UN-GGIM EUROPE

UNITED NATIONS COMMITEE OF EXPERTS ON GLOBAL GEOSPATIAL INFORMATION MANAGEMENT

Core Spatial Data Theme Land Cover Recommendation for Content

Working Group A - Deliverable of Task 1.b

Version 1.1 - 2023-12-07

Version History

Version number	Date	Modified by	Comments
1.0	2022-05-05	WGA	Consolidated draft, for review by geographical and statistical community
1.1	2023.12.07	WGA	Comments from geographic and statistic community taken into account Definitive deliverable

Warning: in the following parts of this document, the paragraphs written in grey e.g. "This document has annexes containing more detailed explanations " are common to all core spatial data themes; they aim to provide context and objectives of core data. The paragraphs written in black are specific to core spatial data theme Land Cover.

Content

1 Executive Summary

In September 2015 the countries of the United Nations adopted the 2030 Agenda for Sustainable Development; a set of goals to end poverty, protect the planet, and ensure prosperity for all as part of a new sustainable development agenda. Each goal has specific targets to be achieved over the next 15 years. The 17 Sustainable Development Goals (SDGs) of the 2030 Agenda are supported by 169 targets and 230 indicators.

Geospatial data supports the measuring, achieving and monitoring of many of the goals and targets set by the 2030 Agenda. The 2030 Agenda demands new data acquisition and integration approaches to improve the availability, quality, timeliness and disaggregation of data. Goal 17 explicitly emphasizes the need for developing capacities and partnerships. In this context the success of the 2030 Agenda depends on senior administrators owning and leading the geospatial efforts in their respective countries.

In Europe, building on the INSPIRE Directive redirecting the focus on a cohesive spatial data infrastructure without gaps in content and discrepancies in quality, stakeholders are working on geospatial standardization and increasing richness of data through Core Data Recommendations for Content that correspond to the first phase of WGA work program. Core data is primarily meant for fulfilling the common user requirements related to SDGs in Member States and European institutions.

According to INSPIRE, Land cover (LC) is an abstraction of the physical and biological cover of the Earth's surface including artificial surfaces, agricultural areas, forests, (semi-)natural areas, wetlands, water bodies, according to its physical characteristics.

Land Cover is clearly core data as it is one of the themes the most widely required by users to decide, implement and monitor SDG policies, at least for SDG 15 (life on earth), SDG 2 (no hunger), SDG 13 (climate change), SDG 11 (sustainable cities), SDG 6 (water quality).

However, Land Cover is not reference data as there is no obvious corresponding real-world object; Land Cover is generalised data that may be designed according various modelling approaches, each of them having its interest and drawbacks. This variety of approaches and the evolving production processes make difficult to provide common detailed recommendations for content for this theme; this issue had already be met by INSPIRE.

WGA has identified two main requirements: the trend for large scale data to be produced at national level and the will to keep the advantages of Corine Land Cover time series. Regarding the large scale data, this document proposes some common key characteristics, such as geographic extent on whole territory, hierarchical classification of pure land cover, production frequency better than 6 years but still lets a lot of flexibility .

WGA has conducted a lot of investigations on this theme whose main learning are documented in chapter about “considerations for future” and in the annex about methodology.

2 Foreword

2.1 Document purpose and structure

2.1.1 Purpose

This document provides the main characteristics of core data for theme Land Cover with focus on the recommendation for content. This document aims to help decision makers (from governments, data producers, national coordination bodies, etc.) to define their policy regarding the improvement of existing data and production of new geospatial data. It addresses digital data.

This document has Annexes containing more detailed explanations targeting the technical people who will be in charge of implementing or adapting core data recommendations (e.g. for production purpose, as source of other standards, etc.).

2.1.2 Structure

The executive summary synthesizes the main conclusions of the Working Group A (WGA) process and results to develop the recommendation for content. It is meant mainly for high level decision makers.

The foreword reminds the general context of core data, the first step achieved by WGA (i.e., selecting core data themes), and it explains the general principles set by WGA to develop the recommendations for content of core data specifications for all selected themes.

The 'recommendation for content' document itself includes four chapters:

- Overview: it provides the general scope of the theme and describes the main use cases addressed;
- Data content: it provides the main characteristics of the recommended content, such as the list of core features and attributes (for vector data), as well as data capture and quality rules;
- Other recommendations: e.g. Coordinate Reference System, Metadata, Delivery;
- Considerations for future: this chapter addresses some key trends or significant user requirements that cannot be considered as core today but that might be considered in future.

The 'recommendation for content' document is meant for medium level decision makers. It is written in natural and not too technical language.

The technical explanations included in annexes describe the relationship between the recommendation for content and the corresponding INSPIRE specification, and contain any other appropriate information useful for this theme.

2.2 Core data context

2.2.1 Rationale for core data

The following background of harmonised pan-European data was identified.[1]

[1] Extract from the Report by the Preparatory Committee on the establishment of the UN-GGIM: Europe Regional Committee, European Commission Ref. Ares(2014)1491140 - 09/05/2014.

Authoritative geospatial data are used to support both the implementation of public policies and the development of downstream services. Moreover, geospatial data are required to be homogenous to enable the implementation of public policies in a coherent and coordinated way among countries and at regional or global level. Likewise, significant opportunities exist if services developed by industry can be exploited without requiring country specific adaptation.

The INSPIRE Directive has set up the legal and technical framework for harmonisation of the existing data related to the themes in annexes I, II and III. INSPIRE specifications provide common data models that ensure a first step towards interoperability, however ensuring homogeneous content is outside their scope, as they contain no indication about levels of detail, very few recommendations about quality, and as most features and attributes are "voidable", i.e. to be supplied if available or derivable at reasonable cost.

This background led the UN-GGIM: Europe Regional Committee to setup in 2014 the Working Group A on Core Data to deal with core data content and quality, production issues, funding and data availability.

Recommendations for content of core data will complement INSPIRE data specifications by defining the priorities on the core content that is encouraged to be made available in Europe in order to fulfil the main user requirements that are common to many countries, with focus on the SDG related ones.

Core data availability may be ensured either through upgrading of existing data when feasible or through production of new data when necessary.

2.2.2 Core data scope

In its first phase, WGA selected core data themes according to the following criteria: core data is the geospatial data that is the most useful, either directly or indirectly, to analyse, to achieve and to monitor the Sustainable Development Goals.

Among the 34 INSPIRE data themes, 14 have been considered as core including theme Land Cover.

More information about the selection process and results may be found in document 'Core Data Scope - Working Group A - First Deliverable of Task 1.a - Version 1.2' on http://un-ggim-europe.org/content/wg-a-core-data

2.3 Document objectives and principles

2.3.1 Encouraging content availability

This deliverable provides recommendations for national governments and data producers, aiming to help them to define their priorities for enriching existing data or producing new data. This deliverable is meant mainly for data producers; however, it defines the recommended result and target but not the production process.

2.3.2 Complementing INSPIRE

Core data specifications are built upon INSPIRE data specifications. On one hand, they often simplify INSPIRE by selecting core feature types and attributes and by restricting or clarifying the scope; On the other hand, they enrich INSPIRE by recommending specific levels of detail, quality rules and sometimes data model extensions. Besides, the INSPIRE common terminology is thoroughly used for naming core features and attributes.

Regarding the levels of detail, the ELF (European Location Framework) project terminology has been used. The ELF levels of detail are the following: Global, Regional, Master level 2, Master level 1, Master level 0. These terms are defined in the glossary.

Regarding delivery, core data may be supplied according to several ways. It is expected that, very often, the core data recommendations will be used to enrich and upgrade existing products. In this case, core data will be available through these improved products. Core data may also be delivered through INSPIRE conditions (specifications and services).

2.3.3 Status of core data recommendations

This document contains recommendations that are not legally binding. However, some recommendations are more important than others. This order is indicated as follow:

Core Recommendation X

It is first priority recommendation, considered as both necessary and achievable in principle. Ideally, it should encourage involved stakeholders to launch short-term actions (typically within a couple of years).

Core recommendations are usually addressing only technical aspects and are meant for the organisations in charge of producing this theme. The set of core recommendations defines the basic expectations on core data.

