for refining the rules allocating HILUCS classes as well as consolidating the list of input data. The plenary meetings included new users (80+ in total, up from the 46 in the initial stage) to discuss methodological options based on preliminary results and ask specific questions of the relevant experts. These actions increased the project's visibility, the comprehension of the methodological choices according to the panel's needs, and the product's appropriation, but also contributed to higher expectations.

2.3. LU processing chain

Figure 4 presents an overview of the developed processing chain, which is fully available in a GitHub repository (i.e., https://github.com/tgrippa/WALOUS_UTS).

The processing chain maps LU at the scale of mutually exclusive cadastral parcels' polygons (i.e., the finest scale requested by the panel of users). The cadastral parcel plan graphically represents and assembles a map of the heritage documentation on the Belgian territory. In terms of layers of information, this dataset consists mainly of real estate (property makers defining the parcels, which are the mapping units of this study, and building footprints), street names, addresses, places, and administrative boundaries. The federal authority (SPF-Finances 2019) publishes this plan annually on 1 January. The orange boxes in Figure 4 depict how statistics are extracted from the three various formats of data inputs. Indeed, the mapping scheme integrates various types of geographical data sets that provide LU information at the time of observation, including

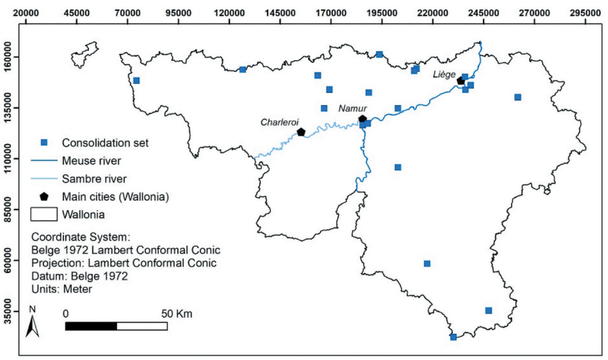


Figure 3. LU consolidation set provided by the external experts.

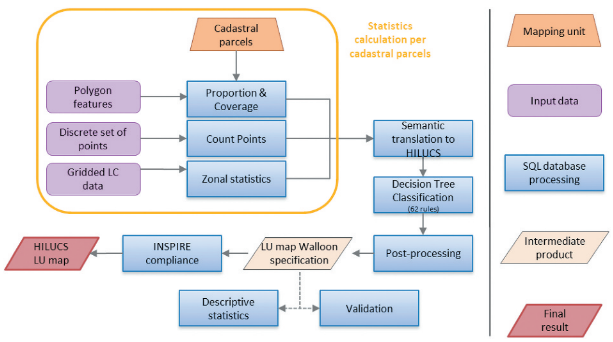


Figure 4. LU classification processing chain.

polygon features (e.g., natural areas), a discrete set of points (e.g., coordinates of population and firm registers according to their postal addresses), and gridded data (e. g., LC). Table 1 provides an overview of the data inputs. In total, the process integrates 24 distinct databases, including the mapping unit and the raster layer of LC as well as 19 thematic data inputs and three specific data inputs integrated for dealing with areas outside of the cadastral parcels (i.e., mainly the networks). The extraction of statistics includes proportion and coverage calculations for polygon features, counts of elements, and a synthesis of attributes (e.g., total number of residents) for discrete data as well as zonal statistics (area ratio and mode) extraction for the gridded LC data. A decision had to be taken in the case of polygon features in order to avoid the observed co-registration issues between the different input data used and the cadastre. This decision included considering an input polygon feature as relevant for the classification of each mapping unit only if it covers at least 15 m2 and overlaps at least 5% of this mapping unit. Without these thresholds, multiple misclassifications were observed, especially between forests and agricultural lands. If necessary, these thresholds could easily be refined within the code.

Lookup tables, which are also available on the GitHub repository, translate complex multi-class dataset legends into the LU nomenclatures. The processing chain derives two LU nomenclatures: one including the subdivision of agricultural and residential lands as well as the increased visibility of natural elements (Walloon users' specification; see Appendix 1), and a second one being strictly compliant with INSPIRE HILUCS. For the latter, the processing chain provides options 1 and 2, as defined in the INSPIRE LU data specification (INSPIRE Thematic Working Group Land Use 2013). Option 1 depicts the majority class per cadastral parcels, which facilitates cartographic representation. Option 2 depicts the complete information by providing several LU per spatial unit for advanced users.

An important aspect for users was the transparency of the classification process; thus, a decision tree approach was deemed to be the most useful for mapping the dominant LU of the parcel. The fact that the queries provide a single piece of LU information does not mean that only one LU exists in the given cadastral parcel as different usages often co-exist. However, the decision tree approach provides easily understandable data, with limited complexity in the data symbology. Still, the process creates an HILUCS-compliant attribute in order to keep all the information useful for an expert's use of the data (i.e., all classes encountered for each cadastral parcel). The user-driven process helped consolidate the decision tree classification (see Figure 1 for the consolidation areas). In total, the decision tree contains 62 hierarchical conditional SQL queries, which are available in the GitHub repository. Figure 5 presents it with a synthetic view.

