Prerequisites and criteria for aligning national/sub-national land monitoring activities

Deliverable n°: 4.1
Version: 2
December 2013

The activities leading to these results have received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 261562.
Content of the deliverable

Part 1: Database merging and grid approaches
Part 2: Synchronisation
Part 3: Nomenclature and Data models

This deliverable is a collation of the results of three HELM tasks. It covers the major technical means required to operationalise the alignment and integration of national/sub-national land monitoring activities across Europe. It reviews techniques of database merging applied to remote sensing and auxiliary surface data to enhance the information available. In this regard besides the well established polygon mapping, it also covers grid approaches as a promising alternative. Moreover, the deliverable investigates how the various actors involved in national/sub-national level land monitoring can move toward synchronisation among each other and with data requirements of the European actors, especially the European Environment Agency. The third chapter addresses Nomenclature and Data models. This text presents criteria and boundary conditions for a potential European land monitoring system/data model. It thereby links the findings of the HELM project and the EAGLE group (EIONET Action Group on Land Monitoring in Europe).

Authors:

<table>
<thead>
<tr>
<th>Part 1: Database merging and grid approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geir-Harald Strand (Norwegian Forest and Landscape Institute)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part 2: Synchronisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gergey Maucha (FÖMI, Institute of Geodesy, Cartography and Remote Sensing, Hungary)</td>
</tr>
<tr>
<td>Stefan Arnold (subcontracted by Environment Agency Austria) and Barbara Kosztra (FÖMI)</td>
</tr>
</tbody>
</table>

Contributors:

This document contains contributions by the entire HELM consortium. If needed, more specific information about contributions is provided at the beginning of the chapters.
Database merging and Grid approaches
Content of the deliverable

Chapter 1: Database merging and grid approaches

Chapter 2: Synchronisation

This deliverable covers the major technical means required to operationalise the alignment and integration of national/sub-national land monitoring activities across Europe. It reviews techniques of database merging applied to remote sensing and auxiliary surface data to enhance the information available. In this regard besides the well established polygon mapping, it also covers grid approaches as a promising alternative. Moreover, the deliverable investigates how the various actors involved in national/sub-national level land monitoring can move toward synchronisation among each other and with data requirements of the European actors, especially the European Environment Agency. Version 2 of this deliverable will also contain a third chapter addressing Nomenclature/Data models.

Authors:

| Chapter 1: Geir-Harald Strand, Norwegian Forest and Landscape Institute |
| Chapter 2: Gergey Maucha, FÖMI, Institute of Geodesy, Cartography and Remote Sensing, Hungary |

Contributors:

This document contains contributions by the entire HELM consortium.
Database merging and grid approaches
Executive summary

The European land monitoring community is currently using ancillary databases mainly for support to image interpretation and for validation of land cover maps. The method is in both cases mostly visual inspection. Database merging is extensively used only in a few countries and regions, but not widely adopted in European land monitoring. The trend at the national level is, however, a steady progress towards better access to ancillary data and - as an effect - also towards increased use of database merging in land monitoring.

The ancillary data so far used by the land monitoring community is primarily topographic maps and orthophoto. In some countries and some specific sub-national regions the institution responsible for land monitoring has access to a broader range of ancillary data, including Land Parcel Information System (LPIS) data, building registers, cadastral data, national forest inventories and a number of more specialized data sets. The main obstacle for the remaining countries is access restrictions, partly linked to the cost of obtaining access. Other issues may be the fitness-for-use of the data, data quality or temporal mismatch between the cycles of the land monitoring program and update cycle of the ancillary data; but these concerns are mostly mentioned by countries with limited or no access to ancillary data.

Most countries seem to fall into one of two distinct categories: (1) Full access or (2) no access to relevant ancillary data. Full access seems to be the rule when a (national or regional) spatial data infrastructure has been implemented. The access to ancillary data is therefore expected to improve as more countries and regions develop their spatial data infrastructures.

Examples of database merging from five European countries revealed a common methodological approach structured into two phases: Integration of ancillary data into a common data structure (or data base), and aggregation of land monitoring products, e.g. CORINE Land Cover (CLC). The fundamental model is iterative, with a possibility to reintegrate derived products into the data structure and aggregate data by stepwise procedures.

Figure 1: The iterative data integration and generalization process identified as the common structure used in European bottom-up approaches to land monitoring (simplified from Figure 4).
The integration phase of the database merging can be described or understood as “populating” a set of predefined spatial units with information retrieved from ancillary data bases. The spatial units used in all the reported examples were 25 meter pixels.

The database merging process is thus closely linked with the “grid approach” where a pre-defined set of regular grid cells are “populated” with data from ancillary data sources. The grid approach differs from the more familiar raster data structure in the sense that each grid cell is considered to represent a polygon, rather than a point. Grids are increasingly important for harmonizing spatial data among statistical agencies, and also in the landscape monitoring community. Grids have a potential to become an important tool for harmonized land monitoring by providing a platform where data from bottom-up and top-down approaches can be combined.

A few European countries are already exploring the grid approach to land monitoring at the national level. The benefit is better access to and integration with data from other sectors (demographic, socio-economic and environmental information) providing a firm basis for development and exploration of explanatory analysis and predictive modeling.
Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAD</td>
<td>Computer Aided Design</td>
</tr>
<tr>
<td>CLC</td>
<td>CORINE Land Cover</td>
</tr>
<tr>
<td>CLC2000</td>
<td>CORINE Land Cover for the year 2000</td>
</tr>
<tr>
<td>CLC2006</td>
<td>CORINE Land Cover for the year 2006</td>
</tr>
<tr>
<td>CORINE</td>
<td>Coordination of Information on the Environment</td>
</tr>
<tr>
<td>DB</td>
<td>Database</td>
</tr>
<tr>
<td>DTM</td>
<td>Digital terrain models</td>
</tr>
<tr>
<td>EEA</td>
<td>European Environmental Agency</td>
</tr>
<tr>
<td>EFGS</td>
<td>European Forum for GeoStatistics</td>
</tr>
<tr>
<td>EIONET</td>
<td>European Environment Information and Observation Network</td>
</tr>
<tr>
<td>ELCAI</td>
<td>European Landscape Character Assessment Initiative</td>
</tr>
<tr>
<td>ESPON</td>
<td>European Spatial Planning Observation Network</td>
</tr>
<tr>
<td>ESS</td>
<td>European Statistical System</td>
</tr>
<tr>
<td>FAO</td>
<td>Food and Agriculture Organization of the United Nations</td>
</tr>
<tr>
<td>GMES</td>
<td>Global Monitoring for Environment and Security</td>
</tr>
<tr>
<td>HELM</td>
<td>Harmonization of European Land Monitoring</td>
</tr>
<tr>
<td>INSPIRE</td>
<td>Infrastructure for Spatial Information in the European Community</td>
</tr>
<tr>
<td>LAEA</td>
<td>Lambert Azimuthal Equal-Area Projection</td>
</tr>
<tr>
<td>LCSS</td>
<td>Land Cover Classification System</td>
</tr>
<tr>
<td>LCML</td>
<td>Land Cover Meta Language</td>
</tr>
<tr>
<td>LISA</td>
<td>Land Information System Austria</td>
</tr>
<tr>
<td>LPIS</td>
<td>Land Parcel Information System</td>
</tr>
<tr>
<td>LUCAS</td>
<td>Land Use/Cover Area frame Survey</td>
</tr>
<tr>
<td>LULC</td>
<td>Land Use and Land Cover</td>
</tr>
<tr>
<td>LULUCF</td>
<td>Land Use Land Use Change and Forestry</td>
</tr>
<tr>
<td>METLA</td>
<td>The Finnish Forest Research Institute</td>
</tr>
<tr>
<td>NFI</td>
<td>National Forest Inventory</td>
</tr>
<tr>
<td>NFLI</td>
<td>Norwegian Forest and Landscape Institute</td>
</tr>
<tr>
<td>NSI</td>
<td>National Statistical Institute</td>
</tr>
<tr>
<td>OODM</td>
<td>Object Oriented Data Model</td>
</tr>
<tr>
<td>PLCC</td>
<td>Pure Land Cover Classes</td>
</tr>
<tr>
<td>SDI</td>
<td>Spatial Data Infrastructure</td>
</tr>
<tr>
<td>SEEA</td>
<td>System of Environmental-Economic Accounting</td>
</tr>
<tr>
<td>SYKE</td>
<td>The Finnish Environment Institute</td>
</tr>
<tr>
<td>UNSTATS</td>
<td>United Nations Statistics Division</td>
</tr>
<tr>
<td>YKR</td>
<td>The Finnish urban structure monitoring system</td>
</tr>
</tbody>
</table>
Specific objectives

The specific objectives of this report are described under Task 4.1 in the FP7 HELM project document. Task 4.1 of the FP7 HELM Project will examine

- the use of ancillary data in land monitoring.
- how database merging can foster the alignment of national land monitoring activities.
- the suitability of grid approaches based on data obtained through database merging.

Methods

Review of past questionnaires

The results from the Joint Eurostat/EEA questionnaire on Land Cover/Use sources was reviewed, emphasizing databases reported as sources for land cover information.

The Catacause method

The Catacause method was used to address a set of seed questions (Table 1) during the FP7 HELM project meeting in Malaga in March 2012. The seed questions were divided into three sub-groups, each assigned to a moderator (“Knowledge Carrier”) located in a “Niche” (corner or place in a large room or small extra room). Each participant joined one of the Niches in order to discuss the seed questions with the Knowledge Carrier. The Knowledge Carrier did to some extent act as an instructor but mostly a moderator. The participants could change Niches individually whenever they wanted. This ongoing moving around brought about some repetition in the discussions, indicating particular interest or need for information regarding a certain matter. The Knowledge Carriers observed such patterns, but also took note of surprises and issues which remain open.

The Knowledge Carrier assigned to each subtopic was assisted by a Rapporteur taking notes for the final documentation of the discussions.

Table 1: Seed questions used in the Catacause during the HELM meeting in Malaga, March 2012

<table>
<thead>
<tr>
<th></th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Which ancillary databases do you consider relevant for land monitoring?</td>
</tr>
<tr>
<td>2</td>
<td>How and to what extent have you used ancillary databases in the production or validation of CLC?</td>
</tr>
<tr>
<td>3</td>
<td>Can you provide (and assess) examples where ancillary databases have been used in bottom-up processes in land monitoring?</td>
</tr>
<tr>
<td>4</td>
<td>Can you provide (and assess) examples where ancillary databases have been used to parameterize land cover or land use databases?</td>
</tr>
<tr>
<td>5</td>
<td>What are in your opinion the important preconditions/obstacles for the use of ancillary databases in land monitoring?</td>
</tr>
<tr>
<td>6</td>
<td>Do you have any experience with using grids for monitoring? Give examples.</td>
</tr>
<tr>
<td>7</td>
<td>What do you consider advantages/disadvantages of using a grid approach in land monitoring?</td>
</tr>
<tr>
<td>8</td>
<td>What are the problems and prerequisites for harmonizing grids across national borders?</td>
</tr>
</tbody>
</table>
Comparison of selected national databases

The FP7 HELM participants from countries where national databases were available, or at least considered, as ancillary data were asked to provide information on the content and availability of such data. Three themes were selected (buildings, roads and agricultural land). A concise questionnaire was used to ensure comparable information. Participants from seven countries (Austria, Finland, Germany, Netherlands, Norway, Switzerland and United Kingdom) provided input for this exercise. Participants from countries with more than one organization involved in FP7 HELM coordinated the exercise among themselves.

Collection of examples of database merging

The FP7 HELM participants known to have experience with database merging in the production of national or pan-European land monitoring information were asked to describe their methodology and procedures following a simple but systematic form. The results were compared in order to identify joint experiences. Participants from six countries (Austria, Finland, Netherlands, Norway, Switzerland and United Kingdom) provided input for this exercise. The descriptions were mainly, but not exclusively related to bottom-up production of CLC.

Evaluating the suitability of grid approaches

Information about the use of grids was obtained from national and international systems having experience from the use of grid approaches. Information about FP7 HELM has been provided to the European Forum for GeoStatistics (EFGS). EFGS is a voluntary cooperation between 32 National Statistical Institutes (NSI) working closely together with Eurostat in order to develop and harmonize the use of grids in their statistical work. EFGS also provided background documents and reports documenting their efforts to harmonize their spatial statistical systems using grids. Similar information was obtained from the European Spatial Planning Observation Network (ESPON) where the use of grids as statistical units has been considered for some time.
Ancillary data and database merging – an overview

Ancillary databases are (in this context) databases with a content that provides relevant land cover or land use information for a monitoring system, without having this use as the primary intention. Examples are topographic maps and orthophoto obtained for various other purposes but used – through visual inspection - as supporting information in a land monitoring system (e.g. CLC). Other examples of ancillary data used in land monitoring are building registers, road maps, cadasters, land parcel information systems and hydrological data. A number of more specialized data sets are also used when available. A special case of ancillary data usage occurs when land cover or land use data are compiled for use at one scale and later applied in a monitoring system working at a different - usually smaller - scale.

**Ancillary:** subsidiary; providing something additional to a main part or function

The use of ancillary data in land monitoring can be divided into two categories:

1. **Visual inspection** of ancillary data in order to assist an analyst during image interpretation or validation.
2. **Database merging**, i.e. computer-assisted compilation of land monitoring data from existing databases.

Ancillary data is one of the three data sources for land monitoring (Figure 2). The other two sources are field observations and image interpretation. Ancillary data can be used together with image interpretation (and also during field observation). The direct use of ancillary data is the practice labeled as database merging (although, in its simplest form, it can be reduced to generalization).
Overview of relevant ancillary data types

Topographic maps (or digital topographic data) and orthophoto are the data considered by the FP7 HELM participants to be most relevant for land monitoring (CLC production in particular). These two data types are widely available and reported by FP7 HELM participants as useful in database merging as well as for visual inspection during image interpretation.