Good Practice X
It is second priority recommendation; if adopted, it will provide significant added value to core data; it indicates a relevant trend to be adopted as much as possible. It encourages involved stakeholders to take these recommendations into account in long term, if not possible in short term.

NOTE: some of these good practices may be quite easy to achieve and are already effective in some countries whereas some others may be more difficult to achieve. This is typically the case when these good practice recommendations involve other stakeholders in addition to the organisations in charge of producing this theme, and when they address not only technical aspects but also legal or organisational ones.

A “core data set” should contain the minimum data defined by the core recommendations (and ideally also by the good practices) of this deliverable but may of course contain more and/or better information.

2.4 Abbreviations

CLC	Corine Land Cover

CRS	Coordinate Reference System
EAGLE	EIONET Action Group on Land monitoring in Europe
EEA	European Environmental Agency
EIONET	European Environment Information and Observation Network
ELF	European Location Framework
HELM	Harmonised European Land Monitoring (European project)
HR	High Resolution
LC	INSPIRE theme Land Cover
LPIS	Land Parcel Identification System
LU	INSPIRE theme Land Use
LUCAS	Land Use/Cover Area frame Survey
SDG	Sustainable Development Goal
UN-GGIM	United Nations initiative on Global Geospatial Information Management
WGA	(UN-GGIM: Europe) Working Group on Core Data

2.5 Glossary

2.5.1 Levels of detail

Global	Level of detail defined by ELF: data to be used generally at scales between 1:500,000 and 1:1,000,000, i.e. mainly at international level
Regional	Level of detail defined by ELF: data to be used generally at scales between 1:100,000 and 1:500,000; data mainly for national or regional (European or cross-border) actions.
Master level 2	Level of detail defined by ELF: data to be used generally at scales between 1:25,000 and 1: 100,000); data mainly for regional (sub-national) actions.
Master level 1	Level of detail defined by ELF: data to be used generally at scales between 1:5,000 and 1:25,000; data mainly for local level actions.
Master level 0	Level of detail defined by ELF: data to be used generally at scales larger than 1:5,000; typically, data at cadastral map level, mainly for local level actions.

NOTE: the above definitions are indicative; in practice, detailed data (Master levels) may also be required by national, European or even international users.

2.6 Reference documents

INSPIRE Data Specification on Land Cover– Technical Guidelines 3.1:
http://inspire.ec.europa.eu/id/document/tg/lc

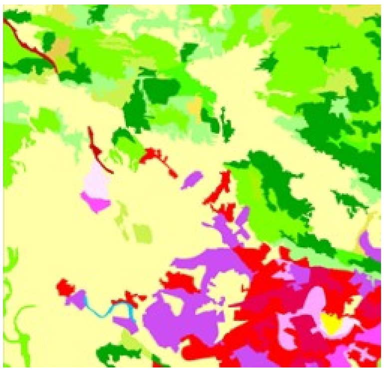
3 Overview

3.1 General scope

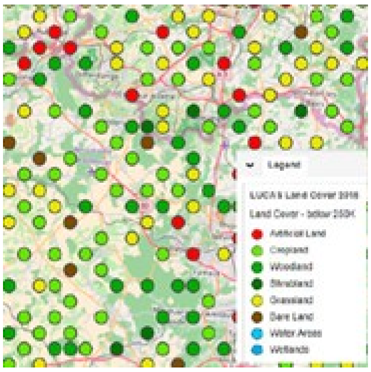
<u>Definition</u>: Land cover (LC) is an abstraction of the physical and biological cover of the Earth's surface including artificial surfaces, agricultural areas, forests, (semi-)natural areas, wetlands, water bodies, according to its physical characteristics.

This definition comes from INSPIRE Data Specifications on Land Cover. This abstraction is shown in a categorisation of the different covers on the Earth's surface, categorisation that is expressed in several ways (traditional classifications and nomenclatures, or object-oriented data models), usually according to the purpose of use of the resulting dataset.

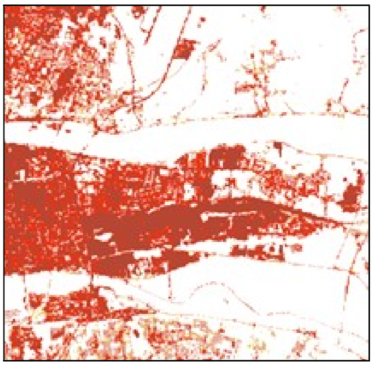
More detailed comparison with INSPIRE is available in Annex A.



Corine Land Cover



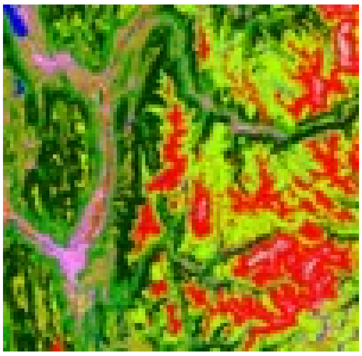
LUCAS



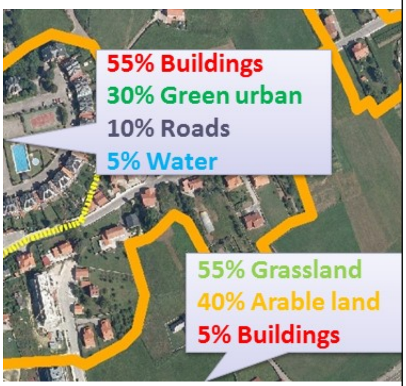
HR imperviousness changes

Examples of pan-European data

Finland



France



Spain

Examples of national data

Figure 1: Examples of Land Cover data

Most common geometries of the land cover dataset are vector polygons, vector points or raster cells and grids whereas lines may be used in specific cases (such as hedges). Land cover data may exist and be required at various levels of detail, from local to global.

General approach

More details about the methodology followed by the UN-GGIM: Europe WGA can be found in annex B.

- **Land cover and land use**

This deliverable follows the INSPIRE principle of clear separation between land cover and land use. Existing land use (including crop maps) and planned land use are under deliverable "Core Spatial Data Theme Land Use".

- **Flexible approach**

Land Cover data are generalised products that may use a variety of modelling options addition and whose production methods are significantly evolving. As a result, it makes it difficult or even irrelevant to recommend a set of well-defined products, as done for the other core themes.

In practice, WGA has identified a trend and requirements for large scale data, i.e. Master level 1 data for which a methodology and some common characteristics have been recommended and a strong interest to keep the temporal series of Corine Land Cover data at regional level. Core data recommendations and good practices are based on these two key drivers.

3.2 Use cases

The simplest use of land cover is to locate places of interest, such as forests, wetlands, built-up areas etc. especially at medium and small scales. This enables to find a first assessment of relevant place for a new construction project, e.g. to delimit the coarse area where to build a new road.

Land cover data and its dynamics may be used to understand and forecast propagation of various natural or human phenomena. For example, the degree of soil sealing helps to predict the evolution of a flood flow over time; the nature of land cover is necessary to identify the areas prone to erosion or to predict how pollution will propagate; the evolution of built-up areas will inform deciders and citizens about the land take issue. Typically, land sobriety is often promoted to reduce soil artificialisation and therefore, there is need to be able to quantify consumption of natural and agriculture spaces.

Land cover and, even more, existing land use are key aspects for the assessment of human foot print, they enable location of human activities and to evaluate their pressure on environment and natural resources, such as land or water.

Some types of land cover are likely to have a bigger role in the evolution of biodiversity (e.g. permanent grasslands, ponds and wetlands). Land cover is strongly required to understand ecosystems, it is strongly linked with habitats and biotopes and with species distribution. This knowledge is necessary to take relevant decisions: which agricultural land to be protected from urbanisation when preparing a new spatial plan? Which agricultural practices to be recommended? etc.

Land Cover data are is also quite critical for understanding and monitoring climate change by contributing to assess the greenhouse gas emissions or storage.

From land cover, it is possible to derive other data, such as ecological networks or ecosystem accounts that may be used in operational phase, for instance to assess the environmental impact of a project; it is also possible to derive various indicators that are quite useful for the reporting to some European Directives and may be in future for SDG indicators.