Automated post-processing in the form of SQL queries corrects well-identified and regular misclassifications and derives subclasses. The well-known misclassifications include garden parcels classified as agricultural lands due to the owners' incorrect or out-of-date declaration of the nature of the land in the cadastre. Most of these parcels are annexed to residential areas in very dense neighbourhoods, which help in the automatic detection. A neighbourhood's analysis of population density calculates for each residential parcel the population density in a buffer of 200 metres, which can then be used to subdivide residential areas into four subclasses. Finally, the post-processing step also manages the creation of the HILUCS-compliant attributes as well as the final table export.

Following the post-processing, two toolboxes allow for the validation of the final map as well as for the extraction of a series of descriptive statistics. The validation toolbox reports confusion matrices, overall accuracy, Cohen's kappa metric, and the F1 score.

The processing chain has been designed and implemented in such a way that it is as automated as possible

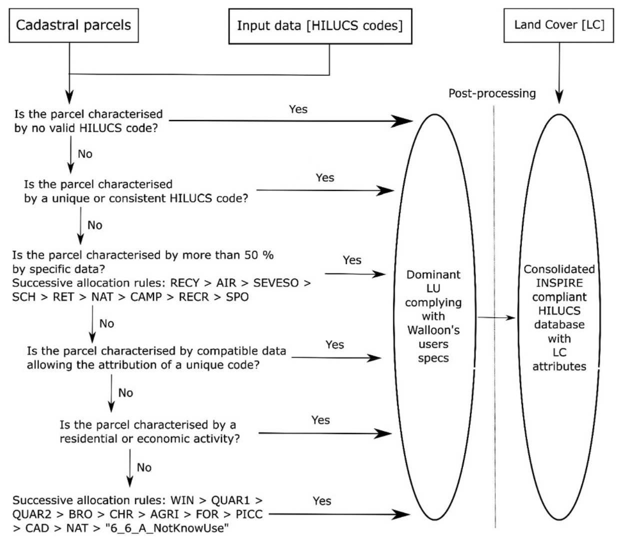


Figure 5. Synthetic view of the hierarchical set of queries for mapping dominant LU. Table 1 explains the acronyms.

while remaining easily executable and modifiable by a user not highly specialized in programming. Thus, it is possible, for example, to add new sources of information and modify the HILUCS lookup tables or the rules' priorities in the decision tree classification. The processing chain is provided as a series of Jupyter Notebook files (Kluyver et al. 2016), which are widely used in the computer sphere as they offer the advantage of being able to combine, in the same file, the programmer's and/or user's guide and the actual code. The code can be directly executed in the Jupyter Notebook, thereby improving interactivity with the code and debugging. In addition, Jupyter Notebooks can display the content of the tables stored on a PostgreSQL database directly, without having to go through a database client software.

The processing chain uses several software packages and programming languages. The spatial and non-spatial data tables are integrated into the PostgreSQL database (http://www.postgresql.org/) with PostGIS extension for spatial analysis functions (Strobl 2008). Working with database management software (DBMS) with spatial functions instead of traditional GIS analyses (intersections, unions, joints of attribute tables within GIS software such as QGIS or ArcGIS) eased the data organization and manipulation of very large tables. DBMS such as PostgreSQL easily manages very large volumes of spatial and non-spatial data. The SQL language is used for queries of the database. Python automates the different steps as well as the automatic generation and launching of SQL queries (Van Rossum and Drake 1995). Whenever possible, steps are coded as functions stored in Python scripts. Some pre-processing of the input data is done in GRASS GIS (Neteler et al. 2012), but in an automated way via the Python API of the software.

Although cost management also plays a role, the choice for completely free and open source software was motivated mostly by the idea that public software should be open because it was financed by public money (https://publiccode.eu/) and it ensures durability and independence for the administration itself as well as allowing public scrutiny of procedures. If all software used is at least on par with, if not more advanced than, the equivalent proprietary products, there is no reason to use the latter.

2.4. Data challenge

Providing highly detailed LU data according to INSPIRE LU data specification and users' needs for the 3,909,501 cadastral parcels and the 808 km2 of non-cadastre land proved to be a complex task as various challenges emerged related to big data, data properties, calibration, and validation.

In recent decades, the rapid evolution of geographic information data handling, along with aerial and satellite data acquisition capabilities with the automation of data extraction algorithms and techniques, is providing increasingly abundant and up-to-date information on LU (Diaz-Pacheco. and Garcia-Palomares 2014). For example, the WalOnMap web platform, gathering all official geodata of Wallonia, contained 372 datasets as of August 2020, many of which relates to LU to some extent. The user-driven building process allowed for the selection of 24 datasets. Integrating these into the mapping process, given their varying formats, scales, timeliness, qualities, and contents (Table 1), makes it challenging to get coherent output on the 69 classes' final map. These variations strongly influence the methodological choices, the overall implementation of the processing chain, and the hierarchical order of queries.