Land parcel data (Land Parcel Information System (LPIS) in the EU and similar data outside the EU) was especially mentioned by FP7 HELM participants as a dataset needed by the land monitoring community. Some countries have access to these data, but access is more often than not restricted due to the content of personal data about the owner and/or user of the parcel. Some countries where land parcel data are accessible to the land monitoring community (e.g. Netherlands and Norway) have solved this issue by removing the sensitive information from the data before they are made available for land monitoring. The land monitoring service in Finland, on the other hand, has full access to LPIS data and is itself responsible for appropriate handling of sensitive data.

LUCAS (Land Use/Cover Area frame Survey) sample data and LUCAS photo are used as ancillary data for validation of CLC in Italy and Germany. Norway has implemented a national adaption of LUCAS as the methodology for land cover monitoring, and used this as ancillary data supporting image interpretation in the compilation of the land cover map of mountain areas – which again was the input in the production of CLC2000.

Ancillary data on buildings are reported by several countries as useful in order to identify built-up land and determine the land use. This information can in some cases be retrieved from digital topographic maps (e.g. the Dutch Top10NL) where buildings are represented as objects. Other countries have specific registers of buildings – sometimes included in the cadaster - with sufficient spatial reference allowing the data to be used in land monitoring systems.

Digital terrain models (DTM) are available for use in several countries, e.g. Finland, Switzerland, Czech Republic and Norway. The main use of DTM in land monitoring is linked to automatic and semi-automatic image interpretation.

Hydrological data are used to delineate water bodies and watercourses in the land monitoring in Germany, Switzerland and several Nordic countries.

Road data are used to determine the land use of built up land in Germany, Switzerland and several Nordic countries.

Other thematic maps occasionally used as ancillary data in land monitoring are

- Soil maps - used in Finland to distinguish between peat land and mineral soils.
- Storm damage maps – used in Switzerland
- Glacier maps – used by Iceland and Switzerland to delineate glaciers in CLC
- Forest maps and National Forest Inventories (NFI) – used in Finland, Iceland, Switzerland and Norway
Several countries have large-scale databases containing a mixture of topographic and infrastructural data. These databases are usually maintained by local authorities but (more or less) within standards set by a supervising national body, often the national mapping authority. These databases are also available as national (copy) datasets compiled from the local data by the national authority in charge. Examples are found (among others) in Germany (ATKIS), Iceland (IS50V), Switzerland (TLM), Finland (Topographic database) and Norway (FKB). These large-scale databases can include detailed road data, hydrology, cadastral information and building registers and are valuable sources for land monitoring when available to the land monitoring community.

The results from the Joint Eurostat/EEA questionnaire on land use/cover 2010 contain information about available land use/cover and ancillary data provided by 19 European countries. The results from this survey closely resemble the information provided by the FP7 HELM participants, described above. Detailed maps with land cover and land use information as well as various registers linking land use to points or parcels are available and considered relevant for land monitoring in many European countries.

The format and detailed content of the national datasets are rarely standardized at the pan-European level. The data are, however, so detailed that harmonization most likely can be achieved when the data are aggregated to broader categories for use in pan-European land monitoring. None of the countries where bottom-up procedures have been implemented reported any major problem linked to this issue.

Ancillary data – access and availability

The main advantages of using ancillary data are that they: 1) represent existing knowledge; and 2) are updated on a regular basis, usually as part of the administrative routines in a responsible authority. One such example is building registers. Building registers are usually maintained on a regular basis by the local authorities, as part of the legal process linked to granting permits for construction, demolishing or changing the use of buildings. When this process is operational and the register is available as a national database, it provides excellent and updated information about land use as well as land cover.

Table 2: Issues identified as obstacles to wider use of ancillary data.

<table>
<thead>
<tr>
<th>Category</th>
<th>Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal</td>
<td>Access and licensing issues</td>
</tr>
<tr>
<td></td>
<td>Sensitive (restricted) information (eg LPIS)</td>
</tr>
<tr>
<td>Administrative</td>
<td>Harmonization / coordination within the country</td>
</tr>
<tr>
<td></td>
<td>Temporal mismatch (updating cycles)</td>
</tr>
<tr>
<td></td>
<td>The “Federal issue” (Budgets, ownership, coordination)</td>
</tr>
<tr>
<td>Technical</td>
<td>Formats (especially the use of CAD formats by some institutions)</td>
</tr>
<tr>
<td></td>
<td>Spatial resolution (incompatible details)</td>
</tr>
<tr>
<td></td>
<td>Missing metadata</td>
</tr>
<tr>
<td></td>
<td>Capacity (having sufficient computing power)</td>
</tr>
<tr>
<td></td>
<td>Insufficient quality (not updated)</td>
</tr>
</tbody>
</table>
The main obstacle for use of ancillary data is data access. The datasets in question exist in most of the countries, but the use is restricted and the usefulness is limited. The obstacles to wider use of ancillary data are listed in Table 2.

**Ancillary data as “copy data bases”**

In some countries good ancillary data exists and are available to the land monitoring community, but access is not granted directly to the database itself, but rather to a copy derived from the main database at certain intervals. There are both advantages and disadvantages to this approach.

Major advantages of using copy databases are:

- The secondary usage does not interfere with the primary use of the database.
- The copy database can be somewhat simplified removing specialist information of little or no use to the land monitoring community and other secondary users.
- Sensitive information about individual persons or businesses can be removed and the database presented in an anonymized form.
- Additional quality control can be carried out during the production of the copy.
- Added standardization can be achieved between the copy data bases.

The disadvantage of the copy database approach is the introduction of a lag time between the maintenance of the primary database and the availability of the maintained data in the copy database. This is rarely a problem when the copy database is derived several times each year, but reduces the temporal quality of the data seriously when the copy database is updated at longer intervals than a year. A second disadvantage is that the copy procedure represents another processing step where errors may be introduced.

**Study of selected examples of ancillary data**

Three sets of thematic data (*buildings*, *roads* and *agricultural areas*) were selected for more detailed studies. Structured information about the availability of these datasets at the national level was provided by FP7 HELM participants from seven countries (Austria, Finland, Germany, Netherlands, Norway, Switzerland and United Kingdom). The compiled information is found in Attachment 1, where the replies from the countries also can be compared easily.

Detailed data on *buildings* in vector format and linked to detailed attribute lists are available in all seven countries. The information is free (within the national Spatial Data Infrastructure; SDI) in four of the countries, while access requires substantial payments in two countries. There is also a licensing system for these data in Germany as well as in Finland. The buildings are encoded with building type according to national code lists, but general land use information categories can be inferred from these code lists. The detailed versions of these datasets are maintained through administrative updating, but the accessible version can be updated at less frequent intervals (up to two years is reported).

Detailed data on *roads* are available in vector format in all seven countries. The information is free (within the national SDI or through a general open access policy) in five of the coun-
tries. There is no access to the data in one country, and a licensing system in another country. Data representation is mostly by centerline. Roads are encoded according to national code lists, but sufficiently detailed to allow use in harmonized land cover/land use data products. The databases are maintained through administrative routines, but the public version may be a derived product only maintained at intervals.

Detailed data on agricultural areas is available in polygonized vector format in five countries. Switzerland has a system based on intensive point sampling with 100 meter intervals. The information is free in four of the countries, restricted in two countries, and subject to licensing in the last country. The polygon data are detailed (large scale and detailed encoding) and maintained through administrative routines, often linked to the agricultural subsidy system. The point data available in Switzerland are also detailed, but are obtained through a sampling program run by the Swiss NSI.

The different cycles at which the data are collected in the countries presents a particular challenge for land monitoring at the pan-European level because this monitoring is based on the idea of presenting a “snap-shot” of the situation across the continent.

Conclusions

The general impression from the study is that ancillary data, when available, are sufficiently detailed to provide a sound basis for database merging as a tool to provide harmonized data at the pan-European level, e.g. within the framework of CLC. The reason is that the categories used at the pan-European level are fairly broad. CLC uses concepts like “urban fabric” and “industrial or commercial units”. The more detailed national or regional code lists may not be individually comparable, but can usually be consistently aggregated to the wider pan-European categories. The imprecise definition of some of the pan-European categories is in this respect a bigger challenge than the differences between the national code lists.

Data access is limited in many countries. Most countries seem to fall into two distinct categories: (1) Full access or (2) no access to relevant ancillary data. Full access seems to be the rule when a (national or regional) spatial data infrastructure (SDI) has been implemented, but open access can also be practiced without the formalization of an SDI, as in Finland.

The access to ancillary data is likely to improve as more countries and regions develop their spatial data infrastructures. The introduction of the INSPIRE directive can be expected to accelerate this process and provide better access to ancillary data for the public institutions in charge of land monitoring. It should, however, be noted that “free” in some countries (e.g. Norway) means access where payment for the individual datasets is replaced by a general payment for participation in the national SDI. The cost is still there, but the benefit of the SDI is that access is simplified because it is not necessary to negotiate and buy licenses for the individual datasets. The SDI can also stimulate the owners of previously restricted databases to produce simplified and anonymized copy databases for use in the SDI.

The maintenance of ancillary data is often carried out as part of the administrative routines in a public authority, but access is granted by providing copy data sets produced at certain intervals often involving a delay. Users of statistical surveys are faced with similar challenges
when the survey data are obtained over several years, as in the Swiss Land Use/Land Cover (LULC) survey and in the national forest inventories found in many countries (Tomppo et al. 2008).

Pan-European land monitoring practiced today is supposed to provide a snap-shot of the situation at a certain moment in time. Satellite imagery is therefore the preferred data source. Imagery is, however, also usually a mosaic of images obtained at different dates. Realistically, at least for countries with challenging weather conditions, these images may be obtained over several years. Replacing image interpretation by ancillary data is therefore not necessarily worsening the temporal accuracy in land monitoring, but the temporal aspect will change and has to be interpreted and explained differently.
Database merging

*Database merging* in a land-monitoring context is the computer-assisted compilation of land monitoring data from existing databases. FP7 HELM participants performing any kind of database merging as part of their land monitoring activities were asked to provide a description of their approaches. Descriptions were provided by FP7 HELM participants from Austria, Finland, Netherlands, Norway, Switzerland and United Kingdom. These six descriptions are found in Attachment 2 of this report.

There are considerable commonalities between the approaches used by the five countries. The basic structure is a process where areas expected to belong to different classes in an output data set (e.g. CLC) is retrieved individually from a collection of input data sets. Each class is geometrically generalized and the results merged together following a prioritized sequence. This basic structure, illustrated in Figure 3a (left), is the approach used in Switzerland and the United Kingdom. Austria also follows this structure, but is using a single high resolution database (LISA) as input data.

![Diagram of database merging](image)

Figure 3: (a) The basic structure of database merging (left) and (b) with an initial data integration phase (right)

Finland, Netherlands and Norway also follow the basic structure illustrated in Figure 3a (left), but have added a “data integration phase” at the outset of the process as shown in Figure 3b (right). First, the ancillary data bases are merged together into a single, integrated dataset with common spatial units. Incidentally, all three countries use 25 meter pixels as the common geometrical units. The data integration is followed by procedures implementing the same basic sequence of “retrieval – generalization – merging” operations as in Figure 3a.
What is revealed here is a common database merging process with two phases: (1) An initial “data integration” phase and (2) a subsequent “data aggregation” phase. We can also notice that the two phases can be used independently of each other. The grid approach, described in a separate chapter below, is a stand-alone application of phase 1. Simple generalization of a single input map is actually a stand-alone application of phase 2.

Figure 4: Database merging described as an iterative approach. Phase 1 is the “integration” of input data into a common data structure (or database) while Phase 2 is the “aggregation” through retrieval, generalization and merging of land monitoring classes.

A closer study of the national examples furthermore shows that the approach can be implemented as an iterative process. This is illustrated in Figure 4. An example of such an iterative process is found in Netherlands where the structure is repeated twice: once an integration of databases with the topographic polygons as geometrical units and a second integration phase using 25 m grid cells as units. Data integration is followed by a generalisation/merging process in both iterations. Another example is the methodology developed for CLC2000 in Norway. Initially, various ancillary input data were integrated (using 25 m grid cells as the common spatial units) and stored in an integrated database. CLC Level 1 classes were retrieved from this database, generalized and merged together. The resulting CLC Level 1 map was returned to the integrated database and used as a framework when CLC Level 3 classes were retrieved, generalized and later merged in order to form the CLC2000 map.

Based on the experience from five countries, Figure 4 is found to give a good general description of the process involved in database merging. The structures shown in Figure 3a and 3b are just simplifications of this general process. The commonality between the approaches
used in the five countries supports the idea that database merging can be used to foster the alignment of land monitoring practices between the countries.

**Classification and characterization**

There are, fundamentally, two different approaches to land cover and land use mapping: *Classification* and *characterization*. Classification is the approach where spatial units are assigned to categories in a pre-defined classification system. The method is well-known from Corine Land Cover and many national land monitoring systems. Characterization, on the other hand, is the approach where the spatial units are described using a set of independent diagnostic criteria (Di Gregorio and Jansen 2000, Gomarasca 2009). Variants of this method are the Object Oriented Data Model used in Spain, the normalized nomenclature used in Norway, and the FAO LCCS used (among others) in Albania (Nikolli 2010).

Database merging, as described here, involves both approaches. The “integration” phase of the database merging process is an example of *characterization*, when the predefined spatial units are “populated” with attributes drawn from the various input data sets. The “aggregation” phase of the database merging process, on the other hand, is *classification*.

**Attribution of land cover/land use units**

*Attribution* of land cover/land use units is a special case of the general database merging, limited to Phase 1 (“integration”). The predefined spatial units are in this case the existing geometry of a land cover or land use map and ancillary data are added as attributes of these spatial units. A simple example is to link a land cover map to a map of buildings, count the number of buildings in each land cover polygon and add the number of buildings as an attribute to the polygons.

*Attribution* is sometimes used in the iterative aggregation process described above. In the Norwegian CLC2000, data integration and generalization (as described above) was used to create pure agricultural classes (2.1.1 Non-irrigated arable land and 2.3.1 Pastures). These areas were re-integrated into the database in order to obtain more detailed land cover data describing the composition of land cover inside each polygon. The information was used to identify mosaic polygons (2.4.2 Complex cultivation patterns and 2.4.3 Land principally occupied by agriculture).