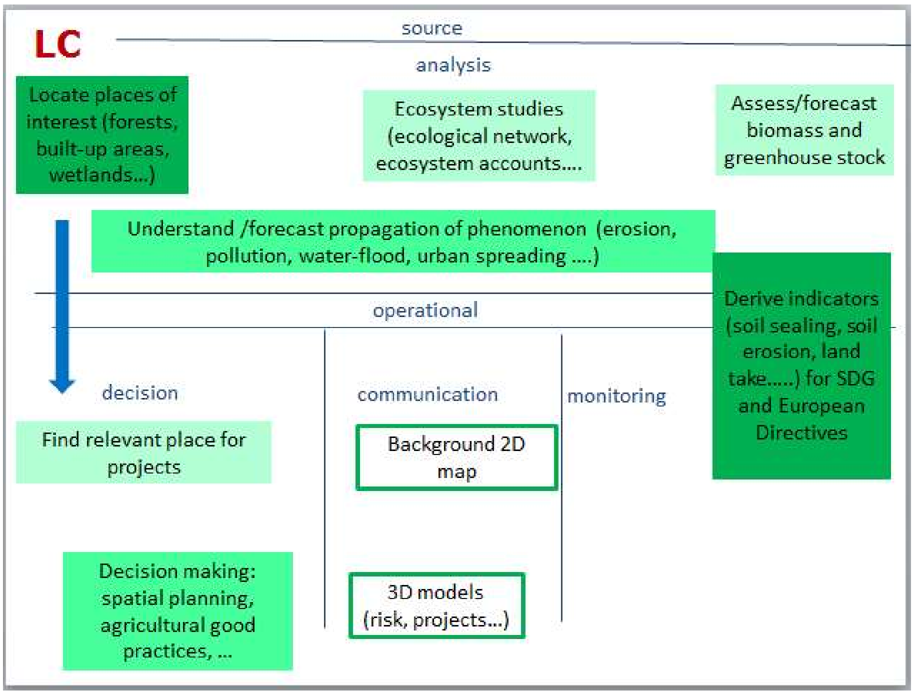


Figure 2: Map of use cases for theme Land Cover

4 Data content

4.1 Core data at national level

4.1.1 Methodology,

Production of large scale Land Cover data is necessary to support SDG related policies and to enable their monitoring on a fair and reliable basis. There is no perfect Land Cover product able to satisfy all user requirements at reasonable cost but a set of various approaches having their advantages and drawbacks.

Production of large scale Land Cover data implies significant investment. So, it should be ensured that the results will be used by all target stakeholders, ideally by all potential users or at least by those considered as priority ones.

The rationale for choices (main characteristics of the data, production method) should be decided in a collaborative way or at least clearly explained and widely communicated; for instance, a relevant balance between user requirements and feasibility has to agreed.

<u>Good Practice 1</u>

Common use at national level of large scale Land Cover data should be ensured; an agreement on such product characteristics should be found by coordination between involved stakeholders and if possible, recognised by regulation.

NOTE: The Working Group has conducted wide investigation in order to decide of relevant recommendations for core data; this investigation includes the identification of main Land Cover characteristics and some assessment of the benefits and drawbacks of possible choices about these characteristics.

The results from this investigation may be a good starting point for data producers having to design a new or upgraded Land Cover product. It would provide them a minimum list of questions to wander about and some rationale for choice, i.e. some methodology to be followed. This is documented in annex B of this deliverable (mainly chapter 2 about key findings).

4.1.2 Geographic extent

Core Recommendation 1

Geographic coverage of core Land Cover data should be whole territory.

NOTE: The whole land territory includes of course the land itself but also inland waters.

4.1.3 Spatial resolution

The general recommendation is to target scale around 1:10,000. However, for vector data, the Minimum Unit of Mapping may vary according to the user requirements (e.g. the public policies to be considered) and to the landscape characteristics. For instance, the Minimum Mapping Unit may be smaller in urban areas than in rural ones, it may also be smaller in a small parcel country such as Malta than in large forests of Nordic countries.

There are some countries which actually generated LC geometries referenced to cadastral or LPIS parcel geometries, which in practice means a 0.05 ha spatial resolution or equivalent cartographic scales associated with cadastral resolution. This trend should be encouraged in order to favour integration of Land Cover data with other Core data Themes such as Cadastral data.

For raster products, the Minimum Mapping Unit is the pixel; this generally implies the same MMU on whole territory. Resolution close to 1 m is considered as balanced compromise between good rendering of some Land Cover classes (such as built-up areas) and reasonable data volume to be handled by users.

4.1.4 Type of geometry

Core Recommendation 2
Land Cover data should be provided on pixels and/or on polygons.

NOTE 1: A spatial representation by polygons or pixels ensures that the whole territory is covered by LC information, what is a frequent user requirement.

NOTE 2: If affordable, a sufficiently dense sampling of points might be a complement, especially for statistical or monitoring purposes. For more details, see chapter 6.6.1.

4.1.5 Semantic - Classification

Core Recommendation 3
Each spatial object should have an attribute with its (main) Land Cover class.

NOTE 1: As stated above, the spatial object may be a pixel (raster data) or a polygon (vector data)

NOTE 2: The attribute is about the LandCover class; this implies that the classification should be only about land cover and should not mix land cover and land use concepts, as recommended by EAGLE and INSPIRE principles.

NOTE 3: In addition, the data may include more information about Land Cover. For instance, the other (secondary) LC classes may be documented, using the mosaic concept of INSPIRE. This may be useful for instance due to the generalisation process or due to the use of cadastral parcels as polygons.

NOTE 4: This core recommendation does not forbid collecting Land Cover and Land Use in same product but these two concepts should be modelled in separate attributes. Collecting LC and LU on same spatial objects offers wider opportunities of data analysis but it may generate some conflicts for the choice and delineation of these spatial objects, as there is no systematic correlation between LC and LU.

Good Practice 2
It is recommended to use a hierarchical classification based on or matching with EAGLE concepts, at least at its highest level(s).

NOTE 1: A hierarchical classification is a good way to manage several levels of detail in semantic and to enable some flexibility between various levels of government. On one hand, the national classification may be enriched by lower levels in order to produce more detailed data LC at local level. On the other hand, the national data might be merged according the highest level(s) of the classification in order to derive pan-European or cross-border harmonised data.

NOTE 2: The highest levels of EAGLE classification are based on 3 classes and 8 sub-classes: abiotic (artificial / natural materials), biotic (woody / herbaceous / lichen, mosses and algae / succulent and others), water (liquid / solid).

Abiotic: physical rather than biological; not derived from living organisms.

More details may be found in:

https://land.copernicus.eu/eagle/files/explanatory-documentation/eagle-docu

NOTE 3: The Land Cover classification to be used for core data should ideally be a national standard based and compatible with European classifications, such as the PLCC (Pure Land Cover Classification) proposed by INSPIRE or as the classifications used under Copernicus.

4.1.6 Temporal aspects

Core Recommendation 4
Land Cover data should have a well-defined reference date or period.

NOTE 1: Land Cover is often used to assess landscape evolution. This is why it is important to document in a reliable way the reference date.

NOTE 2: If the reference date is not the same on whole national territory, this implies either that Land Cover data should be provided in several data sets, each of them having same temporal information or that the temporal information has to be delivered with each object.

NOTE 3: It is advised to provide a reference date as much as possible and this is quite easy when the LC data is derived from image(s) taken on the same day. Providing a reference period may be envisaged if the production process is based on the exploitation of a temporal series of images.

Good Practice 3
The update frequency of core Land Cover data should ideally be equal to or better than 3 years and in any case be better than 6 years.

NOTE: The general trend is that users are requiring frequently updated data and are often mentioning 3 years or even less. However, this target may still be difficult to be achieved.

4.1.7 Quality

Core Recommendation 5
Geometric consistency with core topographic data, such as transport network and hydrography, should be ensured.

NOTE 1: It is advised to specify the Minimum Mapping Width (MMW) for linear features. This will contribute to specify the elements from transport and hydrography data that are expected to be represented as land cover features. For a target scale around 1: 10,000, a MMW of 5 m is reasonable.