3. Results

The final output is a fully automated processing chain providing a spatial database containing LU and LC attributes for the complete Walloon territory. Attributes classify LU according to the Walloon and INSPIRE specifications (see Appendix 1 for the complete nomenclature). Figure 6 illustrates the dominant LU according to Walloon users' needs over the city of Louvain-la-Neuve. After the general Level 1, Levels 2 and 3 subdivide the land according to more detailed economic and administrative uses. Level 4 provides a further discrimination of agricultural lands into grasslands, arable lands, and Christmas tree farms. The bottom of Figure 6

Table 1. Input Data Sources and Their Main Characteristics in Terms of Format, Timeliness, Quality, INSPIRE Readiness, and Property Rights.

Database (interest) [acronyms]	Format	Timeliness	Quality issues	HILUCS compliance	Property rights
Cadastral parcels (mapping unit)	Polygons	Yearly (1 January)	1–2.5 m geometric accuracy/ incomplete (95.22% of the territory)	/	Open data
Cadastral matrix (declared use of land) [CAD]	Table	Yearly (1 January)	1–2.5 m geometric accuracy/ incomplete (0.36% of parcels without declared use)	Uncertainty for the conversion of approx. 10% of the 240 declared use	Strict terms of use
WALOUS Land Cover map (change detection) [LC]	Raster	2018	91.5% overall accuracy	/	Open data
National Register of Natural Persons (Residential LU) [RNPP]	Points	Yearly (1 January)	Limited quality of the geocoding	Residential LU	Strict terms of use
Company register (Commercial/Industrial LU) [DBRIS]	Points	Continuous update	Limited quality of the geocoding/ validity of the declaration (multiple codes)	Ok	Strict terms of use
Land Parcel Information System (Agriculture) [AGRI]	Polygons	Yearly	Validity of the declaration and quality of the control	Agriculture LU	Publicly available according to terms of use
IGN Top10vGIS elements: Specific activities incl. airports [AIR], sport fields [SPO], recreative areas [RECR] and quarries [QUA1]	Polygons	Continuous update from 2017 to 2020	Spatial and temporal variability of the updating	Ok	Strict terms of use
Other regional datasets: Public forest lands[FOR][1]/Brownfields[BRO]2/Recyparks [RECY]2/3 databases on nature conservation [NAT]2/Schools[SCH][1]/Retirement homes[RET]2/ SEVESO2/Camping [CAMP]2/Wind turbines[WIN][1]/Quarries [QUA2][1]/Christmas trees [CHR]2/various elements of the buildings & infrastructure database [PICC]2/3 databases for dealing with non-cadastre areas2	Polygons	Uncertain updating frequency, theoretically continuous update	Spatial and temporal variability of the updating Variable geometric accuracy qualities	Ok	[1]Not published yet or [2]Publicly available according to terms of use

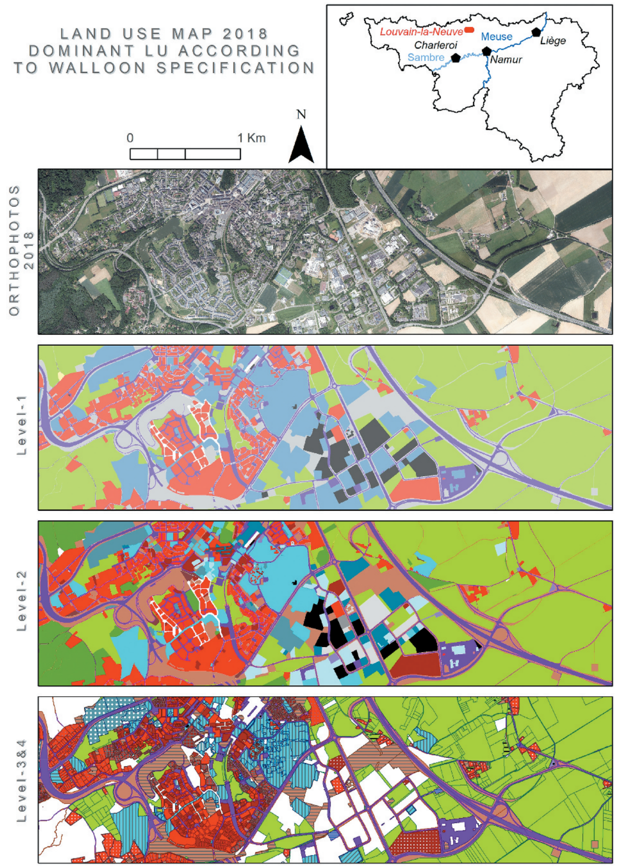


Figure 6a. Illustration of the city of Louvain-la-Neuve, highlighting the dominant LU classification and nomenclature (Walloon specification).

presents a hybrid level grouping Levels 3 and 4. White areas in the map for Levels 3 and 4 indicate areas where no information is available at the given levels.