*Attribution* can also be used to check the quality of land monitoring data. The polygons of the Norwegian CLC2006 map were attributed with data from the detailed Norwegian land resource maps. The resulting attribute data set was treated as a statistical table and aggregated in order to provide a profile of land cover composition within each CLC class. The results showed that the CLC product matched with the definition of the CLC classes (Aune-Lundberg and Strand 2010).

*Attribution* is also expected to be one of the usages of the upcoming GMES High Resolution Layers produced as part of GMES Initial Operations Land. High Resolution Layers linked to CLC maps will provide information about the content of the CLC polygons which can be summarized and stored as attributes of the polygons.
The grid approach

A grid is a spatial model falling somewhere between the vector model and the raster model. The grid is a spatial partition of the Earth’s surface into non-overlapping and non-empty parts. Similar to the raster model, the grid consists of spatial units (“cells”) of uniform size and shape. The grid cells can, however, be thought of as representing “polygons” rather than as “pixels”.

The grid is not “mapped” in any traditional sense, but instead “populated” with information from various sources. This information could include but is not restricted to information about land cover and land use. The process can be direct, when the grid cell is small and information is obtained by remote sensing, but can also involve counting or statistical processing of register data for various kinds of observations made inside each grid cell. One method is to calculate spatial coverage based on a geometric overlay between the grid and various polygon datasets (e.g. CLC). Ancillary databases provide important and relevant information about the content of the grid cells. Examples are number and composition of buildings in the grid cell, or the length of roads within the grid cells.

The grid approach is closely linked to database merging, and can be interpreted as another example of phase 1 in the database merging model presented above (see Figure 4).

The grid approach in European land monitoring

EEA periodically converts the CLC dataset into a grid in order to carry out analysis.

Iceland uses grids for various monitoring activities like forest (woodland and plantations), soil, plants, birds and LULUCF but there is not one common grid and the different initiatives usually cover only part of the country.

The Netherlands have a gridded land cover data base. The origin was pixels, due to the use of satellite imagery, but the grid cells are now fixed and often populated with data from sources other than remote sensing.

Norway has developed a standardized national grid (Strand and Bloch 2009) with flexible resolution (the owner is Statistics Norway). The Norwegian Forest and Landscape Institute (NLFI) provides land cover and land resource data for a 5 km and a 1 km database based on this common grid. Statistics Norway and other public authorities also make data available using the same grids.

Austria: Have started research on grid mapping with various types of data, grid cells 200 – 2000m.

Finland (Finnish Environment Institute) produces its national Corine Land Cover as 25 m raster classification. There are also other information systems like the urban structure monitoring system (YKR) which utilizes a grid approach, at least partially. YKR is an information system for land use experts, making it possible to monitor long-term changes of urban areas as well as to carry out analysis and research on urban areas. YKR contains information about
people, housing including summer cottages, employment and land cover and land use in a 250 by 250 meter grid. The Finnish Forest Research Institute (METLA) provides information concerning forests in a grid format. The National Land Survey has produced SLICES land use data using 25 and 10 m grid size for the years 1999, 2005 and 2010.

Switzerland uses a 25 m grid (due to satellite images) of forest mixture classes. Starting in 2012 there will also be a 5 m grid representing tree type (coniferous/deciduous) with updating foreseen every three years, according to the availability of new aerial images. The national area statistics uses a 100 m grid measured at the center point, with 72 LULC classes. Population (census) is also available on a 100 m grid if more than a minimum number of persons are living there.

EFGS (European Forum for GeoStatistics) is a voluntary cooperation between several NSIs across Europe (and now spreading to South-America) that have started using grids for integration and harmonization of spatial data. Eurostat is informally involved in this cooperation. The steering committee consists of the national contact points from the NSIs that participate in the European Statistical System (ESS) project GEOSTAT 1A: Representing Census data in a European population grid mainly funded by Eurostat. ESS can loosely be interpreted as an equivalent to EIONET and GEOSTAT 1A as an equivalent to CLC.

The System of Environmental-Economic Accounting is an internationally agreed standard for harmonized statistics on the environment and its relationship with the economy, organized by the United Nations Statistics Division (UNStats). With respect to environmental-economic accounting, the grid cell turns out to be a fairly common statistical unit for spatially referenced environmental data, including land cover and land use, used in the national accounting systems Vardon et al. (2011).

The European Spatial Planning Observation Network (ESPON) is an EU program aiming to support the “reinforcement of regional policy with studies, data and observation of development trends” European Commission (2007). Potential and actual use of grids in a spatial planning context is described in an ESPON report addressing the modifiable areal unit problem (Anonymous, no date) and gridded data structures are given renewed attention in recent results from projects organized under ESPON (e.g. Milego and Ramos, 2011).

There is furthermore a movement towards increased use of grids in the landscape research community, linked to the concepts landscape character assessment and the development of a European landscape classification system. The European Landscape Character Assessment Initiative (ELCAI) running in the period 2003 – 2005 documented a widespread use of grids in national landscape descriptions throughout Europe (Wascher 2005) and led to the development of a proposed pan-European landscape description based on a grid approach (Mücher et al. 2010).
Advantages and disadvantages with grids

Grids have the advantage that several parameters/attributes from various sources can be combined in statistical, spatial analyses, environmental models and surveys, i.e. land cover, landscape and land use parameters can be combined with economic, social and even epidemiological information to reveal interesting relationships or associations.

Grids can be perceived as a kind of object oriented data model (OODM). The grid cell is an (abstract spatial) object and various attributes (e.g. houses: number of houses, use, building material, year of construction etc.) can be attached to the object. There are vast possibilities for integrating many different aspects of the land that open up by following such an approach.

The disadvantages include the fact that a grid cell is an artificial spatial unit and it can be difficult to interpret the information that is presented. There is also concern about the “impurity” of grid cells since they (especially large grid cells) often will contain a mixture of different types of land cover.

Another drawback of the grid is the fact that information on gross changes is lost (Figure 5). Net change is the calculated difference in area between two specific points in time while gross change is the total area modified during this period. One problem with the grid approach is that we only get to know the net changes, while these may actually hide more complex (and interesting) patterns of gross changes.

<table>
<thead>
<tr>
<th>Gross changes vs. net changes:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Built-up</td>
</tr>
<tr>
<td>Built-up</td>
</tr>
<tr>
<td>Built-up</td>
</tr>
</tbody>
</table>

Figure 5: Imagine a grid cell with 25 hectare Built-up land, 25 hectare Agriculture and 50 hectare Forest (left). After some years the content has changed to 37,5 hectare Built-up land, 25 hectare agriculture and 37,5 hectare forest. The net change is that the coverage of Built-up land has increased 12,5 hectare and the coverage of Forest has decreased 12,5 hectare. The gross change is not known at the grid cell level. One scenario is that 12,5 hectare Forest land was developed into Built-up land (center). Another possibility is that 12,5 hectare Agricultural land was developed into Built-up land, while 12,5 hectare Forest land was cleared and developed as new Agricultural land (right).
Harmonizing grids across national borders

**Issue 1: A common grid**
Regarding harmonization at the European level, it seems that the location of individual European countries is a challenge with respect to establishing usable grid databases. EEA as well as INSPIRE recommends the use of the multipurpose ETRS89 Lambert Azimuthal Equal Area 52N 10E grid. This grid is described in detail in INSPIRE (2010). EFGS is also using this grid for pan-European data.

Even if there is a common European grid, it will probably not be used at the national level where it is more convenient and user friendly to use a grid based on the standard national projections. Production of gridded data for a common European grid therefore have to follow some agreed and standardized methodology. This could be to populate the common grid (for sharing data) independently of populating the national grid – i.e. the national ancillary data constitute the basis and independent gridding procedures are used in order to populate the national and the European grid. Another method could be to convert the data from the national to the European grid using an agreed convolution method.

**Issue 2: Populating the grid**
There are many ways to populate a grid (Strand and Bloch 2009). The land cover class assigned to a grid cell could be the land cover observed at the center of the grid cell or at a random point within the grid cell; or it could be the dominant land cover inside grid cell. It is also possible to measure each land cover class and either assign a presence/absence indicator to the grid cell or attach a value representing the coverage (e.g. percentage) for each land cover class found in the grid cell. With respect to cross-border harmonization, it is imperative to agree on a common methodology for populating the grids.

**Issue 3: Borderline grid cells**

Pan-European grids will contain grid cells intersecting two or more countries. There is a need for a methodology and administrative routines for populating these grid cells.

**Issue 4: Definitions**

The issues concerning classification systems and class definitions also apply when grids are used as the spatial representation of land monitoring data.
References and documentation

This section includes both literature referenced in the text and more detailed documentation of the national database merging approaches found in Attachment 2.


Heggem, ESF. (2011) CORINE Land Cover: An automatic generated land cover map, Kart og Plan, 77: 274 – 280 (in Norwegian with summary in English)


Attachment 1: Ancillary data at the national level

Availability of selected ancillary data at the national level

A structured form was sent to the FP7 HELM participants, requesting information about three selected ancillary datasets (buildings, roads and agricultural areas). Replies were provided from Austria (AT), Switzerland (CH), Germany (DE), Finland (FI), Netherlands (NL), Norway (NO) and United Kingdom (UK).

<table>
<thead>
<tr>
<th>Buildings</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>An almost complete (&gt;95%) building register is managed by communities under the guidance of the Statistical Office (XY-coordinates) and most buildings are included in the national cadaster (polygons), available through the National Mapping Authority</td>
</tr>
<tr>
<td>CH</td>
<td>A complete building register is included in the national cadaster, available through the National Mapping Authority</td>
</tr>
<tr>
<td>DE</td>
<td>A complete building register is included in the national cadastre, available through a central distribution authority of the Cadastral and Surveying Authorities of the Länder (German Federal States).</td>
</tr>
<tr>
<td>FI</td>
<td>A complete building and dwelling register is available from the national Population Register Center (VRK)</td>
</tr>
<tr>
<td>NL</td>
<td>A complete building register is included in the national topographic map (Top10NL)/Key Registration Topography Basic Registration Addresses and Building (BAG), source municipalities, available through Cadaster</td>
</tr>
<tr>
<td>NO</td>
<td>A complete building register is included in the national cadaster, available through the National Mapping Authority</td>
</tr>
<tr>
<td>UK</td>
<td>A building layer is part of the OS OpenData VectorMap District product</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Access</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>licensed access to polygons (quite expensive) and licensed access to grid-data (100m, 125m, 250m, ...): e.g. 13.000€ for 100 m raster of count of buildings for total Austria</td>
</tr>
<tr>
<td>CH</td>
<td>Free access as part of the national SDI (liable to pay costs)</td>
</tr>
<tr>
<td>DE</td>
<td>License (cost)</td>
</tr>
<tr>
<td>FI</td>
<td>Accessible, but accessibility requires the payment of a substantial yearly fee</td>
</tr>
<tr>
<td>NL</td>
<td>Free access (for non-commercial users) as part of the national SDI</td>
</tr>
<tr>
<td>NO</td>
<td>Free access as part of the national SDI</td>
</tr>
<tr>
<td>UK</td>
<td>Free access, download as 100km bundles</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geometry</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Vector: Buildings are represented as vector polygons, Grid: Buildings are represented as points</td>
</tr>
<tr>
<td>CH</td>
<td>Vector (3D)</td>
</tr>
<tr>
<td>DE</td>
<td>Vector: Buildings are represented by ground plots (polygons).</td>
</tr>
<tr>
<td>FI</td>
<td>Vector: Buildings are represented as points. Some parts of the data are sensitive (information on individuals) and are not accessible for all users.</td>
</tr>
<tr>
<td>NL</td>
<td>Vector: Buildings are represented as polygons</td>
</tr>
<tr>
<td>NO</td>
<td>Vector: Buildings are represented as points</td>
</tr>
<tr>
<td>UK</td>
<td>Vector: Buildings &gt; 175m² are represented as areas</td>
</tr>
<tr>
<td>Attributes</td>
<td>AT</td>
</tr>
<tr>
<td>------------</td>
<td>----</td>
</tr>
<tr>
<td></td>
<td>Polygons: no attribute</td>
</tr>
<tr>
<td></td>
<td>Grid-data:</td>
</tr>
<tr>
<td></td>
<td>100m: number of appartements, number of houses</td>
</tr>
<tr>
<td></td>
<td>250m: Building type (codelist – 11 categories, implying land use)</td>
</tr>
<tr>
<td></td>
<td>Appartement type: 4 categories</td>
</tr>
<tr>
<td></td>
<td>Number of rooms: 1, 2, &gt;=3</td>
</tr>
<tr>
<td></td>
<td>Size &amp; primary/secondary home: 4 categories</td>
</tr>
<tr>
<td></td>
<td>Base area (sq.m)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maintenance</th>
<th>AT</th>
<th>CH</th>
<th>DE</th>
<th>FI</th>
<th>NL</th>
<th>NO</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Continuously updated by local administration as part of their management of building applications (building permit, address registry)</td>
<td>Continuously updated by local administration as part of their management of building applications</td>
<td>Updated yearly by the Cadastral and Surveying Authorities of the Länder (German Federal States).</td>
<td>Continuously maintained by local administration (municipalities) as part of their management of building applications. The updates of the national register are made by the VRK based on the information provided by the municipalities</td>
<td>Top10NL; every 2 years BAG: Continuously updated by local administration as part of their management of building applications</td>
<td>Continuously updated by local administration as part of their management of building applications</td>
<td>Base data continuously updated by the national mapping agency, free version updated at intervals</td>
</tr>
</tbody>
</table>
# Roads