NOTE 2: A possible way to ensure this consistency is to use core topographic data in the production process of land cover (e.g., for preliminary segmentation, for validation process).

Good Practice 4
Production methods have to be adapted to the target semantic accuracy and to expected level of details in the classification.

NOTE 1: A detailed classification increased the risk of confusion between classes and ensuring both detailed and reliable LC data may be costly. A relevant balance has to be found between the classification level of detail, the target semantic accuracy and the production cost.

Good Practice 5
Backward compatibility with previous products should be ensured as much as possible.

NOTE 1: One of the main use cases of Land Cover data is the assessment of land cover changes, what can be done only if data is captured at different dates according similar or at least compatible specifications. Typically, Land Cover data for statistics requires comparable time series.

NOTE 2: However, ensuring backward compatibility may raise issues, if it implies to keep the drawbacks of an old-fashioned product, though the technical progresses would enable to get rid of them.

4.2 Core data at European level (Regional)

Many of the SDG related use cases require Land Cover data across several years, with as much as possible persistent specifications in order to enable reliable comparison between several years.

The pan-European product Corine Land Cover offers a very valuable input by supplying both several versions of Land Cover data (1990, 2000, 2006, 2012, and 2018) at resolution 25 ha and a layer of changes at resolution 5 ha between two consecutive versions.

For the future, the Corine Land Cover second generation initiative is planning the CLC Legacy product whose purpose is to ensure the backward comparability with previous versions of Corine Land Cover.

Good Practice 6
Existing CLC data sets and the upcoming CLC Legacy instance are considered as main input for core data at Regional Level, due to the key reason that they allow backward comparability between reference years, are a de facto standard for LC European databases, and constitute also key information for understanding the dynamic of LC&LU in Europe.

NOTE 1: It is up to the Copernicus CLC 2nd generation programme to define the detailed specifications of the CLC Legacy product, taking into account both user requirements and feasibility constraints.

NOTE 2: Corine Land Cover is not available on whole Europe but only on EEA39. Unfortunately, it seems very difficult to make missing countries benefit from this temporal series of Land Cover data at regional level.

NOTE 3: The current Corine Land Cover classification is composed of 44 classes, some of them mixing Land Cover and Land Use concepts (what is not in line with the recommendations of this deliverable). However, the benefits of backward comparability are considered as overcoming the drawbacks of an old-fashioned classification.

5 Other recommendations

5.1 Coordinate Reference System (CRS)

5.1.1 Horizontal component

Good Practice 7
Core data should be stored and managed in a CRS based on datum ETRS89 in areas within its geographical scope, either using geographic or projected coordinates.

NOTE 1: Geographical scope of ETRS-89 excludes oversea territories, such as Canary Islands or French Guyana or Madeira Islands and Azores Islands. In these cases, it is recommended to use a CRS based on ITRS (International Terrestrial Reference System).

NOTE 2: Storing and managing data in CRS based on international datum facilitates the import of measures from modern sensors, ensures that data is managed in a well-maintained geodetic framework and of course, facilitates the export of data into international CRS (e.g. those mandated by INSPIRE).

NOTE: If core data at regional and global levels has to be provided as a single data set on an area including oversea territories, it is recommended to use as CRS geographic coordinates with any realisation of the International Terrestrial Reference System (ITRS), known as International Terrestrial Reference Frame (ITRF). At small or medium scales, all ITRS realisations can be considered as equivalent, as deviations between them are negligible compared to data accuracy.

5.2 Metadata

<u>Good Practice 8</u>
Core data should be documented by metadata for discovery and evaluation, as stated in the INSPIRE Technical Guidelines for metadata and for interoperability.

NOTE: This is an INSPIRE recommendation (only the INSPIRE Implementing Rules are legally binding for the Member States belonging to the European Union, but the Technical Guidelines are considered necessary to make the European Spatial Data Infrastructure work in practice). For the other countries, this is a way to make their data easily manageable by transnational users.

5.3 Delivery

It is expected that core data will be made available through improved existing products (or new products) or as INSPIRE data, and perhaps as specific core products (delivery issues still have to be investigated by the working group).

<u>Good Practice 9</u>
Core data corresponding to INSPIRE theme Land Cover should be made available according to the INSPIRE Technical Guidelines for interoperability, for metadata and for services.

NOTE: this is an INSPIRE recommendation (only the INSPIRE Implementing Rules are legally binding for the Member states belonging to the European Union, but the Technical Guidelines are considered necessary to make the European Spatial Data Infrastructure work in practice). For the other countries, this is a way to make their data easily manageable by transnational users.

In a first step, Land Cover data is often produced in raster format, with the semantic information carried at pixel level. This is typically the case when Land Cover data is produced by automatic classification from aerial or satellite ortho-images. In a second step, the information may be submitted to generalisation and validation processes in order to get vector data.

<u>Good Practice 10</u>
In case the production process provides both raster and vector data, it is recommended to provide both layers to users.

NOTE 1: The raster layer may be a “heat map” coming from Artificial Intelligence process and providing for each pixel, the best candidate LC class and its probability.

NOTE 2: On one hand, the generalisation process is quite useful, by decreasing noise and smoothing data directly derived from images. The final vector data is of smaller volume and is generally considered of simpler use. This is why supplying vector data to users is recommended.

NOTE 3: On the other hand, the generalisation process is also forcing to merge some small areas of a well-identified LC class to another one, according to more or less arbitrary rules and is by nature implying some loss of information. This is why the intermediary raster data may be of interest for users willing to keep as detailed information as possible.

6 Considerations for future

6.1 New production methods

Most of current Land Cover data have been produced by partly automatic processes from orthoimages coming from planes or from satellite sensors. The arrival of Sentinel images (resolution: 10 m, frequency visit around 5 days, free access) has opened lots of opportunities for the production of Land Cover data. For instance, in the context of CLC second generation, the Copernicus program has launched the production of CLC Backbone, a simple pure land cover product that may be currently derived in an automatic (or almost automatic) way from Sentinel-2 images.

However, there is still place for improvements in order to exploit better the full potential of Sentinel images and to overcome their limits (such as lack of observation due to clouds). Artificial Intelligence with the progress of machine or deep learning is good candidate to enable more automatic processes and more reliable results. Research on these automatic methods should obviously be promoted. Investments for collecting and sharing field data that might be used as training data or as uncontested reference for comparing classification methods should also be encouraged.

A promising topic is the availability of temporal series of Sentinel images that enable to monitor the vegetation cycle. The Copernicus programme is already proposing some phenology services and more products (e.g. crop maps) are under development. The combination of optical observations (Sentinel-2) and of radar observations (Sentinel-1) is also a topic for further investigation.

Use of other Earth Observation satellites and of Voluntary Geographic Observation (e.g. to collect in-situ data) are also topics of interest for future.

All concerned stakeholders are incited to initiate or pursue watchdog, research and knowledge exchange activities. In addition, as this deliverable "Core spatial data theme Land Cover" has strongly taken into account the feasibility aspect, most of the recommendations are based on current state-of-the-art regarding land cover data production methods. An update of this deliverable might be envisaged in a few years in order to take into account the on-going progresses.

However, it is out of scope of the UN-GGIM: Europe Working Group on Core Data to propose solution to this political issue.

6.2 Comparing and combining land cover data

6.2.1 Comparing data with the EAGLE concepts

There are various products on land cover data with heterogeneous specifications, especially regarding land cover classifications. However, there is often need to compare and combine these various products, for instance when a user is willing to use national data from several countries (or local data from several zones) or when a user is willing to use data on same area but with different classifications.

Combining data from these various products require having some means to understand and compare the products and especially the land cover classifications. This is already more or less available through the EAGLE concepts and tools: the main idea is to describe land cover classes as a set of basic object and parameters (descriptive approach or object-based approach). These principles have been globally adopted by the 19144-2 standard on Land Cover Meta Language.