The LU map with Walloon specification has been validated at Level 1 using the validation toolbox. A stratified random sampling approach has been used to create an independent set of 1200 parcels classified using the photo interpretation of aerial images, Google street view, and internet research (to assess company activities). For 240 parcels, the expert was not able to define dominant LU with greater than 80% certainty (subjective assessment). As such, the final validation dataset contains 960 parcels (i.e., an average 137 parcels per class). Overall accuracy (OA) is 83.14% (area wise), the F1 score is 0.86, and the Kappa index is 0.76. Classes 1 (F1 score of 0.94) and 2 (F1 score of 0.87) get the highest scores while Class 6 is of low quality (F1 score of 0.06) due to the inner definition of Class 6 (e.g., abandoned areas, unknows use) versus the capacities of the expert to attribute such classes during the validation of the dataset creation. Class 1 is the most represented class in terms of surface in Wallonia (80% of land, forest and agriculture) and is globally well classified. Not considering Class 6 given the classification uncertainty of its elements, but taking into account sampling probability and using an accuracy approach weighted by the area represented by each class (Radoux and Bogaert 2017) results in an OA of 96.10% ± 1.09 at Level 1.

The descriptive toolbox provides some key numbers for the regional application. In terms of the percentage of cadastre areas while crossing all data inputs, 0.23% have no LU information at all, 18.34% have only one piece of LU information, and 50.95% are characterized by identical LU information from different sources at Level 3. The maximum of different LU classes attributed to one parcel when including all levels is 10, resulting in a mean number of 1.79 classes per parcel. In terms of classification results of dominant LU at Level 1 (area wise), Wallonia is predominantly characterized by primary production (80.96%), but also includes secondary production (0.67%), tertiary production (1.55%), transport networks, logistics, and utilities (2.62%), residential lands (6.79%), natural areas (2.13%) and other LU (5.27%).

IWEPS (2020b) provides annual LU regional statistics via the WALSTAT online interface (walstat.iweps.be).

Figure 6b. (continued)

Using only the declared use of land (240 categories) from the cadastral matrix, which has the strong advantage of having been published annually for years, IWEPS synthetized this information into a nomenclature of 16 classes (IWEPS 2018). Table 2 compares the 2018 WALSTAT statistics with the WALOUS LU product (in terms of area percentage).

However, when comparing these numbers, grouping them into the 16 categories of the original nomenclatures can lead to discrepancies. In that IWEPS (2018, 2020b) has extensively documented the relevance and limitations of its analyses. However, several observations can be made regarding the main differences. Residential use is higher in WALOUS LU, as this class also includes compatible uses. WALOUS integrates the population register and thus detects mixed uses, including residential use, whereas WALSTAT prioritizes commercial activities and services over residential usages to provide only a single LU per parcel. The lower value of community services in WALOUS LU can be explained by the classification of military zones. After consulting with the end-users and given the very high biological value observed in several Walloon military camps, it was decided to classify these as natural areas. Still, the semi-natural areas' value is lower in WALOUS LU than in WALSTAT; this is due to different definitions. With the exception of the military camps, WALOUS LU allocates cadastral parcels to natural areas only if official nature conservation databases intersect them. Meanwhile, the grouping of declared use from the cadastral matrix in the semi-natural areas class of WALSTAT is less strict as it also integrates unprotected areas. Integrating the Land Parcel Information System (LPIS) in WALOUS LU and the post-processing of misclassified agricultural lands into residential gardens leads to a better classification of arable lands and grasslands. Finally, the inclusion of three datasets for dealing with non-cadastre areas reduces the percentage of land with unknown use. The WALOUS LU approach thus brings additional geo-information for an improved LU classification. Still, a more in-depth comparison will be

Table 2. Comparison of current LU regional reference statistics from WALSTAT versus the new WALOUS LU database.

IWEPS WALSTAT classification	WALSTAT 2018 (% of land)	WALOUS LU 2018 (% of land)	Difference (WALOUS – WALSTAT)
Residential Use	6,4	6,9	+0,5
Commercial, Financial Professional and Information Services	0,3	0,5	+0,2
Community Services	1,1	0,5	−0,6
Cultural, Entertainment and Recreational Services	0,9	0,6	−0,3
Farming Infrastructure	0,6	0,6	0
Secondary Production	1,0	0,7	−0,3
Mining and Quarry, Waste Management and Abandoned Areas	0,2	1,7	+1,5
Transport Network and Logistics	0,4	1,0	+0,6
Arable Lands and Permanent Crops	28,6	27,4	−1,2
Grasslands	23,2	24,7	+1,5
Other artificial lands	0	0	0
Forests	29,3	28,9	−0,4
Semi-Natural Areas	2,6	2,1	−0,5
Wetlands	0,3	0,0	−0,3
Water Areas	0,2	0,4	+0,2
Unknown Use	5,2	4,0	−1,2

carried out in the future to improve the new LU database in subsequent iterations.

The software solution is as modular as possible and open source (see the GitHub repository: https://github.com/tgrippa/WALOUS_UTS). At each step of the mapping scheme, particular attention has been dedicated to ensuring that methods are as reproducible as possible. Well-established open source software has been used, and big data issues have been managed by optimizing code. Given the property rights of the datasets, database access is secured and limited to specific users. Still, the software solution (without the input data) is provided under the GNU-GPL V3 licence, and the final cartographic products are open data under the general conditions (SPW 2016a, 2019) extended by the CPU-Type C licence (SPW 2016b).