## Availability

<table>
<thead>
<tr>
<th>Country</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>A complete road database of drivable roads is currently developed as joint initiative of the regional governments that can be used for all kind of road and traffic management applications (not yet fully available)</td>
</tr>
<tr>
<td>CH</td>
<td>A complete database of roads and paths (highway → hiking path fragments) is available through the National Mapping Authority</td>
</tr>
<tr>
<td>DE</td>
<td>Roads are a part of topographic database ATKIS® of the Cadastral and Surveying Authorities of the Länder (German Federal States). Complete coverage of Germany is distributed by Bundesamt für Kartographie und Geodäsie (BKG, Federal Agency for Cartography and Geodesy).</td>
</tr>
<tr>
<td>FI</td>
<td>A complete road database called DigiRoad is available from the Finnish Transport Agency.</td>
</tr>
<tr>
<td>NL</td>
<td>A complete road register in the national topographic map (Top10NL)/Key Registration Topography National Road database(NWB) – Rijkswaterstaat – Ministry I&amp;M</td>
</tr>
<tr>
<td>NO</td>
<td>A complete road database of driveable roads longer than 50 meters is available through the National Mapping Authority</td>
</tr>
<tr>
<td>UK</td>
<td>A complete road database of drivable roads is part of the OS OpenData VectorMap District product</td>
</tr>
</tbody>
</table>

## Access

<table>
<thead>
<tr>
<th>Country</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Not yet accessible</td>
</tr>
<tr>
<td>CH</td>
<td>Free access as part of the national SDI (liable to pay costs)</td>
</tr>
<tr>
<td>DE</td>
<td>License (cost)</td>
</tr>
<tr>
<td>FI</td>
<td>Free access (only a very small extraction fee)</td>
</tr>
<tr>
<td>NL</td>
<td>Free access as part of the national SDI</td>
</tr>
<tr>
<td>NO</td>
<td>Free access as part of the national SDI</td>
</tr>
<tr>
<td>UK</td>
<td>Free access, download as 100km bundles</td>
</tr>
</tbody>
</table>

## Geometry

<table>
<thead>
<tr>
<th>Country</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Vector: Each road is represented by many geometries (center line, outer lines, etc.)</td>
</tr>
<tr>
<td>CH</td>
<td>Vector: Roads are represented by centerline</td>
</tr>
<tr>
<td>DE</td>
<td>Vector: Roads are represented by centerline</td>
</tr>
<tr>
<td>FI</td>
<td>Vector: Roads are represented by centerlines and as a network.</td>
</tr>
<tr>
<td>NL</td>
<td>Vector: Roads are represented by centerline and polygons when roads exceed 2 m width</td>
</tr>
<tr>
<td>NO</td>
<td>Vector: Roads are represented by centerline</td>
</tr>
<tr>
<td>UK</td>
<td>Vector: Roads are represented by centerline</td>
</tr>
</tbody>
</table>

## Attributes

<table>
<thead>
<tr>
<th>Country</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Road type (code list), traffic density, road width, .....</td>
</tr>
<tr>
<td>CH</td>
<td>Road type (code list)</td>
</tr>
<tr>
<td>DE</td>
<td>Several attributes, e.g. road type</td>
</tr>
<tr>
<td>FI</td>
<td>Road and railway type (code list) Road width and pavement type</td>
</tr>
<tr>
<td>NL</td>
<td>Road type (code list) with many additional attributes</td>
</tr>
<tr>
<td>NO</td>
<td>Road type (code list)</td>
</tr>
<tr>
<td>UK</td>
<td>Road type (code list), Road name (free text), Road identifier (classification)</td>
</tr>
</tbody>
</table>
### Maintenance

<table>
<thead>
<tr>
<th>Country</th>
<th>Update Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Road management authorities in regional governments will continuously update the dataset</td>
</tr>
<tr>
<td>CH</td>
<td>Every 6 yr. by the National Mapping Authority</td>
</tr>
<tr>
<td>DE</td>
<td>Continuously updated by the Cadastral and Surveying Authorities of the Länder (German Federal States)</td>
</tr>
<tr>
<td>FI</td>
<td>Continuously updated by the National Land Survey, the Finnish Transport Agency and the municipalities. They have different responsibilities in the maintenance process.</td>
</tr>
<tr>
<td>NL</td>
<td>Top10NL; every 2 years NWB: Continuously updated by the road administration</td>
</tr>
<tr>
<td>NO</td>
<td>Continuously updated by the road administration</td>
</tr>
<tr>
<td>UK</td>
<td>Base data continuously updated by the national mapping agency, free version updated at intervals</td>
</tr>
</tbody>
</table>

### Agricultural land

#### Availability

<table>
<thead>
<tr>
<th>Country</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>A complete map of agricultural land in scale of cadastral parcels (1:1:000) is used for subsidy payments in the ministry for agriculture</td>
</tr>
<tr>
<td>CH</td>
<td>A geostatistical LC &amp; LU dataset – (3 nomenclatures: LU, LC, mixed)</td>
</tr>
<tr>
<td>DE</td>
<td>Agricultural land is a part of topographic database ATKIS® of the Cadastral and Surveying Authorities of the Länder (German Federal States). Complete coverage of Germany is distributed by Bundesamt für Kartographie und Geodäsie (BKG, Federal Agency for Cartography and Geodesy).</td>
</tr>
<tr>
<td>FI</td>
<td>A database (Finnish Land Parcel Identification System, FLPIS) of agricultural land is available from the Ministry of Agriculture and Forestry. The database covers the agricultural areas that get agricultural subsidies.</td>
</tr>
<tr>
<td>NL</td>
<td>A complete coverage is included in the national topographic map (Top10NL)/ Key Registration Topography BRP/LPIS – farmers information on crops for subsidy LGN National Land use database</td>
</tr>
<tr>
<td>NO</td>
<td>A complete map (1:5,000) of agricultural land is available at the Norwegian Forest and Landscape Institute</td>
</tr>
<tr>
<td>UK</td>
<td>Rural Land Registry. 1:5000 registered land parcels in a digital format.</td>
</tr>
</tbody>
</table>

#### Access

<table>
<thead>
<tr>
<th>Country</th>
<th>Access Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>No public access until now (ongoing discussions with legal advisers how to resolve data property rights – Status 10/2012)</td>
</tr>
<tr>
<td>CH</td>
<td>Free access as part of the national SDI</td>
</tr>
<tr>
<td>DE</td>
<td>License (cost)</td>
</tr>
<tr>
<td>FI</td>
<td>Free access. An application form has to be filled in.</td>
</tr>
<tr>
<td>NL</td>
<td>Top10NL: Free access as part of the national SDI BRP/LPIS: Agricultural Ministry EL&amp;I – “free” access LGN: Commercial</td>
</tr>
<tr>
<td>NO</td>
<td>The database resides with the land monitoring institute</td>
</tr>
<tr>
<td>UK</td>
<td>Restricted, held within Rural Payments Agency</td>
</tr>
</tbody>
</table>
### Geometry

<table>
<thead>
<tr>
<th>Country</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Vector: Single agricultural fields are represented as polygons</td>
</tr>
<tr>
<td>CH</td>
<td>Vector: Data is point information on 100m x 100m regular grid, represented as raster cell of 100m</td>
</tr>
<tr>
<td>DE</td>
<td>Vector: Areas are represented by polygons.</td>
</tr>
<tr>
<td>FI</td>
<td>Vector: Areas are represented as polygons</td>
</tr>
</tbody>
</table>
| NL      | Top10NL: Vector: Areas are represented as polygons  
BRP/LPIS: Vector: Areas are represented as polygons  
LGN: Vectors and 25 m raster |
| NO      | Vector: Areas are represented as polygons |
| UK      | Vector: Areas are represented as polygons |

### Attributes

<table>
<thead>
<tr>
<th>Country</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Crop type (code list, implying agricultural use): 5 main categories (arable, grassland, alpine meadows, vineyards, other) &gt;200 sub-categories (e.g. “soy beans”)</td>
</tr>
<tr>
<td>CH</td>
<td>Land cover (27 classes), Land use – (46 classes), CH-standard (mixed LULC, 72 classes)</td>
</tr>
<tr>
<td>DE</td>
<td>Feature types and attributes (implying land cover and/or land use)</td>
</tr>
<tr>
<td>FI</td>
<td>Crop type (code list, implying which plants that has been grown in each polygon in each year).</td>
</tr>
</tbody>
</table>
| NL      | Top10NL- Type of land use (code list)  
BRP: crop type (code list)  
LGN: crop type (code list) |
| NO      | Land type (code list, implying land cover and land use) |
| UK      | Unknown |

### Maintenance

<table>
<thead>
<tr>
<th>Country</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>1*/year updated directly by farmers (with the help of local authorities) as part of their payment proposal. The dataset is a prognostic dataset, as it covers the situation that is planned to be managed in the next year (autumn 2012 → map for agricultural management in 2013)</td>
</tr>
<tr>
<td>CH</td>
<td>Previously updated every 12 years. The cycle starting in 2013 will be completed in 9 years. and future cycles (after 2022) are expected to be completed in 6 years.</td>
</tr>
<tr>
<td>DE</td>
<td>Periodically updated by the Cadastral and Surveying Authorities of the Länder (German Federal States) and BKG.</td>
</tr>
<tr>
<td>FI</td>
<td>Updated by the Ministry of Agriculture and Forestry (MAVI, TIKE offices) based on information provided by the farmers as part the agricultural subsidies system.</td>
</tr>
</tbody>
</table>
| NL      | Top10NL; every 2 years  
BRP: every year  
LGN: every 3-5 years |
| NO      | Continuously updated by local administration as part of their management of the planning act. Changes not linked to the planning act updated by the national authority with a cycle of 5-7 years. |
| UK      | Updated annually based on farmers returns and updates from the national mapping agency. |
Attachment 2: Database merging – national examples

A2.1 Austria

LISA Generalization

Step 0: Input LISA Land use & Land cover (MMU 25m²-50m²)

Step 1 – 4 are iteratively repeated for 3000 m², 1 ha, 5 ha and 25 ha

Step 1: Polygons are exported separately based on their main class in LISA land use:
- Settlement
- Transportation
- Agriculture
- Forest
- Natural areas
- Water

Step 2: Expand and Shrink is applied with different buffer sizes based on the land use class

Step 3: Merging
Layers are merged in the following default-order:
  Water, Transportation, Settlement, Agriculture, Natural areas, Forest

The visual interpretation of the matrix (currently only Forest or Agriculture, otherwise default) is applied per LISA-test area and modifies the merge-order as needed. Forest matrix leads to a higher priority of Agriculture and Natural area, whereas Agricultural matrix prioritises Forest and Natural areas.

Merge order in test areas where the Forest is the matrix:
  Water, Transportation, Settlement, Agriculture, Natural areas, Forest
Merge order in test areas where the Agriculture is the matrix:
  Water, Transportation, Settlement, Forest, Natural areas, Agriculture

The matrix guarantees not to lose landscape-characteristic elements in the merging process (Figure A2.1.1). In the process of generalization objects within a specific distance are merged. If the default order is used, Forest objects are overlaid by Agriculture objects. This can lead to forest objects below MMU size, which would be assigned to their agricultural neighbours. Therefore the matrix is applied.

Figure A2.1.1: Prevent the loss of landscape-characteristic elements
**Step 4:** Holes are assigned to neighbouring objects with the longest shared outline (except Water and Transportation).

**Step 5:** Objects below a specific size are assigned to neighbouring objects with the longest shared outline (except Water and Transportation).

A semantic transformation has to be applied before the geometric generalization to CORINE is taken into account. This semantic transformation is achieved by mapping the classes manually to its CORINE level 3 equivalent. Afterwards the process steps mentioned above are applied using the CORINE classes.

CORINE level II and level I is generated by merging the individual polygons in their sub classes. E.g. merging level 3 classes 1.1.1 and 1.1.2 to level 2 class 1.1.

**Figure A2.1.2:** The work-flow of LISA generalization in Austria

**Generating LISA – Land cover:**

Mostly automated, object based segmentation/classification of orthophotos (combining DSM – digital surface model) and multispectral satellite imagery.

<table>
<thead>
<tr>
<th>Original</th>
<th>3000m² Level III</th>
<th>1ha Level III</th>
<th>5ha Level III</th>
<th>25ha Level III</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Original" /></td>
<td><img src="image2" alt="3000m² Level III" /></td>
<td><img src="image3" alt="1ha Level III" /></td>
<td><img src="image4" alt="5ha Level III" /></td>
<td><img src="image5" alt="25ha Level III" /></td>
</tr>
</tbody>
</table>

**Figure A2.1.3:** Generalization levels used in LISA generalization in Austria
Generating LISA – Land use:

Production of land use (depending on the availability and without defined licensing policy):

- Digital cadastral parcels
- INVEKOS (integrated administration and control system / IACS)
- Water network
- Transportation network
- Building registry
- Other Austrian datasets
  - Dry grasslands
  - Bog and mire conservation database
  - Wetland inventory
  - Inventory of riparian areas

The following basic rule for defining the geometric outline of an object is applied:

- All areas which are bigger than the MMU (minimum mapping unit) and differ in one or more attributes are differentiated from each other. The goal is to have polygons with highest possible homogeneity, whilst preserving the MMU.
- Beside the integration of additional sectorial datasets, research is going on to automatically derive certain land use types from land cover patterns.

A2.2 Finland

The Finnish database merging and generalization process in CLC2000 and CLC2006

Phase 0: The acquisition of the needed spatial data sets and satellite images

Phase 1: Preprocessing and interpretation of satellite images

Step 1.1 Detection and masking out clouds, haze and their shadows
Step 1.2 Radiometric, atmospheric and topographic correction
Step 1.3 Mosaicking of single scenes into calculation units and nationwide data
Step 1.4 Automated and semi-automated interpretation using for example k-NN method (forest parameters), density slicing and decision tree classification
Step 1.5 Detection of land cover change layers (type and degree) based on image-2-image Comparison

Outputs: HR forest layers (tree species %, crown cover %, tree height), these parameters are interpreted for each pixel; vegetation coverage and type maps; image based changes.