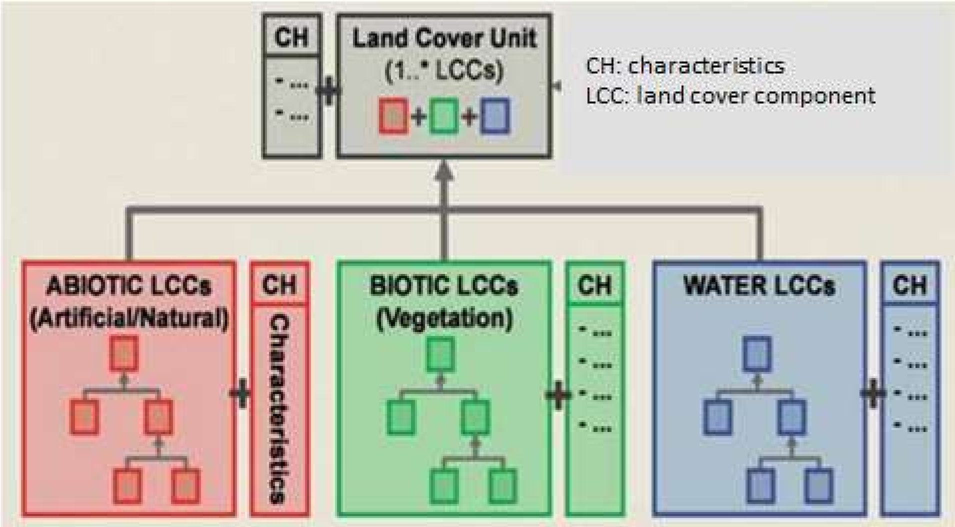


Figure 3: The EAGLE data model

The EAGLE principles have been widely adopted by the INSPIRE themes Land Cover and Land Use. For instance, it is recommended to document the land cover classification of a national or regional product using the standardised Land Cover Meta Language.

In the CLC 2nd generation initiative, the EAGLE concepts are expecting to be used as "translation" tool between pan-European products and products at national or subnational levels.

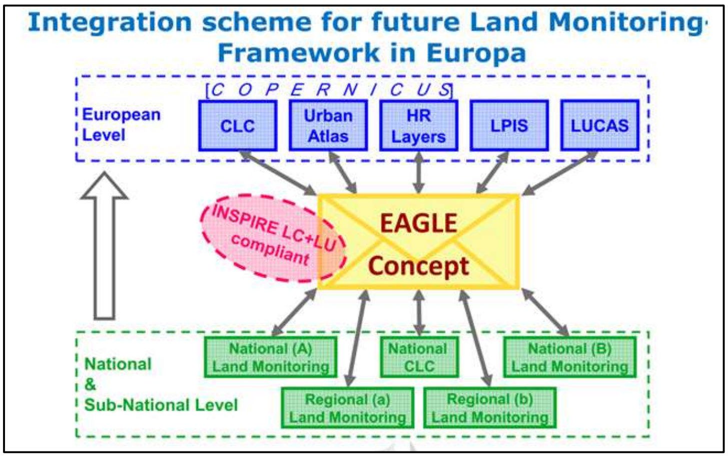


Figure 4: EAGLE Concept use in data exchange

Significant work has already been achieved but there is still place for progress. For instance, the Working Group ISO TC 211 is working on new proposal regarding support for registers (classification systems, schemas, elements, characteristics, etc.) on land cover.

UN-GGIM: Europe Working Group on Core Data encourages the standardisation activities around land cover classification description, the main need being probably about spreading knowledge and adoption of this standard by the concerned stakeholders.

6.2.2 Deriving products from the Data Cube

There are various land cover or land cover related products at European level. The Copernicus initiative on CLC 2nd generation is envisaging among its products, CLC Core a land cover-land use knowledge data cube, the spatial domain being composed of a regular grid of pixels whereas the semantic is composed of various layers, each of them being described according to the common EAGLE concepts. This data cube should enable to store not only the various pan-European products but also national or regional contributions. Several CLC Instances might be then derived from CLC Core, depending on specific requirements (mainly on European Directives). Some European countries are also envisaging a similar approach for their national needs.

The data cube approach looks very promising but it also raises big challenges: how to make a reliable product from various data sources (different dates, different resolutions…) and how to document this heterogeneous quality?

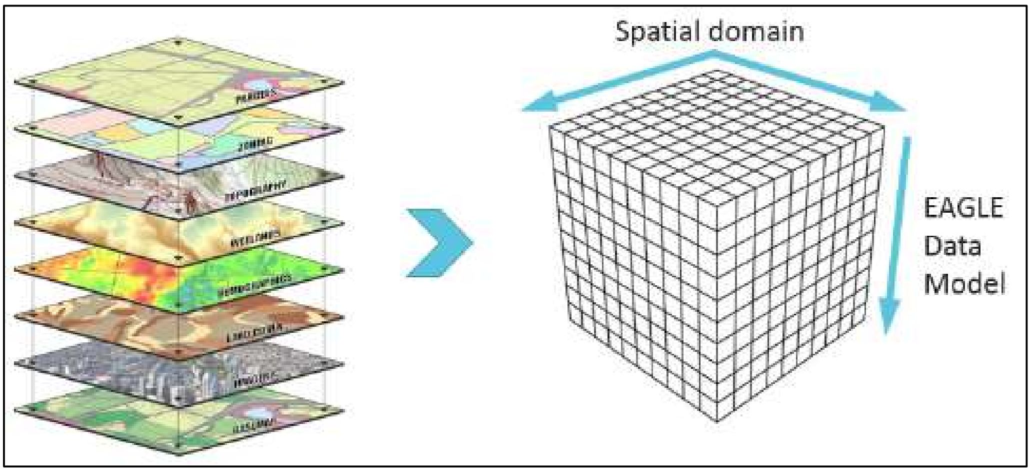


Figure 5: CLC Core data cube

Further researches and tests (with user involvement) will be required in future to provide methodologies, guidelines, tools enabling easy combination of various data and derivation of relevant products from the data cube.

6.2.3 Comparing LC data from various temporal versions

Some users are interested in getting change only products showing the Land Cover evolutions since a previous version. This practice is already in place for a few products (e.g. Corine Land Cover).

In theory, users can compute by themselves the Land Cover changes if they have the whole data sets at the 2 dates of interest. However, in practice, especially with vector data, this computation may be cumbersome and may provide disappointing results, difficult to be exploited (e.g .micro-surfaces just due to lightly different data delineation capture).

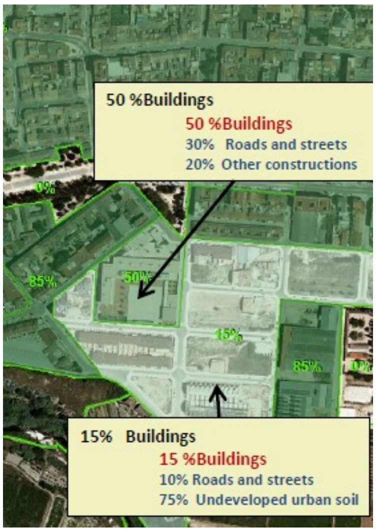
It is questionable if data producers should provide change only products. On one side, it would factorise the efforts to get meaningful products showing the Land Cover evolutions. On the other side, it is unsure that all users are expecting the same kinds of change only data; for instance, how many previous temporal versions should be taken into account: just the last one or more? what should be the MMU for a Land Cover change zone? how to document the change (keep same classification with old and new values or use a classification of changes, such as “glacier to bare soil”?

Knowledge exchange about solutions adopted or envisaged by data producers and feed-back from user experiences would be very useful and should be encouraged.

6.3 Exploiting complex data

For many years, land cover users have worked with CLC data or CLC inspired data, i.e. with “simple” products composed of geometric objects with one semantic attribute whose values were in a single nomenclature, generally mixing land cover and land use concepts.

The current trend is to better and richer products that apply a clear separation between land use and land cover, that may use a descriptive approach (by providing several Land Cover values on a feature) or that may also include land cover parameters. Typically, the INSPIRE data model is allowing all these various ways of providing land cover information.



Example of a descriptive approach

Example of parameter use

Figure 6: Examples of complex Land Cover data

However, these elaborated products are often considered as “complex” by users who meet difficulties to do their habitual tasks with these new and unfamiliar data and who may lack of motivation and imagination to use the potential of these products for carrying on better or new analysis.