4. Discussion

This paper has introduced a fully automated and open-source methodology for mapping INSPIRE-compliant HILUCS data from a rich catalogue of geodatabases, without neglecting the specific needs of local stakeholders. At a time when the replicability of research and the sharing of existing solutions are high on the discussion agenda, the development of free and open-source solutions is becoming paramount (Ince, Hatton, and Graham-Cumming 2012; Nosek et al., 2015). This processing chain, relying on existing open-source geospatial software, proves its ability of being quickly customizable in order to match the continuously evolving requirements expressed by a large panel of users.

Two mains issues emerged during the research. The first issue was how to develop a tool that meets both the requirements from above (INSPIRE regulatory framework) and the (varying) operational needs from below (local actors). The answer took the form of a well-defined user-driven building process. Users were involved from the first day, when they were introduced to the INSPIRE specificati ons and the practical possibilities offered by Walloon data, until the last day, which involved the validation and presentation of the final outputs. What was clear was the need to clarify from the outset the concepts and requirements specified by the directive. Even the distinction between LU and LC needed to be clarified, as reinforced by the historical mixing of information at both the regional (SPW-ARNE 2007) and European scales (Corine Land Cover). Given the desired generalist scope of the LU data, the user sample should leave no one out. Moreover, no data should be omitted. The right to speak and express an opinion had to be fairly accommodated without any single opinion prevailing. In order to guarantee all these elements, it was essential to set up a procedure for collecting needs that was as objective and transparent as possible, followed by several plenary meetings led by a moderation team. Feedback from this group of users was provided back to the decision-making body overseeing the research. The reasoned decisions were then transparently passed down to the panel, who had the chance to react to them. When no response could be provided to a particular need, a dialogue was established with the stakeholder to assess future avenues for improvement or ancillary data to meet it. Taking into account local needs resulted in easily reversible changes, both in terminology and in the automated tools developed. Documented codes were as modular as possible, and versioning was key.

The second issue that emerged in this research was the need to determine how to process so much input data with various specifications and ensure the transmission of accurate information that was up to date as possible and not in contradiction with the realities on the ground. This technical challenge was difficult to address. The experts all had their own opinions, according to their specific needs, about the input data qualities. Source data documentation can be very fragmented or complex to obtain. Geometric accuracy varies from one dataset to another. In addition, as Diaz-Pacheco. and Garcia-Palomares (2014) stated, it is not easy for each piece of information to be integrated to depict the LU configuration at a single reference date (here, 2018). In practice, this single reference date may often be a composite of several adjacent times. In our case, some input data sets were created in 2015 (such as Christmas tree farms) or even earlier (such as PICC dataset), with no traceability of production date per object. Some data providers use local or thematic updating approaches according to available resources or opportunities. Therefore, their dataset does not ensure a complete representation of the situation in a given year, but mixes recent data with older and outdated elements. Still, no alternative to these datasets exists, which makes the study of short-term (one or two years) change more complex, which is a strong constraint for many applications. Ranking the data according to an order of reliability has proven to be difficult, and changes to the decision tree classification (prevalence of rules) have been realized until the very last minute. Again, the modularity and automation of the code proved essential here.

In the end, an INSPIRE-compliant LU map has been provided to the authorities, valorizing the rich catalogue of reference geodata available in Wallonia. As the end-users' needs were strongly taken into account, various derived uses of the new LU data are already being carried out, including regional flood risk assessments, a study of the vulnerability and adaptation to climate change in urban areas, and the modelling of LU change. With an open-source and reproducible mapping scheme, the presented approach offers great opportunities for a continuous LU mapping of Wallonia.

Future work is likely to include the renewal of the database for 2019, 2020, and beyond, with revised methodological choices according to user feedback, training of the administration for internalization of the process, and the development of approaches to change detection for decision-making efforts.

Acknowledgements

This work was supported by the Geomatics Department (SG-DGEO) and the Agriculture, Ressources naturelle et Environnement (ARNE) of the Service Public de Wallonie (SPW) under the Grant "WALOUS". Computational resources benefited from the ULB server infrastructure. The authors would like to thank all 80 + experts participating in the user-driven process.

Disclosure statement

No potential competing interest was reported by the authors.

Funding

This work was supported by the Service Public de Wallonie as part of the WALOUS project.

Notes on contributor

Benjamin Beaumont is a researcher in the Remote Sensing and Geodata unit of the Institut Scientifique de Services Public (ISSeP). He received his MSc in Bioengineering, specializing in Science and Technology of the Environment, at the Université catholique de Louvain (UCLouvain) in Belgium. His current research interest focuses on exploiting Earth observational data for deriving useful thematic information, especially in the field of LCLU mapping, climate reporting, and coal heaps and landfills monitoring.