Phase 2: Preprocessing of digital map data and environmental registers

Step 2.1 Identifying and digitizing of missing data
Step 2.2 Converting vector data into raster format (including needed data treatments e.g. buffering point data representing buildings and reclassification of data)
Step 2.3: To ease the further processing steps datasets are merged into larger units

Outputs: HR raster layers (25 *25 m equal to the satellite images), which can overlap each other
Figure A2.2.1: The data flow and methodology of the Finnish CLC2000 and CLC2006 projects.
Phase 3: Integration of layers and production of national CLC with appropriate metadata layers

Step 3.1: Combining HR layers (outputs of previous phase 1 and phase 2 into single datasets.

Step 3.2: Defining national CLC class for each pixel. The proper class is determined using the priority list, which is based on the quality of the data; spatial accuracy, age, thematic accuracy. The priority list is:

- Digitized artificial surfaces
- Water bodies
- Road
- Other artificial surfaces
- Agricultural lands
- Marshes and peat bogs
- Bare rock and beaches, dunes and sand plains
- Semi-natural areas and forests

Step 3.3: Comparing previous national version of CLC with current national version together with image based changes in order to determine the changed areas. Only possible changes are approved and based on this information the final classification is done.

Step 3.4: The same process is repeated in order to get source and age metadata for each pixel

Output: National CLC data in 25 m raster format

Phase 4: Generalization

Step 4.1: Preparation of the data, removal of very small areas (1-2 pixels) and elimination of narrow features like roads and rivers.

Step 4.2: Generalization of water bodies. During the process small islands are aggregated to water or neighboring islands and small lakes near each other are combined (merged).

Step 4.3: Generalization of artificial surfaces. Individual houses and road network disappear and small non-built-up areas inside built-up areas are merged into built-up classes.

Step 4.4: Generalization of agricultural areas starts with creating heterogeneous classes (2.4.2 and 2.43). Arable lands are generated into 25 ha.

Step 4.5: Generalization of wetlands, where parcels classified either to water bodies or to marshes and small marshes are merged into neighboring water bodies

Step 4.6: Generalization of forest and semi-natural areas is a process, where proper forest class for each parcel is defined iteratively going through each class according to the CLC priority list. Heterogeneous areas are classified as mixed forest.

Step 4.7: The remaining small parcels are dissolved into neighboring areas according to a priority list. If no thematic affinity is found, parcels are split between neighbors.

Step 4.8: The resulting raster dataset is vectorized to form the EU level data

Output: CLC data in 25 hectare vector format
Figure A2.2.2 Finland: National CLC data (25 m raster) (left) and standard (EEA) CLC data (right)

Figure A2.2.3 Standard (EEA) CLC boundaries superimposed on the Finnish national CLC data (left) and on IMAGE2006 (right)
A2.3 Netherlands

The Dutch LGN6 dataset - Production methodology

Based on G.W. Hazeu, in prep. Operational land cover and land use mapping in the Netherlands. In: Land cover and land use mapping in Europe (EARSeL book series on Remote Sensing)

The production methodology for the LGN6 dataset did consist of two main steps: 1) an assignment of LGN classes to Top10vector objects followed by 2) a rasterization process (Hazeu, Schuiling et al. 2010; Hazeu, Bregt et al. 2011). Figure A2.3.1 and Figure A2.3.2 are flow diagrams representing the different steps in the two main phases of the production process.

The object based classification started with the aggregation of Top10vector classes into a limited number of LGN classes. Secondly, urban areas from BBG2003/BG2003 were integrated with the Top10vector database (version2006) to define the urban extent. The integration was based on decision rules taking into account Top10vector, the BBG2003/BG2003 and surface area information. Thirdly, Top10vector objects were aggregated into eight LGN monitoring classes. Land cover changes were attributed by comparing 2003/2004 and 2007/2008 imagery with the Top10_upd database as result. The marked land cover changes (Top10_chg) finally appeared as a binary mask in the LGN6chg database (Figure A2.3.2). The updating methodology is comparable with the update of previous LGN databases and extensively described in De Wit (2003).

The following methodological step was to attribute information to the Top10vector objects on basis of additional databases. LGN5 and Natural Areas database (BKN2007) in combination with Top10vector made it possible to attribute salt marshes, heath land in the coastal areas and raised bog to the Top10vector objects resulting in Top10_lgn6 database (Table A2.3.1). Finally, forest and heath-land areas were selected as input for the classification based on satellite imagery.

Figure A2.3.1: Data flow diagram for the object based LGN6 production phase. The topographical database (Top10vector) is the geometrical basis for updating (Top10 upd) and crop classification (Top10 agr). It is integrated with urban information (BBG/BG2003), land cover changes based on comparison of remote sensing images from different time steps and information on nature areas (BKN2007 and LGN5).

In a separate step, the agricultural areas were classified. The classification started with the selection of the agricultural parcels of the Top10vector database (pastures and agricultural land) (Top10 agr). Agricultural parcels with more than one agricultural crop were manually subdivided on basis of satellite imagery. A multi-temporal classification of satellite imagery on basis of NDVI was carried out. The
differences in phenology for seven agricultural crops were used to classify them (grassland, maize, potato, sugar beet, wheat, other crops and flower bulbs). A non-supervised classification with manual correction was applied. Classification results were compared with CBS agricultural statistics. Finally, each crop parcel was attributed to the main crop type (LGN6crop database) (De Wit and Clevers 2004; Hazeu 2006).

Figure A2.3.2: The second step of the LGN6 production process. The LGN6 grid cells (LGN6ras_basis) were enriched with additional information from GIS databases (Top10_houses and infrastructure, BKN2007 and LGN5) and remote sensing classifications (forest, heath and crops) resulting in the final database (LGN6 and LGN6chg).

The second step started with the conversion of Top10_lgn6 objects into 25*25m grid cells (LGN6ras_basis) (Figure A2.3.2). The location of the grid cells corresponds with the previous LGN versions. Crop information from the LGN6 crop database (LGN6crop) was attributed to the agricultural areas of LGN6 grid database (LGN6ras_basis). Also forest and heath-land unsupervised classification results were attributed. The extent and type of different grassland, dune and swamp areas were defined by combining data from the topographical (Top10vector), natural areas (BKN2007) and LGN5 database information. Next, the buildings (Top10vector buildings) were buffered with 10m, rasterized and attributed to the 25*25m grid cells. For a selection of main roads and railways rasterization and attribution was also applied after buffering them with 5-15m depending on their type. All information was added to the LGN6ras_basis database with the LGN6 and LGN6chg database as final result (Figure A2.3.2). The LGN6crop and the LGN6chg were validated (see for details Hazeu, Schuiling et al. (2010); Hazeu, Bregt et al. (2011).

Summarizing from the methodology described above the following main database merging steps can be distinguished:

- Define the extent of urban areas by merging information of the land use map (BBG) with the topographical map (Top10NL). The combination of BBG attribute information with Top10NL attributes and a surface area threshold defined the extent of urban areas. See Figure A2.3.3 below (left the Top10NL with defined urban area (red/orange) on basis of BBG, right Top10NL)
- Define and attribute on basis of LGN5 and BKN2007 the area of nature classes salt marshes, heath-land in the coastal areas and raised bog
- All information of several sources in the rasterization process (Figure A2.3.2) is merged into the LGN6ras_basis dataset
Table A2.3.1 Datasets used in the final thematic assignment of the LGN6 classes (Hazeu, Bregt et al. 2011).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top10vector*</td>
<td>greenhouses, orchards, tree and fruit nurseries, main roads and railways</td>
</tr>
<tr>
<td>Top10vector buildings</td>
<td>primary and secondary built-up areas; buildings in agricultural areas</td>
</tr>
<tr>
<td>Land use and urban areas of 2003</td>
<td>urban classes</td>
</tr>
<tr>
<td>Natural areas of 2007</td>
<td>sand and dune classes, swamp classes, natural grasslands</td>
</tr>
<tr>
<td>LGN5</td>
<td>fresh and salt water, salt marshes, raised bogs classes, swamp classes</td>
</tr>
<tr>
<td>Satellite imagery**</td>
<td>grassland, maize, potato, sugar beet, wheat, other crops and flower bulbs; forest classes; heath land classes</td>
</tr>
<tr>
<td>Aerial Photographs**</td>
<td>dune classes</td>
</tr>
</tbody>
</table>

* Top10vector geometry and thematic content plays a role for each LGN6 class at an earlier stage.

** Both remote sensing sources were used to detect land cover changes

Figure A2.3.3: Left the Top10NL with defined urban area (red/orange) on basis of BBG , right Top10NL.
A2.4 Norway

Norwegian generalized land resource maps AR50

Step 0: Input data is the detailed land resource map AR5 (1:5,000)
Step 1: Each polygon in AR5 is assigned a class from the (simplified) AR50 nomenclature
Step 2: The polygons belonging to each AR50 class is selected from the AR5 layer and transferred to a separate layer. The result is one layer for each AR50 class.
Step 3: Each layer is generalized by applying a simple algorithm (Strand and Moum 2000) where the polygons first are enlarged through a buffer operation and then contracted through a second (negative) buffer operation and polygons below a threshold size are removed.
Step 4: The layers are superimposed in a prioritized sequence. Resulting holes or small polygons filled or removed by attaching them to neighbors – again following a prioritized sequence.

Figure A2.4.1: The generalization technique described under Step 3 (Strand and Moum 2000)

Figure A2.4.2: AR5 (left) and generalized AR50 (right)
Norwegian CLC2000

Step 0: Input data is a large number of datasets (approx. 30): Land resource data, digital topographic maps, building register, road data and a number of specialized datasets – all vector data.

Step 1: All input datasets were rasterized to a uniform 25x25 meter pixel raster creating a uniform multi-layer raster dataset where every relevant attribute in each of the input datasets was represented as a separate layer. This is the Production Data Set (PDS).

Step 2: A new (binary) raster layer was created for each CLC level 1 class by setting up the relevant criteria for selection from the PDS. Individual pixels could be extracted into several of these layers, and some pixels might not be extracted into any layer. Each layer was converted to a polygon coverage.

Step 3: Each CLC Level 1 layer was generalized by applying a simple algorithm where the polygons first were enlarged through a buffer operation and then contracted through a second (negative) buffer operation (see example 1 above). Polygons below a threshold size were then removed.

Step 4: The five CLC Level 1 layers were combined (in a prioritized sequence). Resulting small polygons or holes not assigned to any CLC class was removed by attaching them to their neighbors. The result was a CLC Level 1 map.

Step 5: Each CLC level 3 class was processed using the same procedure as described in step 2 and 3 above, but using the CLC Level 1 map as a mask limiting the occurrence each CLC Level 3 class to the area covered by the corresponding CLC Level 1 class. Mixed agricultural classes were not used – only the “pure” agricultural classes.

Step 6: The CLC Level 3 maps were merged into a single map.

Step 7: Polygons belonging to agriculture classes were populated with data from detailed land resource map in order to determine the distribution of land cover classes in each polygon. Polygons with mixed agriculture or considerable non-agricultural areas were assigned to the appropriate CLC mixed classes.

Step 8: Visual inspection and manual editing of the final result.

Figure A2.4.3: Simplified illustration of the data flow in the Norwegian CLC2000 project
A2.5 Switzerland

Swiss CLC2000

Step 0: Input data is a large number of datasets:
- **Vector data**: digital landscape model (26 classes), building register, national inventories of protected areas (3), a number of specialized datasets (3)
- **Raster data**: Land use map, forest mixture degree, DTM

Step 1: all vector data is generalized separately with GIS software

Step 2: the empty CLC map is filled stepwise in a defined order with the available data, including the NDVI of Image2000.

Step 3: After each import of a layer, the map is checked visually with the Image2000 as background and adjusted if necessary.

Step 4: manual editing of very few classes during the processing (step 2 -3)

Step 5: cleaning up features according CLC definition (size of polygons, sliver,...)

Step 6: leftover empty area is divided up in the remaining 4 classes, which are separated according the biogeographical area and DTM.

Step 7: Visual inspection and manual editing of the final result.

References: Still not published but expected by the end of the year 2012

Figure A2.5.1: Diagram showing data flow and stepwise generalization and integration in the Swiss CLC production
A2.6 United Kingdom

CLC production in the UK

The CLC1990, CLC2000 and CLC2006 production in the UK used broadly similar semi-automated generalization procedures. The CLC2006 production will be described below, as this is the most advanced version of the approach.

The CLC and LCM specifications created a number of challenges for production of CLC based on national LCMs (Figure A1.6.1). At the spatial level, the UK LCMs have around 100 times the number of objects as the equivalent CLC product. Thematically, the UK nomenclature focuses on land cover and habitats, whereas the CLC nomenclature is a mixture of land cover and land use classes.

Figure A1.6.1: A comparison of LCM2000 and CLC2000 in the United Kingdom.

In most cases, rules were developed to allow thematic conversion in an automated or semi-automated fashion. The generalisation from the 0.5 ha MMU of LCM to the 25 ha is relatively straightforward for most classes. Others however required some form of manual intervention, particularly for the mixed classes, those classes that were based on land use and where CLC classes are impossible to map due to their complex contextual definitions (e.g. airports).

All of the UK CLCs have used some form and amount of visual interpretation of satellite imagery as part of the process. Imagery was made available for CLC1990 from the LCMGB production and for CLC2000 and CLC2006 the IMAGE2000 and IMAGE2006 dataset were extensively used.

The main steps in producing CLC2006 for the UK were therefore (1Spatial, 2007):

1. Extraction of the urban land use classes from CLC2000 (CLC Level-1 class 1.*.*). This layer was updated with appropriate recent land use changes via visual interpretation against IMAGE2006 and topographic maps.