The situation might even be worse with the potential future products of heterogeneous quality derived from the Land Cover- Land Use data cube.

Production of elaborated land cover products should be accompanied by capacity building activities towards users: communication, raising awareness, guidelines for use, training…

6.4 Use of Land Cover data for indicators

Land Cover data has to face the following paradox. On one hand, land cover data is often used to monitor environmental European or national regulations; this reporting is achieved through indicators being defined using land cover data and generally reported on regular basis. The values or evolution of these indicators may have significant impact of the lower-level policy makers (penalties, need of corrective actions…).

On the other hand, land cover data is by nature generalised data and so, not fully reliable data. In practice, any generalisation implies the sacrifice of detailed information to get a more meaningful overview. For instance, in a vector product, small areas under the MMU threshold will have to be merged with a neighbour area. In addition, even with detailed specifications, the production of land cover data always gives place to various interpretations.

Monitoring has to be fair so there is a need for homogeneous data across space to ensure comparability between territories.

Monitoring has to be effective so there is need for good quality enough data in order to ensure that the indicator reflects correctly the reality. Unfortunately, this is not always the case currently as getting correct thematic accuracy is quite costly (e.g. automatic process of satellite images lead to 90% thematic accuracy in best cases). The case is even worth when the purpose is to assess land cover evolutions, these evolutions being generally smaller than the data accuracy. In addition, for evolution indicators, it is very important that the land cover data reflects only the evolutions in real world and not the changes in the production process.

Research to improve land cover data (especially its thematic accuracy) should be encouraged together with knowledge exchange and methodology work to investigate the good practices related to use of land cover data for environmental indicators. For instance, experiences might be conducted to get a quantitative assessment of the reliability of the indicators as a result of the generalisation of the Land Cover maps.

6.5 Production of historical Land Cover data

It may be required for some SDG related policies to understand the evolution of land cover during a long period and not first from the last twenty or thirty years for which digital land cover data are available.

Many providers may have sets of aerial photographs that were taken before the ninety-nineties; they may also have historical maps even for previous centuries.

The production of digital land cover data using these historical images or maps raises lots of difficulties both in georeferencing and in semantic interpretation.

However, some countries may have initiated such production at least for research purposes. Sharing these experiences should be encouraged in order to gain better knowledge about potential methods and difficulties and about the costs-benefits to be expected.

6.6 Other geometries for Land Cover data

6.6.1 Point representation

The representation by pixels or polygons ensuring whole coverage of territory is most common user requirements. This is why this document recommends a polygon or pixel representation for core land cover data.

However, for some purposes, a point-based approach offers significant advantages. A point-based approach allows a more accurate classification of the land cover, especially in complex areas (e.g. urbanised areas) where a polygon-based approach tends to make small features disappear because they are merged into larger features.

A point-based approach also has advantages for classifying phenomena with unclear margins: in most cases, it is possible to assign an unambiguous classification to a given point, whereas it is difficult to delimit the contour between two phenomena when the transition zone is wide. This precision of classification offers significant benefits for statistical analysis as well as for monitoring the land cover change.

A point approach also makes it easier to avoid erroneous changes due to methodological developments. Unlike a polygon whose delimitation can be affected by the methodology used, the point is not affected in its geometry. This is particularly important for monitoring land cover change.

Therefore, Land Cover point based data may be useful complementary product, in addition to polygon or pixel core data. It is true that this point approach requires dense sampling with a relative high cost.

Research, knowledge sharing, cost-benefit analysis are encouraged to better understand how point-based land cover data may be produced, if it is worth to maintain such a product in addition of core (polygon or pixel) core data and how these two products can be articulated one with another.

6.6.2 Linear representation

Current Land Cover data is generally represented by surfaces (polygons or pixels) or by points. However, due to the generalisation effect, some land cover structures, though being of interest for several use cases, may disappear because not wide enough to be represented as surfaces. This is for instance the case of roads, railways, rivers, hedges, shore. This is not an issue for roads, railways, and rivers that are generally present in other themes (respectively Transport and Hydrography) but hedges and shores may be missing if land cover data is represented only by surfaces.

Potential solution would be to complement classical surficial land cover by inventory of hedges and possibly shore, represented as linear elements. Research and knowledge exchange should be encouraged to investigate how to produce and use efficiently such potential inventories.

6.6.2.1 Volumetric representation

Most if not all current land cover products are providing information as 2D data, i.e. as a single layer of the Earth surface coverage.

However, the statement that observation from above is often considered to offer a correct and complete representation of the physical land cover object (phenomenon) on Earth is not always true as shown by following illustration.

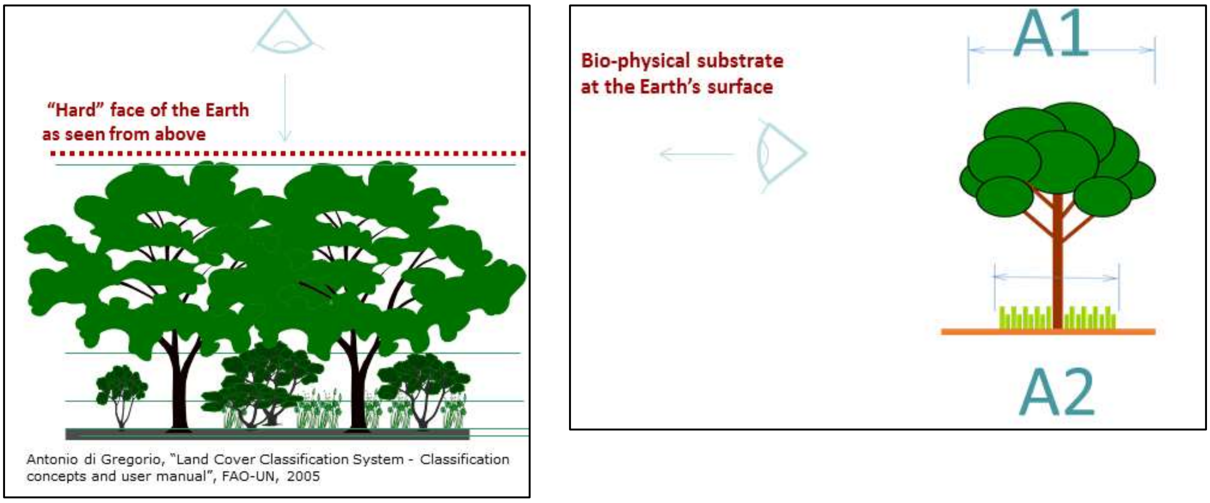


Figure 7: Several strata of land cover (at different heights above ground)

Similar cases may occur in urban areas, with green roofs and land cover being "built-up area – artificialised" at ground level and "vegetation" a few meters above.

These cases where land cover data should ideally be provided according several heights on or above ground, i.e. as 3D data, are addressing a few specific use cases but most of them are very significant for especially for climate change: several strata of trees won't have the same effect on greenhouse gas effects, green roofs (or façades) will mitigate the urban heat islands.

In addition, information about volumes of buildings is useful to assess urbanism densification whereas information about volumes of wood is necessary to measure the biomass and to assess the resource.

Research activities about modelling and production of 3D Land cover data should be considered at least to deal with some specific types of landscapes.

7 Annex A: Relationship with INSPIRE

7.1 Data model

The UML models provided in this annex are only graphical illustrations of the core recommendations and of the good practices present in this document.

The UML illustrations are pure extractions of the INSPIRE conceptual model for theme Land Cover.

The recommendations for content for Core Spatial Data theme Land Cover are represented by highlighting the selected attributes in green.