Taïs Grippa is a geographer and geomatician at the Université Libre de Bruxelles (ULB), Institute for Environmental Management and Land-use Planning (IGEAT). He received his PhD degree in 2019. His research activities mainly focus on using state-of-the-art and open-source GIS, remote sensing, and machine learning solutions for mapping land cover and land use from very high resolution satellite imagery, modelling population density, and producing spatial analyses.

Moritz Lennert is a geographer and senior researcher at the DGES of the ULB. He holds a Master's Degree and PhD in Geography from the ULB and an MSc in Globalization and Development from Queen Mary's, University of London. He has extensive experience in economic geography, regional and urban development, and GIS and remote sensing. He is member of the GRASS GIS Project Steering Committee and a charter member of the OSGeo Foundation, as well as co-founder of its Belgian chapter.

ORCID

Benjamin Beaumont http://orcid.org/0000-0003-1518-161X
Taïs Grippa http://orcid.org/0000-0002-9837-1832
Moritz Lennert http://orcid.org/0000-0002-2870-4515

Data availability statement

The Land Use and Land Cover databases described herein are openly accessible through the Walloon geoportal *WalOnMap*: https://geoportail.wallonie.be/walonmap. Metadata files for the two layers are available on the Metawal portal: http://metawal.wallonie.be/.

The computer code developed for automated decision-tree classification of LU is available at https://github.com/tgrippa/WALOUS_UTS (10.5281/zenodo.4108712)

Geolocation information

All geospatial datasets used and produced in this research cover the EU-NUTS 1 region of Wallonia (code BE3) in southern Belgium. GCS_WGS_84 coordinates are 'Left: 2,843,070 dd, Top: 50,811,920 dd, Right: 6,407,420 dd, Bottom: 49,497,240 dd'.

References

Akter, T., P. Quevauviller, S. Eisenreich, and G. Vaes. 2018. "Impacts of Climate and Land Use Changes on Flood Risk Management for the Schijn River, Belgium." *Environmental Science & Policy* 89: 163–175. doi:10.1016/j.envsci.2018.07.002.

Aubrecht, C., K. Steinnocher, M. Hollaus, and W. Wagner. 2009. "Integrating Earth Observation and GIScience for High Resolution Spatial and Functional Modelling of Urban Land Use." *Computers, Environment and Urban Systems* 33 (1): 15–25. doi:10.1016/j.compenvurbsys.2008.09.007.

Barton, H. 2009. "Land Use Planning and Health and Well-being." *Land Use Policy* 265: S115–S123. doi:10.1016/j.landusepol.2009.09.008.

Bassine, C., Radoux, J., Beaumont, B., Grippa, T., Lennert, M., Champagne, C., De Vroey, M., Martinet, A., Bouchez, O., Deffense, N., Hallot, E., Wolff, E. & Defourny, P. (2020). First 1-M resolution land cover map labeling the overlap in the 3rd dimension: The 2018 map of Wallonia. Data. Spatial Data Science and Digital Earth. 1–11. *Data* 5 (4): 117. https://doi.org/10.3390/data5040117

Beaumont, B., N. Stephenne, C. Wyard, and E. Hallot (2019b). "Users' Consultation Process in Building a Land Cover and Land Use Database for the Official Walloon Georeferential." *2019 Joint Urban Remote Sensing Event (JURSE), Vannes, France*, 1–4. doi:10.1109/JURSE.2019.8808943.

Beaumont, B., T. Grippa, M. Lennert, N. Stephenne, S. Zarougui, J. C. Jasselette, L. Van de Vyvere, and É. Wolff. 2019a. "Vers une méthodologie opérationnelle de cartographie à très haute résolution spatiale de l'occupation et de l'utilisation du sol en Wallonie." *Dynamiques régionales* 8: 15–33.

Beaumont, B., T. Grippa, M. Lennert, S. G. Vanhuysse, N. R. Stephenne, and E. Wolff. 2017. "Toward an Operational Framework for Fine-scale Urban Land-cover Mapping in Wallonia Using Submeter Remote Sensing and Ancillary Vector Data." *Journal of Applied Remote Sensing* 11 (3): 036011. doi:10.1117/1.JRS.11.036011.

Cunha, N. S., and M. R. Magalhães. 2019. "Methodology for Mapping the National Ecological Network to Mainland Portugal: A Planning Tool Towards A Green Infrastructure." *Ecological Indicators* 104: 82–818. doi:10.1016/j.ecolind.2019.04.050.

Diaz-Pacheco., J., and J. C. Garcia-Palomares. 2014. "A Highly Detailed Land-use Vector Map for Madrid Region Based on Photo-interpretation." *Journal of Maps* 10 (3): 424–433. doi:10.1080/17445647.2014.882798.

EEA. (2012). "Copernicus Land Monitoring service—Urban Atlas." https://www.eea.europa.eu/data-and-maps/data/copernicus-land-monitoring-service-urban-atlas

European Parliament, Council of the European Union. 2007. "Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 Establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)." *Official Journal of the European Union* 108: 1–14.