2. Generalisation of the LCM2007 data into CLC classes as far as possible by thematic re-coding and spatial merging. Thematic transformation using a look up table to populate the input national dataset with an equivalent CLC class code. Spatial transformation by merging objects less than 25 ha with adjacent objects of the same class or similar using the CLC similarity table. Remaining clusters of the object < 25 ha merged to form possible mixed classes.
4. Visual checking with IMAGE2006 for geometric error, etc.
5. Detailed visual / interactive edit, using IMAGE2006 and topographic maps to complete conversion of spatial and thematic specification.
6. Edge match between separate production units and quality checking.
7. Final quality checking, editing, and delivery of CLC2006 stock product.
Synchronisation
Table of content

Table of content ........................................................................................................................................ 1
1. Executive summary .................................................................................................................................. 2
2. Specific objectives .................................................................................................................................... 2
3. Methods .................................................................................................................................................. 2
4. Results .................................................................................................................................................... 3
   4.1. Existing and potential user requirements ......................................................................................... 3
   4.2. Existing land monitoring cycles ........................................................................................................ 6
   4.3. Feasibility and constraints related to synchronisation of data gathering ........................................... 9
5. Proposals .................................................................................................................................................. 11
1. Executive summary

Synchronization of data gathering is an effective tool in the optimization of resources in monitoring activities, having high potential benefits in the future harmonized European land monitoring. Planning of interconnected land monitoring activities requires well-defined update cycles, well-organized information and data flow should satisfy needs in vertical and horizontal cooperation.

Synchronization of data gathering would include in this environment:

- Synchronized collection and sharing of Earth Observation imagery (satellite / aerial)
- Synchronized collection and sharing of in-situ data

2. Specific objectives

Specific objectives of the task are:

- to gather and examine user requirements,
- to explore current situation regarding the scope (spatial and thematic) of land monitoring initiatives,
- to explore current situation regarding the periodicity of land monitoring (LULC) initiatives.

Based on this knowledge the task describes the feasibility and constraints of the synchronization of data gathering, provides proposals to improve the time- and cost-effectiveness of harmonized European land monitoring.

3. Methods

To collect information about the topic the participating experts from different countries formed small groups (so-called “niche”) and each group with a leading moderator (knowledge carrier) discussed a specific part of the task, following a list of so called seed questions. The groups were separated in different rooms or locations, however the participants were able to change groups depending on their interest on the specific theme. All of the niches had predefined questions, so-called “seed questions”, which ones provided the guidelines for the subtopics. After each small niche session each moderator held a presentation to the plenum to share insights and results of the small group conversation. The moderators with the help of rapporteurs took notes about the information and opinions in each niche to provide relevant bases for the finally created report.

The seed questions for each niche were as follows:

**Niche 4: Existing and potential user requirements**

- What kind of existing annual, multi-annual and non-periodic user requirements do you know ((sub/trans)-national / EU / Global)?
- What kind of meaningful potential annual, multi-annual and non-periodic user requirements can you imagine / propose?
Ways to improve the synchronisation of data gathering

- What are the aspects determining the periodicity of user requirements (legal, budgetary, scientific etc.)?

Niche 5: Existing land monitoring cycles
- What kind of existing annual, multi-annual and non-periodic land monitoring cycles do you know ((sub/trans)-national / EU / Global)?
- Provide a critical analysis of update frequency of these (is it meaningful to gather data so frequently / is data collection frequent enough?).
- Do you know any examples of synchronization of data gathering (horizontal / vertical)?

Niche 6: Feasibility and constraints
- What kinds of constraints define the periodicity of data gathering (financial / organizational / legal / scientific)?
- Which are the constraints of (horizontal / vertical) synchronization of data gathering?
- What are your proposals to overcome recognized constraints?

4. Results

The result section is organized following the system of seed questions. Each subchapter contains the essence of the small group conversations concentrating on the main subject of the task.

4.1. Existing and potential user requirements

In Niche 4 participants tried to collect existing and potential user requirements of land monitoring activities (especially timing aspects) and tried not to consider pessimistic aspects such as administrative difficulties or financing.

Main aspects determining periodicity

Participants identified the following main aspects determining the periodicity of user requirements:

Legal / financial aspects:
- Reporting obligations (national / EU / global),
- Monitoring linked to national budget (1-year period),
- Monitoring linked to EU budget (6-year period)

Scientific aspects:
- Monitoring linked to crop growth (annual / seasonal / monthly),
- Monitoring linked to other annual cycles (e.g. melting of glaciers)

Existing reporting obligations and uses determining user requirements collected:
- Water Framework Directive (6-year period)
- Habitats Directive (6-year period)
Ways to improve the synchronisation of data gathering

- Kyoto protocol (annual greenhouse gas inventories required: reporting on land use, land-use change, and forestry - LULUCF; national communications in every 4-5 years)
- EEA reports and indicators:
  - Land & Ecosystem accounts
  - State and Outlook of Environment Report - SOER (5 years - fixed in founding regulation of EEA)
  - Other reports: Year of Biodiversity, Year of Forest, Year of Water, etc.
- Forest Europe (4 years, last report in 2011)
- Forest Resource Assessment - FAO, global (5 years, last report in 2010)
- Other national reporting obligations (to EUROSTAT, UN, etc)
- CORINE land cover inventory (6-year cycle, last report in 2006)
- Crop fore-casting
- Policy assessment studies (effectiveness of policy measured by change in land-related indicators)
- Monitor spatial developments
- Effect of climate change on land (adaptation of land cover)
- Control of agricultural subsidies

Meaningful potential user requirements

Participants estimated optimal monitoring cycles for specific land cover related themes as follows:

- Urban areas: 1 (max 3) years
- Forestry: 5-6 years
- Crop monitoring: 1 year / crop season
- Subsidies (LPIS): 1 year reporting (3-5 year for ortho-photos)
- Grassland (semi-natural) 3-5 years
- Glaciers: 1 year
- Nature reserves, mountains: 15-25 years

It has been noted that different climatic zones / geographic locations may affect the optimal monitoring cycle.

Conclusions on user requirements

Ideal (scientific) update frequency is dependent on:

- Scale: Larger level of geometric detail requires more frequent updates, while continental-level monitoring does not require updating so often.
- Thematic content ("class" or "attribute"): More dynamic classes/layers, such as built-up need more frequent monitoring than most natural classes, such as bare rock. Similarly, some change processes (eg. forest growth on northernmost territories) are meaningless to be monitored so often, as process is hardly detectable even by remote sensing.
Ways to improve the synchronisation of data gathering

In the practice, user requirements are mostly determined by legal / financial / political aspects:

- **Periodic**: Requirements linked either to reporting obligations or budgetary cycles,
- **Non-periodic**: Many user requirements are non-periodic (eg. policy audits, production of land use/cover statistics, requests for particular analysis). Non-periodic requests do, however, rely on the availability of data. Users having these non-periodic requests expect a continuous or periodic monitoring program to be on-going in the “background” providing data for these non-periodic requests any time.

**Special national particularities and/or best practices mentioned, which should be taken into consideration*:**

- **Finland (FI)**: The National Forest Inventory (NFI) of FI (11th update) takes totally 5 years. Every year 20% of the total sample area measured in the field with about 10 000 field sample plots annually and they cover the whole of FI, not only a specific area. Thus results can be calculated annually for the whole of Finland. The needs for annual data are increasing because of the obligation of greenhouse reporting.
- **Austria (AT)**: 3 years period forest inventory including the whole country.
- **France (FR)**: 10 years period forest inventory (all over the country) including forest types. (Currently under test: how to reach 5-6 year cycle with limited detail.) Ortho coverage in every 5 years, in this way many classes reported in 5 year period: fruit plantations, buildings. Beside it parallel continuous data collection by cadastral.
- **Italy (IT)**: 3 year period for forest inventory covering the whole country. Another 3 year period for land cover update (user requirement from JRC).
- **Portugal (PT)**: Because of extended forest fires 1 year would be ideal.
- **Spain (ES)**: In agricultural areas 1 year cycle area frame sampling on selected spots (the aim to monitor the evolution and dynamics of agricultural land in Spain.) Requirement in SIOSE (Spanish Land cover/Land use Information System): “as frequent as possible” regarding national needs. The urban areas (and agricultural lands) monitored around 1 year cycle, but other classes around 4 year cycle (optimally around 3 year). Note, Mediterranean areas are more dynamic, the effect of climatic/geographical characteristic influences the update cycles.
- **Belgium (BE) / Flanders**: They do CLC only as obligation. A RS based forest survey is performed every year or every two years. No forest types are mapped, the survey is independent from Forestry (cutting / planting and “upper green” / “lower green” distinction). Urban green is part of it. It is not based on the expressed need of one organization; it is an offer to raised needs. (In the remaining part of BE (outside the Flanders region) no land monitoring projects.) The limits of land monitoring depend on the financial and technical availability of RS data.
- **Switzerland (CH)**: Big campaign in forest inventory every 10 years, but the system is changed: 1/9 of the country is updated every year (each year an evenly distributed representative sample for the country). Area frame survey based on RS and field survey. The survey is a statistical one relies on sample points placed every 100m grid. As attributes 3 nomenclatures are applied: LC, LU, and mixture of both. The updating frequency was previously 12 years, currently is a mixture of 6 and 3 year period. Topographical maps are updated every 3 years. The “only” 2-3 year cycle orthophoto
coverage mentioned as a constraint. CH did a survey to check the needs of land cover data, but no results yet.

- Netherlands (NL): One specialty is the water level monitoring (NL under see level). The characteristic of special kind of plants on wet areas provide the opportunity of monitoring, which one is done in 1 year update cycle. For the knowledge of phenological cycle of the crops (development of crops) monthly update cycles are applied (agriculture crop monitoring – it is provide information for companies about the exact harvest time). In precision agriculture for the monitoring of nitrate and pesticides infiltration (amount of nitrates that go into the water) also monthly update cycles are introduced.

*Note, since most of the participants came from the data provider side, it was not so easy to identify user requirements, participants first idea was usually bound to existing mapping projects.

4.2. Existing land monitoring cycles

Examples of existing land monitoring cycles collected by participants

- **Global/EU/EEA**: CLC, LUCAS, HRLs, Euroregional map, Forest Europe (reporting on national level, but for European purpose), Forest Resource Assessment (global), GlobCover, NATURA2000 monitoring (habitat conservation status),
- **National**: Forest inventories – Land use (CZ, DE, FI, ES, UK), database of buildings (FI), environmental data for EEA, environmental database (FI), BBG LU (NL), LGN LC (NL), topographic database (NL), Hydrology data (FI), LCM (UK), SIOSE (ES), DLM-DE (DE), LISA (AT), LPIS (FI), DEM (CZ, FI),
- **Sub-national**: cadastral data, forest fire databases (eg. PT), flood monitoring, LPIS, ATKIS layers (DE), Forest regional map (FL), Forest inventory LC (ES), Federal Forest Inventory – land cover (DE), Land resource map (NO), , ZABAGED - Fundamental Base of Geographic Data (CZ), Digital terrain Model (CZ)

**Norwegian examples**

**Land resource maps**

The basic maintenance of Norwegian land resource maps is done on a day to day basis by municipal administrations as part of the daily routine in public management (eg: When a developer reports that the construction of a house is completed, the public officer updates the building register and the map before he/she closes the case). This data set, maintained by the local authority, is considered to be the original map.

Every year, a copy is submitted to the national authority (NFLI), contributing to an updated national land resource database. NFLI will (over a few days) correct technical mistakes (inconsistencies) and return a corrected version to the local authorities.

For 15% of the administrative units, NFLI locks their data for a period of several weeks and conducts a thorough revision based on aerial photographs before the corrected version is returned to the local authorities. This “periodic maintenance” is on average carried out every seventh year for each local authority, contributing to a periodic overhaul of the dataset.

**Monitoring (area frame surveys)**
Ways to improve the synchronisation of data gathering

Approximately 10 - 20% of the sampling units (depending on the survey – we have several) are surveyed every year.

Examples of synchronization

- Norwegian public institutions cooperate on running a periodic aerial photography campaign with regular coverage of the entire country,
- National LC mapping is executed in a way that data can be aggregated into European CLC (FI, SE, UK),
- NL: LGN data are just started to be synchronized with the topographic data coming from each region (as soon as there is an update available on province level, it is integrated into LGN),
- Flanders: changes in built-up are reported by municipalities so that they can be integrated to regional databases,
- HU: Orthophoto acquisition is already synchronized, scale agreed between LPIS, Forestry and topographic mapping (represented by 3 different institutions),
- LUCAS survey has been synchronized with CLC2012,
- ES: Synchronized gathering of RS data
  - SPOT and LIDAR data are acquired at the same time,
  - Orthophoto flying plans are jointly decided by the provincial and federal governments and even linked to LPIS updates,
  - Reference data are common, information needed for several purposes. This synchronisation is accomplished via dedicated co-funding that provincial governments receive only if they adhere to the schedule. 66% of the costs are covered centrally, the resulting data are then co-owned by both administrative levels (federal and provincial).

Critical analysis of existing land monitoring cycles

Ideal update frequency depends on scale and thematic content

- National / sub-national systems are closer to scientific optimum (if well financed);
- Synchronization is (usually) easier within a country than between countries and easier between EU institutions than between different administrative levels;
- Vertical synchronization (e.g. between national / EU) is mainly driven by EU (existing surveys like CLC, reporting obligations);
- Horizontal synchronization (between countries) is occasional.
Niche 5 team has provided a complex visualization method for critical analysis of land monitoring cycles concerning synchronization (see Table 1).