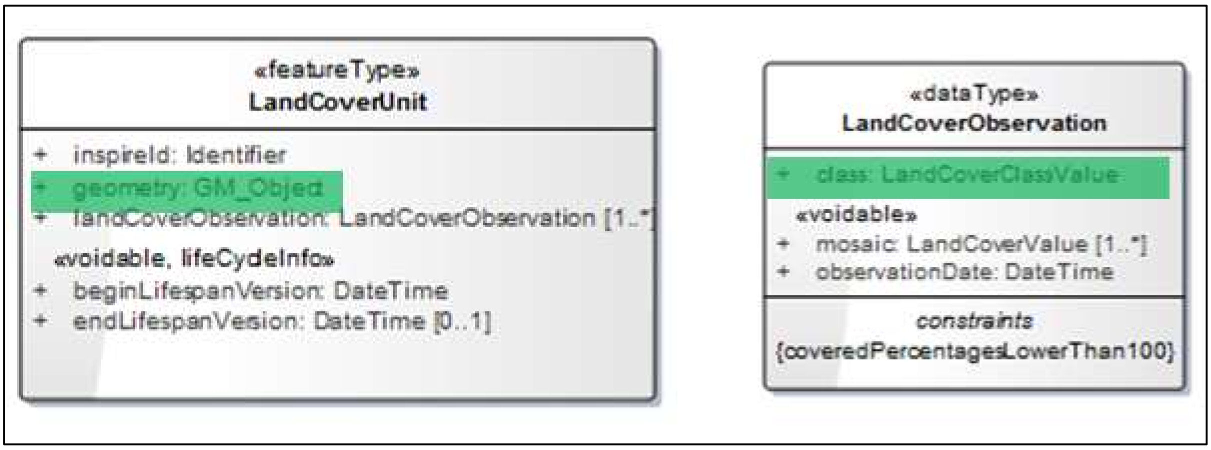


Figure8: Comparison between core data and INSPIRE

The INSPIRE data model provides a rich framework for exchanging the various existing land cover data sets in a very well documented model. In opposite, core data is expecting quite simpler products as shown by the above illustration.

For instance, in INSPIRE, temporal aspects may be documented at observation level whereas core data is rather expected whole data set with a single reference date (to be documented just as metadata).

7.2 Other topics

The INSPIRE data specifications are quite flexible and in practice, they don't provide any idea of the expected content: no requirement regarding the level s of detail, the geometric representation, the classification to be used or the production frequency.

This deliverable provides mainly recommendations about the expected levels of detail with for each of them some considerations about geometric representation (pixels or polygons), about the type of classification to be used and about the production frequency. However, there remains a lot of flexibility especially for the large scale data (Master Level 1)

8 Annex B: Methodology

8.1 Understanding the context

In order to understand the context about land cover production and use, UN-GGIM: Europe has investigated the following initiatives.

8.1.1 At European or international level

8.1.1.1 EAGLE and ISO 19144-2

The EAGLE group was created by the EIONET community in 2009 in order to improve the current design of land monitoring for Europe, solving the limitations of traditional hierarchical classifications. Main principles of the EAGLE group consist in a clear separation between land cover and land use concepts and in using an object-oriented description instead of classification. The EAGLE group has provided not only concepts for defining Land Cover classification in an unambiguous way but also practical tools to implement these concepts (e.g. a matrix allowing to describe any existing land cover classification according to the EAGLE data model).

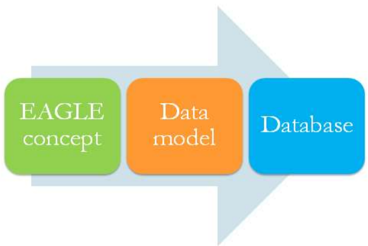


Figure 9: The resulting products of the EAGLE group

The EAGLE data concepts are very similar to the ones of the ISO TC211 standard 19-144-2 called Land Cover Meta Language (LCML).

8.1.1.2 INSPIRE

Following the EAGLE principles, the INSPIRE Directive has defined two separate themes for Land Use and Land Cover.

The INSPIRE Data Specification on Land Cover has modelled the various ways to describe Land Cover data in order to be applicable to the wide range of existing Land Cover datasets So national and European Land Cover data providers can use many modelling choices for INSPIRE dissemination services, allowing various geometric representations, different classifications & data modelling approaches.

INSPIRE Data Specifications on Land Cover, based on the existing submitted reference material and use cases, allow to represent land cover data produced by traditional classification, but also other data including object oriented and parametrized information regarding land cover characteristics, and geometric land cover data which may be vector points and polygons, or raster cells, corresponding to the common methods of observation used in both pan-European and national land cover mapping and monitoring.

As a result, WGA has got from INSPIRE Data Specification on Land Cover a valuable framework providing the potential options for land cover modelling but limited help, regarding thematic content, the INSPIRE specifications being very flexible about levels of detail, geometric representations, classification…

8.1.1.3 Corine Land Cover second generation

The main pan European LC/LU product for Europe is Corine Land Cover (CLC), initiated at the end of the 80s decade last century by the European Environment Agency (EEA), and which became part of the portfolio of the Copernicus Land Monitoring Service in 2013. There are several long time series on key land cover (LC) and land use (LU) aspects at European level (1990-2000-2006-2012-2018 and changes).

With more demanding Community policies, Copernicus Land Monitoring Service is going to upgrade CLC in the second generation of European LC/LU product, under the name of CLC+ or CLC second generation. This next generation Corine Land Cover products (CLC+) is being designed as a logically structured series of subsets, ranging from VHR input layers (CLC-backbone, vector) into a core LC/LU graphical database (CLC-core, grid) to instances, which become on-demand end user products (CLC+, CLC-legacy or heritage).

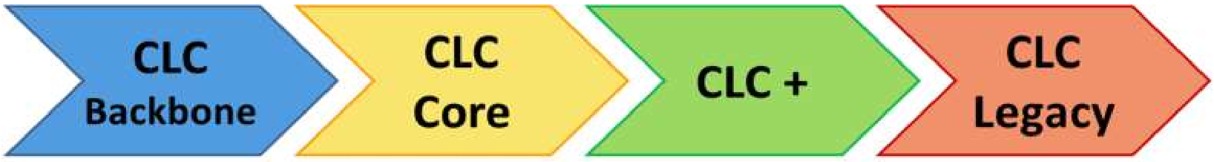


Figure 10: the family products of CLC 2nd generation

It will be based on the conceptual framework elaborated by the EIONET EAGLE group, and will provide a basis to cope with challenging (European) user requirements for LC/LU information for the coming 10 to 15 years. CLC+ is expected to support the LULUCF reporting obligations (obligatory from 2021 onwards), at 0.5 ha MMU, and, building upon the EAGLE concept, will facilitate more analytic mapping of LC, as well more complex querying, and enrichment of the data and by flexible integration with other datasets.

More detailed information on CLC+ can be found here: https://land.copernicus.eu/user-corner/technical-library/upcoming-product-clc

8.1.2 At national or sub-national level

UN-GGIM: Europe WGA has also performed some investigation about the current initiatives, on-going at national or sub-national levels. This investigation has been done through some users or producers interviews, through bibliography, through the workshop co-organised in November 2018 by the EEA, by EuroSDR and by the INSPIRE Knowledge Exchange of EuroGeographics and of course through the expertise of WGA members.

Presentations and minutes of the workshop may be found on:

http://www.eurogeographics.org/event/land-useland-cover-products-challenges-and-opportunities

8.2 Key findings

8.2.1 Main characteristics of Land Cover data

The main characteristics of a land cover dataset are the following:

- Geographic extent: is the data set covering the whole territory or only specific areas?
- Level of detail: what is the expected scale or scale range for data use? In practice, the level of detail is expressed by the MMU (Minimum Unit of Mapping) and/or by the ground pixel size.
- Geometric representation: land cover data may be represented as polygons, as regular or irregular set of points or as pixels.
- Choice for modelling semantic information: is the data using the traditional approach (homogeneous polygons with a single classification) or the descriptive object-based approach (several land cover components and parameters on same geometry)? Is the data going to mix land cover and land use concepts or not?
- Semantic information: which classification will be used? How detailed is it (i.e. how many classes)?
- Production cycle frequency: how often will the data set be updated?