Grippa, T., M. Lennert, B. Beaumont, S. Vanhuysse, N. Stephenne, and E. Wolff. 2017. "An Open-source Semi-automated Processing Chain for Urban Object-based Classification." *Remote Sensing* 9 (4): 358. doi:10.3390/rs9040358.

Hersperger, A. M., E. Oliveira, S. Pagliarin, G. Palka, P. Verburg, J. Bolliger, and S. Gradinaru. 2018. "Urban Land-use Change: The Role of Strategic Spatial Planning." *Global Environmental Change* 51: 32–42. doi:10.1016/j.gloenvcha.2018.05.001.

Huang, B., B. Zhao, and Y. Song. 2018. "Urban Land-use Mapping Using a Deep Convolutional Neural Network with High Spatial Resolution Multispectral Remote Sensing Imagery." *Remote Sensing of Environment* 214: 73–86. doi:10.1016/j.rse.2018.04.050.

Ince, D., L. Hatton, and J. Graham-Cumming. 2012. "The Case for Open Computer Programs." *Nature* 482 (7386): 485–488. doi:10.1038/nature10836.

INSPIRE Thematic Working Group Land Use. 2013. "D2.8.III.4 INSPIRE Data Specification on Land use—Technical Guidelines." *European Commission Joint Research Centre* 1–253. https://inspire.ec.europa.eu/id/document/tg/lu

IWEPS. (2018). "Calculs IWEPS sur base de données Bodem/Sol au 1er janvier 2018—SPF Finances/Administration Générale de la Documentation patrimoniale; Nomenclature IWEPS-CDPT-DGO3, 2010—Note méthodologique sur les fiches d'utilisation du sol."

IWEPS. (2020a). "Indicateurs statistiques—Densité de population." https://www.iweps.be/indicateur-statistique/densite-de-population/

IWEPS. (2020b). "Indicateurs statistiques—Utilisation du sol." https://www.iweps.be/indicateur-statistique/utilisation-du-sol/

Kluyver, T., B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, L. Kelley, et al. 2016. "Jupyter Notebooks—a Publishing Format for Reproducible Computational Workflows." In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, edited by F. Loizides and B. Schmidt, pp. 87–90. IOS Press: Amsterdam, The Netherlands, Jupyter Development Team. doi:10.3233/978-1-61499-649-1-87.

Lennert, M., T. Grippa, J. Radoux, C. Bassine, B. Beaumont, P. Defourny, and E. Wolff. 2019. "Creating Wallonia's New Very High Resolution Land Cover Maps: Combining Grass GIS OBIA and OTB Pixel-based Results." *The International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 151–157. *XLII*. doi:10.5194/isprs-archives-XLII-4-W14-151-2019.

Liu, S., Z. Qi, X. Li, and A. Gar-In Yeh. 2019. "Integration of Convolutional Neural Networks and Object-based Post-classification Refinement for Land Use and Land Cover Mapping with Optical and SAR Data." *Remote Sensing* 11 (6): 690. doi:10.3390/rs11060690.

Neteler, M., M. Hamish Bowman, M. Landa, and M. Metz. 2012. "GRASS GIS: A Multi-purpose Open Source GIS." *Environmental Modelling and Software* 31: 124–130. doi:10.1016/j.envsoft.2011.11.014.

Nosek, B. A., G. Alter, C. Banks, D. Borsboom, S. Bowman, S. J. Breckler, S. Buck, et al. 2015. "Promoting an Open Research Culture." *Science* 348 (6242): 1422–1425. doi:10.1126/science.aab2374.

Radoux, J., A. Bourdouxhe, W. Coos, M. Dufrêne, and P. Defourny. 2019. "Improving Ecotope Segmentation by Combining Topographic and Spectral Data." *Remote Sensing* 11 (3): 354. doi:10.3390/rs11030354.

Radoux, J., and P. Bogaert. 2017. "Good Practices for Object-based Accuracy Assessment." *Remote Sensing* 9 (7): 646. doi:10.3390/rs9070646.

Rozenstein, O., and A. Karnieli. 2011. “Comparison of Methods for Land-use Classification Incorporating Remote Sensing and GIS Inputs.” *Applied Geography* 31 (2): 533–544. doi:10.1016/j.apgeog.2010.11.006.

Science for Environment Policy. (2016). “No Net Land Take by 2050?” *Future Brief 14*. European Commission DG Environment. http://ec.europa.eu/science-environment-policy

SDI4Apps. (2017). “SDI4Apps—Final Report.” *Author*.