### Table 1: Complex visualization method for critical analysis of land monitoring cycles

<table>
<thead>
<tr>
<th>Scope/Update frequency</th>
<th>Sub-national</th>
<th>National</th>
<th>EU/EEA-level</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 10 yrs</td>
<td>DTM</td>
<td>Soil</td>
<td>Geological mapping</td>
</tr>
<tr>
<td></td>
<td>ATKIS forest parcel LU (DE)</td>
<td>ATKIS grassland/arable LU (DE)</td>
<td>Hydrology, DEM (FI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CLC 33x, 324-31x</td>
</tr>
<tr>
<td>4-10 yrs</td>
<td>Federal forest Inv. LC (DE)</td>
<td>Forest Inventories – Land use (CZ, DE, ES, FI, UK)</td>
<td>European CLC</td>
</tr>
<tr>
<td></td>
<td>Forest inventory LC (ES)</td>
<td>LCM (UK)</td>
<td>Forest Resource Assessment (Global)</td>
</tr>
<tr>
<td></td>
<td>ZABAGED - Fundamental Base of Geographic Data (CE)</td>
<td>LGN LC (NL)</td>
<td>NATURA 2000 monitoring</td>
</tr>
<tr>
<td></td>
<td>Digital terrain Model (CZ)</td>
<td>Topographic database (FI)</td>
<td></td>
</tr>
<tr>
<td>1-3 yrs</td>
<td>Rivers (IT, ES)</td>
<td>BBG LU (NL)</td>
<td>HRLs</td>
</tr>
<tr>
<td></td>
<td>Digital surface model (IT)</td>
<td>Forest+habitat (AT)</td>
<td>LUCAS</td>
</tr>
<tr>
<td></td>
<td>Forest regional map (FL)</td>
<td>SIOSE (ES)</td>
<td>Forest Europe</td>
</tr>
<tr>
<td>yearly</td>
<td>LPIS</td>
<td>Topographic dataset (NL)</td>
<td>CLC 1xx, 31x-324</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DLM-DE (DE)</td>
<td>Urban Atlas</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Forest fires</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>EuroRegionalMap</td>
</tr>
<tr>
<td>&lt; 1 yr / continuous</td>
<td>Forest fires (eg. PT)</td>
<td>Database of buildings (FI)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Flood monitoring</td>
<td>Digiroad (FI)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cadastral data</td>
<td>DEM (CZ)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Env. info (FI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATKIS roads + built-up (DE)</td>
<td>Land resource map (NO)</td>
<td>Environmental data for EEA OS mapping (UK)</td>
</tr>
</tbody>
</table>

**Legend:** Horizontal axis (columns) represent the spatial scope of initiatives (sub-national/national/EU/global), vertical axis (rows) represent the update frequency. Inventories are placed according to their actual implementation, inventories marked with blue characters represent the ideal situation. Red arrows show data input directions from one inventory to another one. Meaning of black arrows: the same situation is valid for other spatial levels.

**Explanations to Table 1**

High-frequency monitoring needed for e.g. built-up, roads, forest fires, agricultural crops, clearcuts, floods, and storm damages. Low-frequency monitoring needed for e.g. natural vegetation, rivers, and forest growth. Moreover, CLC classes should be treated separately, not as a whole inventory. E.g. 1xx would require yearly update, while updating 33x classes is sufficient in every 10 yrs. Forest clear cutting can be monitored every year (being an abrupt change, it is simple to detect), while forest growth is detectable only in 6 (Central-EU) – 30 (IS, Scandinavia) year cycles (being a gradual, difficult-to-detect process). The concept of GIO Land High-Resolutions Layers (HRLs) serving as a complement of CLC is a good example of the different update frequencies. The concept says that lower-level of detail database (CLC) is sufficient to be updated every 6 years; while spatially much more (but thematically much less) detailed HRLs are useful to be updated every 3 years. From the second point however, also this concept can be treated critically: some CLC classes are not needed to be monitored so frequently and not all HRL thematic layers are dynamic enough.

---

1 Not all inventories mentioned are inserted to the matrix. E.g. Austrian LISA, which is still in a concept phase, update frequencies are not decided yet. Most likely it will be different for different classes.
to be worth so frequent monitoring efforts. Topographic databases are usually updated in this “ideal” way, layer by layer, depending on dynamics of the class. E.g. ATKIS roads and built-up layer is updated continuously, within one year period, while forest (as land use) layer is updated in more than 10-years -long cycle. Turnaround time is also an important aspect, firstly because it affects feasible update frequency, secondly by being part of the user requirements. Timing is often driven by reporting periods and dates rather than the occurrence of measurable change. For example, there have been requests to the SATChMo project for annual change mapping, but there is very little if any change within some sample areas over such a short period. Also, the level of change is saturated by uncertainties in the approach.

4.3. Feasibility and constraints related to synchronisation of data gathering

Identified factors supporting feasibility of synchronisation:
- Open data policy,
- Centralized data acquisition,
- Existence of INSPIRE/EAGLE- working groups,
- Existence of Interest/thematic working groups,
- Legal obligation for land monitoring,
- Clear legal assignments.

The following legal / organizational / financial constraints have been identified:
- Missing legal framework, organizational issues and level and recognition of responsibilities, lack of clear legal assignments.
- Even if funding/staff is available on each horizontal unit (eg. in each federal state), it might not available at the same time
- Especially in new member states of EU, the legal mandate for monitoring is very important. In these countries often “top-down approach” is preferred.
- For MS with long LM history internal orders are already enough and with the experiences experts can justify the need and added value,
- Lack of communication / miscommunication between stakeholders,
- The personal element: building barriers, defending own business, own carrier,
- Changing responsibilities in ministries and institution and no clear legal mandate,
- No or limited access to data, when ownership of data in other ministries or governmental organizations. Discussions also determined by position of responsible in institutions or interests and scientific background,

The majority of participants did not see substantial constraints from scientific / technical point of view. But...
- Financial constraints lead to scientific constraints (loss of experience and trained staff),
- Vegetation mapping (land cover, forestry, crops) need in vegetation season imagery, while topographers prefer off-season (out of vegetation period) imagery,
Ways to improve the synchronisation of data gathering

- Continuous bad weather conditions may block land monitoring by RS.
- The previous two problems are not similarly true for all climatic zones. Different geographical regions therefore allow and require different synchronisation strategies.
- With the increasing need of data exchange, harmonization and collaboration for integrated analysis new scientific challenges arise.

National examples
While idea of synchronization is easily accepted by technical and scientific people, administrative and financial obstacles hinder actual synchronization process. An example is that of a national orthophoto campaign, which is performed by federal states without practically any synchronization. Federal survey cycles (1-3 yrs) are dependent on available budget and human resources. It is therefore meaningless to order them to collect data if they cannot process them before data gets outdated.

In other examples scientific/practical reasons stand against synchronization efforts. Ortho acquisition cannot be synchronized among specific users, because topographers need data from the leafless period, while foresters/LPIS want data from vegetation period. This problem does not typical in Mediterranean area due to climatic conditions. In Scandinavia, data gathering is done in short cloud-free time windows (timing is weather-driven, not policy requirement-driven). Different geographical regions therefore allow and require different synchronisation strategies.

Other examples:
- In Switzerland, the main constraint for the synchronization of data gathering is legal: if politicians push, the rest (especially organizational and financial aspects) will follow. In fact, CLC and GMES data are not well accepted by the authorities as they are not enough accurate. It must be why the legal aspect seems to be the main obstacle to increase the periodicity of data gathering. Moreover as Switzerland is a small country, it is easy to make more accurate databases at federal levels. The technical approaches depend on the scale, thus the issues at regional or national levels differ. On second thought, the scientific aspect could be also a constraint. The way to fill the gaps between different levels (international, national and regional) must be defined probably by EEA.
- In Bosnia, CLC is the only land cover database. Bosnia doesn’t have much experience in land monitoring activities, mainly because of a lack of funding. Bosnia is also faced to organizational issues: regions are working on CLC but a national authority is missing to homogenize techniques and data. Sometimes a top-down approach should be recommended to make the organization easier through the country.
- Portugal is faced to the same kind of problems: it needs horizontal harmonization through the country before thinking of harmonization at the European level.
- Organizational issues in Germany: Germany’s orthophoto campaign, performed by federal states without practically any synchronization. Federal survey cycles (1-3 yrs) are dependent on available budget and human resources (it is therefore meaningless to order them to collect data if they cannot process them before data gets outdated).
- The organization is different in Wallonia and Flanders (Belgium). Unlike Flanders, Wallonia doesn’t have a central agency that merges all the small agencies’ outcomes. The lack of a unique contact in Wallonia makes the communication harder between the two regions. (Ex: a company flew almost all Belgium for a
Ways to improve the synchronisation of data gathering

Walloon agency but it was not allowed to sell the images to FGIA. FGIA could not have access to those images neither through the company nor through the agency and they had to find another way to get orthophoto.)

5. Proposals

Based on experiences summarized in previous chapters following three strategies is proposed for facilitating synchronization:

1. An obligation / proposal from EU side on update frequencies with fixed initial dates would be useful (e.g. 6 years for more complex inventories, such as CLC; 3 years for thematically simple but fast executable datasets, such as HRLs). Once such cycles are fixed, any activity (EU, national, sub-national) could voluntarily adapt. If this is done by prolonging one cycle until the initiative is synchronised this would not be difficult to achieve. Complete annual Sentinel Satellite coverage could support such activities as an encouragement to engage / synchronize.

2. Continuous maintenance of maps, including change logging, which provides data for answering monitoring questions (monitoring=change detection). Snapshots can be created any time when required. Topographical datasets are typically maintained this way.

3. Continuous but cyclic monitoring programs (e.g. 5 year cycles with 20% of the observations updated each year) – typical of orthophoto campaigns for LPIS. Disadvantage: one cannot create a single-date snapshot. Advantage: no need for large staff for short period, but doable with continuous activity of low staff.

Other useful short ideas/proposals:

- Define reference data for geometry (e.g. topographic map objects or standard European LAEA grid), which can then be filled up with data from different sources.
- By using “time stamp” for data (age of pixel/object) even temporally heterogeneous databases can be handled by users. Do we actually need to synchronize everything or should we learn how to integrate temporal, spatial and thematic heterogeneous classes?
- LPIS as one of the datasets with the highest update frequency and with wide coverage could be used to update other datasets (can also be used in validation or as geometric reference). Legal constraints should be solved.
- Frequency needs to be higher on the regional level, since management of land is there. On the European level only statistics are needed.
Data models and Nomenclature

Enhancing the European Land Monitoring System – Collection of Criteria for a Future Data Model
**Authors:** Stephan Arnold, Barbara Kosztra

**Commentators:** Geoff Smith, Geir-Harald Strand, Gerard Hazeu, Pavel Milenov

**Major contributors:** Geoff Smith, Linda Aune-Lundberg, Jenica Hanganu, Kolbeinn Árnason, Herbert Haubold, Christoph Perger, Marc Zebisch, Alexandra Fonseca, Michael Bock, Jo Van Valckenborgh, Stijn Van der Linden

**Further contributors:** This document contains contributions by all participants of the entire HELM consortium.

---

This report and the other outcomes of HELM can be downloaded at the project website www.FP7HELM.eu
Content

1. Introduction ........................................................................................................................................ 4
   1.1. Scope ........................................................................................................................................... 4
   1.2. Preface .......................................................................................................................................... 4
   1.3. Background and Motivation ......................................................................................................... 4

2. Stakeholders, Reporting obligations, User needs ........................................................................... 5
   2.1. European stakeholders .................................................................................................................. 6
   2.2. Reporting obligations .................................................................................................................... 7
   2.3. User needs and other driving factors ............................................................................................ 7

3. Given situation .................................................................................................................................... 8
   3.1. General statements ........................................................................................................................ 8
   3.2. CLC specific aspects ...................................................................................................................... 9
       3.2.1. Geographical aspect ............................................................................................................ 9
       3.2.2. Contemporary aspect ......................................................................................................... 9
       3.2.3. Status aspect ..................................................................................................................... 11
       3.2.4. Scale aspect ...................................................................................................................... 11
       3.2.5. Land use / Land cover conflicts in change mapping ............................................................ 12
       3.2.6. Semantic inconsistencies, overlaps and gaps in class definitions ................................... 12
       3.2.7. Country-specific discrepancies in interpretation of CLC classes ....................................... 13
       3.2.8. Sectoral land information missing from CLC ..................................................................... 13
       3.2.9. Summary of CLC analysis .................................................................................................. 14

4. Collection of criteria for the new data model .................................................................................. 14
   4.1. Structuring the criteria .................................................................................................................. 14
   4.2. Criteria in detail: The new data model should ........................................................................... 15
       4.2.1. Conceptual aspects criteria ............................................................................................... 15
       4.2.2. Thematic content criteria .................................................................................................. 16

5. Outcomes and proposal – the data model ....................................................................................... 20
   5.1. Structure and content of the data model/matrix ......................................................................... 20
   5.2. How to use the EAGLE concept ................................................................................................. 21
   5.3. Assets and benefits of new concept ............................................................................................. 22

6. Follow-up proposals ....................................................................................................................... 23
   6.1. Paradigm shift in data modeling .................................................................................................. 23
   6.2. SuperCORINE ............................................................................................................................ 23
   6.3. Methodological harmonization .................................................................................................. 24
   6.4. Standardization ........................................................................................................................... 25
   6.5. Legal framework for implementation ......................................................................................... 25

7. Conclusion/ Key messages ............................................................................................................... 25

8. References and Literature ............................................................................................................... 26

9. Acknowledgements ........................................................................................................................... 27

10. Annex – Who is EAGLE .................................................................................................................. 28
1. Introduction

1.1. Scope
Task 4.3 focuses on the collection of criteria and boundary conditions for a potential European land monitoring system / data model and should result in a draft proposal consistent with the findings of the HELM project and the EAGLE group (EIONET Action Group on Land Monitoring in Europe). Further, this task report outlines the potential benefits of such a proposed data model and gives a general outlook on its future usage. However, this task does not claim to tackle the feasibility of implementation and related political or technical obstacles which shall be discussed separately.

1.2. Preface
Since the HELM project was set up until the task 4.3 was opened according to the project’s time plan, significant progress has been made regarding the development of a draft data model by the EAGLE group. As this was already one step ahead of the sole working out of basic criteria, the title of this task was modified from “Work out basic criteria for a Europe-wide nomenclature / data model” to “Enhancing the European Land Monitoring System - collection of criteria for a future data model”. Also, approved by REA\(^1\) and the project lead EAA\(^2\), the task sheet topics were modified to cover the required criteria for a land monitoring system in a broader and more generic view. The results that had been achieved so far were thus incorporated into the HELM task 4.3.