8.2.2 Wide variety of requirements and approaches

Land cover data is required by various stakeholders, at various levels of detail and for a wide range of domains (greenhouse gas, urban spreading, ecosystems…) and purposes (understanding, deciding, monitoring). For instance, several regulations at European and national levels require land cover data. This is leading to a wide range of specific requirements towards land cover data. These requirements coming from different stakeholders may conduct to contradictions:

- Need for comparable generic data across space for fair monitoring at high level of government versus need for detailed specific data at lower levels of government
- Need for persistent data across time (for reliable evolution indicators) versus wish of better products for next versions
- Need for good data (detailed geometry and semantics, high production frequency) that is is very costly to produce and not always affordable

This variety of requirements combined with the fact that **there is no obvious real-world object corresponding to land cover data** has conducted the land cover stakeholders to design a wide range of modelling approaches.

By nature, Land Cover data implies some generalization; as a consequence, there may be a wide variety of generalisation choices, for several of the characteristics listed above. This is the case of geometrical representations: grid points or pixels, polygons that may be homogeneous or not, that may be based on cadastral parcels or not… These geometric representations may also vary according to the expected level of detail. They may also depend on the choice of using same geometry (or not) to carry land cover and land use information.

This is also the case of semantics with a wide range of choices regarding classifications. There are many existing classifications, at European, national, regional or even local levels. These classifications may be one dimension (mixing LC/LU), 2 dimensions (separate LC/LU), multi-dimensions (including other characteristics), they may be very basic with 5-6 classes or very detailed with several hierarchical levels. Semantic content may also be enriched by parameters.

As a result, there is no perfect product able to satisfy all requirements; there is neither a minimum product able to satisfy most user requirements.

This fact has been recognised by the INSPIRE Data Specifications (no recommended content) and by the CLC 2nd generation approach: its third item CLC+ won't be a well-defined product but a family of CLC instances, each of these instances being derived from the CLC Core knowledge base according to a specific purpose (generally, reporting for a European Directive).

8.2.3 Key drivers

8.2.3.1 Importance of the production process

The specifications of Land Cover data can't be decided without having in mind the envisaged production process; in practice, the considerations of feasibility and costs & benefits may be driving the choice of the land cover main characteristics as much as or even more than user requirements. Using satellite or aerial images, exploiting them by automatic segmentation or by human photo-identification, using or not ancillary data, using or not field survey will significantly influence the level of detail that may be achieved, both in geometry and in semantic.

Typically, there is a current trend to get relatively cheap land cover products by exploiting free Sentinel images by automatic processes. However this choice entails a coarse classification (with a few classes) and relatively poor semantic accuracy.

If Cadastre is in charge of collecting land cover data, sometimes with an obligation for owners to declare this kind of information, the cadastral parcel may be reasonably considered as candidate geometry; this may be less obvious choice in other cases.

8.2.3.2 Regulations

European or national regulations are also a key driver for specifying new Land Cover data products. For instance, the pan-European product CLC Backbone aims to fulfil the requirements of the LULUCF directive; this why a pixel of 0.5 ha was chosen.

Nevertheless, it is often difficult to specify a land cover product based on the requirements of a regulation, that is too generic to enable to decide accurately on each land cover characteristic. In addition, a new land cover product requires significant investment, so it is generally better to design it in order to make it benefit to several use cases, which leads to compromises.

In practice, regulations are a key driver more to devote funds for land cover data than for its detailed design.

8.2.4 Separation between land cover and land use

Land cover and land use are related and often combined in practical applications and datasets, and it is usual that those combined datasets emphasize land use aspects in intensively used areas (e.g. settlements) and land cover aspects in extensively used areas (e.g. natural vegetation). Due to the broad variety of applications of LC and LU information, there are many existing classification systems and nomenclatures, which contain to some extent a mixture of land cover and land use class definitions. This is the case for instance in Corine Land Cover.

However, the EAGLE and INSPIRE principles of clear separation between land use and land cover concepts have spread among concerned data producers and most of the new initiatives related to new production of land cover data set have adopted these concepts.

However, there are different cases:

- Data producer is responsible only for land cover : it will produce just a pure land cover data set
- Data producer is responsible both for land cover and land use
    - It may provide separate products, land use and land cover having their own geometry and production cycle
    - It may provide a combined product, land use and land cover sharing same geometry and same production cycle but being documented through two different attributes, one about pure land cover one about pure land use.

8.2.5 User requirements

As said above, user requirements are quite various and it is not really feasible to provide an overview. However, some topics look common to most users.

8.2.5.1 Several levels of detail

There is a general consensus that several levels of details are required for land cover data: small and medium scale data is necessary to get an overview on wide territories and to enable homogeneous monitoring between territories whereas large scale data is required for providing detailed information on smaller territories and to deal with local specificities.

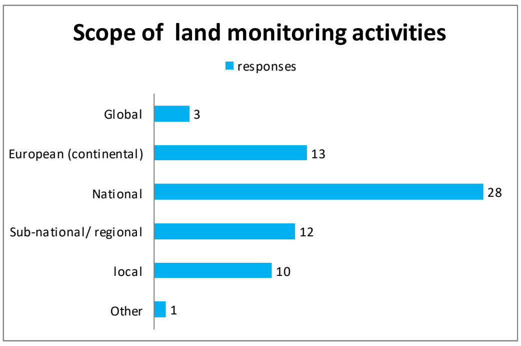


Figure 11: HELM project, scope of land monitoring activities

8.2.5.2 Getting used to separation between Land Cover and Land Use

For many years, users have worked with products mixing land cover and land use concepts in a single nomenclature. These products were familiar and considered as "simple".

Though understanding the interest of separating land cover and land use concepts, users are still a bit reluctant to move to new products or at least, they would welcome some support helping them to handle the new data.

8.2.5.3 Topographic use versus statistic use

The descriptive approach (by providing land cover information as a set of land over components and characteristics) is considered as adapted for statistical applications (such as computing indicators). However, many other users prefer a more traditional approach with homogeneous polygons carrying one land cover class: this is more adapted for mapping and viewing purposes; for instance, this enables them to recognise the underlying features (e.g. a wood, a sport field), this also enables them to combine LC in a meaningful way with other core information, such as roads, railways, rivers… By the way, topologic consistency between land cover data and topographic data is appreciated.

8.2.5.4 Good data

There is clearly no consensus about what is good land cover data, due to the variety of user requirements. However there are some general trends. First trend is toward more detailed data from geometric point of view. This is the case at European level with the migration from CLC to CLC 2nd generation, this is also the case in several Member States that have launched ambitious land cover products for scales around 1:10,000 or even 1:5,000. The second trend is towards most frequent production cycles: 6 years is considered as the minimum acceptable but most users require 3 years or even yearly updates.

Last but not least, the practice of hierarchical classification is appreciated as it enables generalisation of semantics and analysis at several levels.

8.2.5.5 Comparability across time and across space

This is a key requirement for monitoring use cases. To be comparable between territories, indicators have to be based on homogeneous and standardised data. In addition, for evolution indicators, data should be comparable across time, i.e. with persistent specifications.

8.3 From investigation to main decisions

8.3.1 Considering CLC 2nd generation

WGA work has considered the parallel work being done by the CLC 2nd generation initiative and its four foreseen products:

- CLC Backbone (simple product of pure land cover, derived from Sentinel images) is an initiative to get LC data at European level, at Master level 2 scale. WGA has considered CLC Backbone under considerations for future: once the data is available (what is not yet the case), the user feed-back will show if the product meets their requirements and quality expectations. For instance, CLC Backbone is currently not derived from and not consistent with Member States authoritative data.

- **CLC Legacy and its objective to ensure the continuity with the previous versions of CLC has been recognized as very valuable by WGA.** However, these products address only EEA39 and CLC Legacy is not yet accurately defined.
- The principle of CLC Core and CLC Instances has also been recognised of interest by WGA but the work was clearly not mature enough, this is why the ideas of CLC Core and CLC Instances are more relevant as considerations for future, i.e. as recommendations for research and future investigation.

8.3.2 Flexible approach for Master Level 1

Land Cover at Master Level 1 is often driven by requirements coming from national regulations; it is also strongly influenced by national specificities such as size and density of the country, type of landscape, existing LC/LU data producers, etc.

Therefore, proposing a common minimum product for all European Member States has been considered as very difficult or even irrelevant. Instead, WGA is just proposing some advices about a few key points. Though being a bit disappointing, this approach looks the most reasonable. WGA hopes it will help countries willing to upgrade their national data.