SPF-Finances. (2019). “Plan Cadastral.” https://finances.belgium.be/fr/experts-partenaires/cadastre/plan-cadastral

SPW. (2016a). “*Conditions générales d’utilisation des données géographiques numériques du Service public de Wallonie* [PDF].” *Wallonie and Author*. https://geoportail.wallonie.be/files/documents/ConditionsSPW/DataSPW-CGU.pdf

SPW. (2016b). “Conditions particulières d’utilisation d’une donnée géographique numérique du Service public de Wallonie—Type C [PDF].” *Wallonie and Author*. https://geoportail.wallonie.be/files/documents/ConditionsSPW/DataSPW-CPU-TypeC.pdf

SPW. (2019). “Conditions générales d’accès à l’obtention d’une copie numérique des données géographiques numériques du Service public de Wallonie [PDF].” *Wallonie and Author*. https://geoportail.wallonie.be/files/documents/ConditionsSPW/DataSPW-CGA.pdf

SPW-ARNE. (2007). “La Carte d’Occupation Du Sol De Wallonie (COSW).” *WalOnMap*. http://geoportail.wallonie.be/catalogue/290e1fe8-0d99-410e-967b-a02f389b931a.html

SPW-DGO2. (2018). “Les grandes villes wallonnes: À chacune sa mobilité. SPW editions #1, Focus mobilité [PDF].” *Wallonie*. http://mobilite.wallonie.be/files/focus-mobilite-1.pdf

Stephenne, N., B. Beaumont, E. Hallot, F. Lenartz, F. Lefebre, D. Lauwaet, L. Poelmans, and E. Wolff. 2017. “Exposure and Vulnerability Geospatial Analysis Using Earth Observation in the City of Liège, Belgium.” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-1/W1: 149–156. *IV-1/W1*. doi:10.5194/isprs-annals-IV-1-W1-149-2017.

Strobl, C. 2008. “PostGIS.” In *Encyclopedia of GIS*, edited by S. Shekhar and H. Xiong, 51–76. Boston, MA: Springer. doi:10.1007/978-0-387-35973-1_1012.

Van Rossum, G., and J., . F. L. Drake. 1995. *Python Reference Manual*. Amsterdam: Centrum voor Wiskunde en Informatica.

Appendix

1. Code List of INSPIRE HILUCS classes amended by Walloon users

Classes in green were added/moved following the end-users' consultation compared to INSPIRE HILUCS.

Table 2. Code List of INSPIRE HILUCS Classes Amended by Walloon Users.

1_PrimaryProduction	2_4_EnergyProduction	*3_4_4_OpenAirRecreationalAreas*	*4_3_4_OtherUtilities*	6_2_AbandonedAreas
1_1_Agriculture	2_5_OtherIndustry	*3_4_5_OtherRecreationalServices*	**5_ResidentialUse**	6_4_AreasWhereAnyUseAllowed
1_1_1_CommercialAgriculturalProduction	**3_TertiaryProduction**	3_5_OtherServices	5_1_PermanentResidentialUse	6_5_AreasWithoutAnySpecifiedPlannedUse
1_1_1_A_Grasslands	3_1_CommercialServices	**4_TransportNetworksLogisticsAndUtilities**	*5_1_A_PermanentResidentialUseHighDensity*	6_6_NotKnownUse
1_1_1_B_ArableLands	3_2_FinancialProfessionalAndInformationServices	4_1_TransportNetworks	*5_1_B_PermanentResidentialUseMeanDensity*	*6_6_A_NotKnownUseCadastre*
1_1_1_C_ChristmasTrees	3_3_CommunityServices	*4_1_1_RoadTransport*	*5_1_C_PermanentResidentialUseLowDensity*	*6_6_B_NotKnownUseNotCadastre*
1_1_2_FarmingInfrastructure	*3_3_1_PublicAdministrationDefenceAndSocialSecurityServices*	*4_1_2_RailwayTransport*	*5_1_D_PermanentResidentialUseVeryLowDensity*	**7_NaturalAreas**
1_2_Forestry	*3_3_2_EducationalServices*	*4_1_3_AirTransport*	5_2_ResidentialUseWithOtherCompatibleUses	7_1_LandNaturalAreas
1_3_MiningAndQuarrying	*3_3_3_HealthAndSocialServices*	*4_1_4_WaterTransport*	*5_2_A_ResidentialUseWithOtherCompatibleUsesHighDensity*	7_2_WaterNaturalAreas
1_4_AquacultureAndFishing	*3_3_4_ReligiousServices*	*4_1_5_OtherTransportNetwork*	*5_2_B_PResidentialUseWithOtherCompatibleUsesMeanDensity*	
1_5_OtherPrimaryProduction	*3_3_5_OtherCommunityServices*	4_2_LogisticalAndStorageServices	*5_2_C_ResidentialUseWithOtherCompatibleUsesLowDensity*	
2_SecondaryProduction	3_4_CulturalEntertainmentAndRecreationalServices	4_3_Utilities	*5_2_D_ResidentialUseWithOtherCompatibleUsesVeryLowDensity*	
2_1_RawIndustry	*3_4_1_CulturalServices*	*4_3_1_ElectricityGasAndThermalPowerDistributionServices*	5_3_OtherResidentialUse	
2_2_HeavyEndProductIndustry	*3_4_2_EntertainmentServices*	*4_3_2_WaterAndSewageInfrastructure*	**6_OtherUses**	
2_3_LightEndProductIndustry	*3_4_3_SportsInfrastructure*	*4_3_3_WasteTreatment*	6_1_TransitionalAreas	