1.3. Background and Motivation
Information on land cover (LC) and land use (LU) is crucial for any kind of land monitoring activity in a wide range of thematic fields of work (environmental monitoring: nature protection; spatial planning: soil sealing control; agriculture: crop yield estimations; emergency management: natural hazards zones etc.). Besides mapping and observing directly the Earth’s surface (bio)physical condition, status and change, LC / LU data is also important for other monitoring systems that use it as a vital input factor for their data models (climate change: biomass and carbon cycle; renewable energy: location of windmills etc.). LC / LU data is required and used on different scales and for manifold decision making processes on different levels (regional, national, European, global) and with several institutional backgrounds from the individual citizen through municipal government, national administrative or research institutions to European policy makers.

In a globalized world science shows that many natural phenomena and human behavior are connected spatially with each other. For a well-founded decision making process it is necessary to take these factors into account on a broader thematic view and geographical scale. In consequence this demands the interlinking of different sources of information coming from different fields of work and geographical regions. The exchange of specific information between these different domains is essential for solving present-day and future land monitoring challenges. For the majority of land monitoring initiatives information on both LC and LU are important. Therefore most of the existing classification systems contain a pragmatic mixture of LC and LU information. In addition, each given application may emphasize different aspects of either LC or LU, related to their specific requirements.

The long list of diverse applications of LC and LU information has over time led to the existence of many classification systems, nomenclatures and product specifications in the field of land monitoring. Different

\(^1\) REA – Research executive agency (of European Commission)
\(^2\) EAA – Environmental Agency of Austria
data collection methods, different scales, narrow tailored-to-purpose definitions and the lack of completeness for either LC or LU information make the straight transfer of data from one application to another difficult, if not impossible.

The field of land monitoring in Europe still falls short of exploiting the full potential of both the capabilities of the current practitioners and the quality of input information available on different levels. A lack of horizontal (intra-institutional and between national levels) and vertical harmonization between national and European land monitoring schemes leads to inefficient use of resources and to mismatching of products. Also, thematically overlapping user requirements for the same kind of data on different institutional levels in combination with restricted access to that specific information can lead to redundant data production. At the same time, the emergence of new technologies, information resources and the need for more detailed environmental information are challenging the current European system, which is often unable to give timely response to these challenges due to its complexity and inflexibility. Recognizing the improvable situation of land monitoring in Europe from the conceptual point of view has led to the self-initiative founding of the EAGLE group.

Responding to the growing need for higher spatial resolution and higher thematic content of LC / LU data, some European countries have started producing the required products through national initiatives like the examples in AT, UK, NL, DE, ES, HU, NO.

Currently the responsible institutions related to land monitoring are facing the pressure of cost efficiency and the necessity for better consistency between national and European data sets. At the same time they find themselves struggling with technical and semantic limitations of existing European standards (like CORINE Land Cover - CLC) on the other hand. In response to that situation an increasing number of European countries try to meet European community data requirements by developing and/or enhancing their own methods of mapping and data collection in a bottom-up way with the intention to avoid redundant data production.

The information flow generated by these national developments now needs to be integrated with other European “top-down” land monitoring activities such as the Copernicus / GMES GIO Land, and also other European initiatives like the statistical LC / LU field survey LUCAS. All these European activities can benefit in financial, technical and performance terms from national bottom-up contributions.

2. Stakeholders, Reporting obligations, User needs

In Europe, the “end users” such as political decision makers, environmental analysts, spatial planners, public institutions on sub-national / national / European level and citizens require information on the environment. The land monitoring community has to provide these end users with data on the following general topics:
- Status of environment (maps and statistics showing the location, distribution and proportion of LC / LU);
- Changes in the status of environment (statistics, objects, parameters showing differences in LC / LU over time).

4 LUCAS – Land Use & Cover Area Frame Survey
2.1. **European stakeholders**

On the European level several main stakeholders, among them several Directorate-General (DG) of the European Commission, have been identified who have a role as users of LC / LU information. Each of them has their own user needs according to their specific field of work:

**European Environment Agency EEA**\(^{5}\): Beyond the current CLC-based applications the EEA has some requirements for future land monitoring services such as the monitoring of green linear landscape features, biofuel crops, riparian zones, urban sprawl, pressures on Natura2000 habitats and natural hazards among others.

**DG Eurostat**: With the Land Use and Cover Area Frame Survey (LUCAS) the European Office for Statistics collects information on LC and LU on a point-based sample which is oriented on a regular 2x2 km grid. It aims at comprehensive collection of all kinds of LC/LU, but with a more detailed focus on crop types. Within the frame of LUCAS, the requirements of other DGs have been collected by Eurostat to inform the continuation and improved design of the survey. The consultation of EU institutions (EEA, DGs) resulted in the clear need for high quality data that can be aggregated in statistical administrative units (NUTS regions), but also in-situ micro data is required.

**DG JRC**: The Joint Research Centre of the European Commission is active in a very broad field of scientific thematic work. Within the JRC the Institute for Environment and Sustainability (IES) has a strong relation to LC/LU data. The IES carries out research to understand the complex interactions between human activity and the physical environment, and how to manage strategic resources (water, land, forests, food, minerals, etc.) in a more sustainable manner.

**DG ENV**: The objective of the Environment Directorate-General is to protect, preserve and improve the environment for present and future generations. To achieve this it proposes policies that ensure a high level of environmental protection in the European Union and that preserve the quality of life of EU citizens. Important themes are natural protection, air quality, water pollution or habitat monitoring. The DG ENV makes sure that Member States correctly apply EU environmental law. In certain cases DG Environment represents the European Union in environmental matters at international meetings such as the United Nations Convention on Biodiversity.

**DG AGRI**: The Directorate-General for Agriculture and Rural Development is responsible for the implementation of agriculture and rural development policy to promote the sustainable development of Europe’s agriculture and to ensure the well-being of its rural areas. It deals with all aspects of the Common Agricultural Policy (CAP) including farm support, market measures, rural development policy, quality policy, financial and legal matters, analysis and evaluation as well as international relations relating to agriculture. Especially relevant for the land monitoring community and connected to the CAP is the Integrated Administrative Control System (IACS), that lies in the responsibility of each member state.

**DG CLIMA**: The main target of Climate Directorate-General is to monitor the process of climate change and imply appropriate measures to constrain the long-term effects of it. It leads international negotiations on climate, helps the EU to deal with the consequences of climate change and to meet its targets for 2020, as well as develops and implements the EU Emissions Trading System. The CO2 and carbon cycle reporting obligations to Kyoto protocol and its subsequent agreements.

---

\(^{5}\) EEA – European Environmental Agency
The above descriptions of DGs main targets and responsibilities are taken from DGs websites. A more detailed listing explanation of European stakeholder requirements can be found in the report of task 5.1 “European actors needs”.

2.2. **Reporting obligations**

The requirements for LC / LU information are driven by political obligations coming from European legislation or other international agreements. The below is provided to give examples without claiming to be complete:

- **The Habitat Directive (92/43/EEC):** The Habitats Directive (together with the Birds Directive) forms the cornerstone of Europe's nature conservation policy. It is built around two pillars: the Natura 2000 network of protected sites and the strict system of species protection. All in all the directive protects over 1,000 animals and plant species and over 200 so called "habitat types" (e.g. special types of forests, meadows, wetlands, etc.), which are of European importance.

- **The Water Framework Directive (2000/60/EC):** The purpose of this Directive is to establish a framework for the protection of inland surface waters, transitional waters, coastal waters and groundwater which (a) protects the status of aquatic ecosystems and terrestrial ecosystems and wetlands directly depending on the aquatic ecosystems; (d) ensures the progressive reduction of pollution of groundwater and prevents its further pollution, and (e) contributes to mitigating the effects of floods and droughts. For being capable of doing such monitoring, detailed information on land use in river catchment areas is required.

- **The Floods Directive (2007/60/EC):** Its aim is to reduce and manage the risks that floods pose to human health, the environment, cultural heritage and economic activity. The Directive requires Member States to carry out an assessment to identify the river basins and associated coastal areas at risk of flooding. For such zones they need to draw up flood risk maps by 2013 and establish flood risk management plans focused on prevention, protection and preparedness by 2015. The Directive applies to inland waters as well as all coastal waters across the whole territory of the EU.

- **Common Agricultural Policy (CAP14) that requires 5% of the agricultural area as Ecological Focus Areas in connection with the Integrated Administrative Control System (IACS - European agricultural subsidy control system).**

- **Carbon accounting and reporting obligation in Kyoto and post-Kyoto activities.**

The above descriptions (partly shortened) of Directives main targets are taken from European Commissions websites.

2.3. **User needs and other driving factors**

Besides the above mentioned legally binding reporting obligations there are further drivers that induce certain needs of both data users and providers, which stem from national or international initiatives and applications:

- **INSPIRE Directive (Infrastructure for spatial information in the EU)**\(^6\). The overall aim is to create common technical standards for the web-based exchange of spatial data among data providers and users on regional, national and European level. In specific thematic working groups and conceptual committees data specifications have been elaborated with the target of data interoperability and

---

\(^6\) INSPIRE website URL: http://inspire.jrc.ec.europa.eu
EU member states are not requested to set up any new surveying or mapping initiatives, the directive only demands that relevant and already existing spatial data should be made available according and compliant to the data specifications for each theme.

- Eurostat’s Land Use & Cover Areas frame statistical survey (LUCAS). (see also Eurostat as «stakeholder» under chapter 2.1).
- System of Environmental-Economic Accounts (SEEA).
- Streamlining European Biodiversity Indicators (SEBI): The pan-European SEBI initiative was launched in 2005. Its aim was to develop a European set of 26 biodiversity indicators – based on those already existing, plus new indicators as necessary – to assess and inform about progress towards the 2010 targets. One of the principal working methods of SEBI is to build on current monitoring and available data to avoid duplication of efforts and to complement and not replace other activities to describe, model and understand biodiversity and the pressures upon it.
- EEA’s Climate change indicators (CLIM): With a set of 50 indicators over a broad thematic range all kinds of climate relevant influencing factors are collected, e.g. on air pollution, greenhouse gas emission, distribution of sea ice, forest fires, water surface temperature, and many more.

3. Given situation

3.1. General statements

As previously stated in the introduction, the multitude of uses and users of LU / LC data has led in general to development of many classification systems, nomenclatures and mapping initiatives at regional, national and European levels. Exploring the current situation in search of candidates for a European data model, existing and available classification systems can be analyzed by different aspects of their specification: the scale / minimum mapping unit, the thematic detail (the subdivision of classes), thematic completeness (comprehensive list of classes for all kinds of LC and LU), and the thoroughness of class definitions (whether a single class definition includes exhaustive list of different classifiers, parameters or criteria).

Based on a first overview, none of the European level classification schemes currently available appear to follow the specification required for a continent-wide harmonized land monitoring. In most cases a given nomenclature has a special focus on a particular aspect, for example broad differentiation between several vegetation types (LCCS), another one has its focus on urban distinction of soil sealing degrees (Urban Atlas), a third one emphasizes the categorization of landscape into different habitat types or ecosystems (EUNIS), or any of those aspects are present in parts (CLC), but never elaborated enough to completely cover all landscape properties in a single nomenclature.

This task focuses on general purpose inventories, therefore does not examine inventories of limited thematic field (e.g. specific forest inventories) or of limited geographical coverage (e.g. national inventories). CORINE Land Cover (CLC) as the major general-purpose land monitoring inventory with full European geographical coverage and with a significant timeline was considered to be the existing “quasi-standard” for Europe. Critical analysis therefore focuses on CLC (also taking into consideration that this is the inventory most HELM participants have collected experience with).
3.2. CLC specific aspects

The CLC concept and nomenclature was designed in the 1980’s for a visual photointerpretation method, with Central- and Southern-Europe as the main geographical focus. Its extension to become a pan-European database happened without a fundamental conceptual revision of nomenclature. The reason behind keeping it unchanged was the need to maintain backward comparability (heritage) with previous products. Meanwhile, as the CLC concept has been implemented in an increasing number of 39 countries over a 20-year time span, the need for revision of the concept has become evident both for participants and stakeholders to adapt to changing technical circumstances and policy requirements.

This chapter aims to list some of the shortcomings of CLC that should be tackled by a new data model. It must be noted, that some of these shortcomings emerge from circumstances (lack of technological possibilities) and user needs that did not exist by the time of creation of CLC nomenclature, therefore have not been considered by its developers.

3.2.1. Geographical aspect

The CLC nomenclature was originally designed with a limited geographical focus. Landscapes, vegetation types and human activities typical of other areas were therefore not sufficiently represented in terms of thematic content and detail. Such insufficiently represented features are for example:

- Types of moors and heathland (e.g. areas covered by mosses and lichens)
- Types of peatland (vegetation types, cultivated / non cultivated)
- Scattered artificial features (holiday cottage areas, geothermal plants)
- Salt lakes and inland salines (CLC maps salines in a maritime context)
- Salt planes lakes with periodic or episodic water regime, which are used for agricultural production when water withdraws / evaporates ("poljes", typical karstic phenomena of the Dinaric Alps)

This leads not only to the missing of important landscape information, but also to large, heterogeneous areas being mapped as a single class, the most typical example of which found in North Scandinavia. The CLC nomenclature fits Central- and especially Southern-Europe much better than Northern-Europe. Vast and even significantly heterogeneous areas in Norway and Iceland have to be put in one and the same class because appropriate alternatives are missing or the selection of subdivisions based on the existing CLC classes would be meaningless. In Iceland class 322 includes large moss and lichens covered areas and comprises 35% of the country. In Norway large part of the country’s northernmost area is covered by three classes: moors and heathland (322), sparsely vegetated areas (333), peat bogs (412), while in reality a great diversity of distinct habitats characterize the area.

3.2.2. Contemporary aspect

First CLC inventory (called CLC1990 after its average image acquisition date) started in the mid 1980’s and production was spread over a decade. Since then three updates (CLC2000, CLC2006, CLC2012) have been produced with the corresponding change mapping activities covering almost 30 years. During this time not only the geographical coverage or the technical background of the mapping have changed, but also several new practices, techniques and purposes of landscape use have emerged. Changing climate, together with changing political, economic and social environment has led to a rise of previously unseen or much less frequently practiced land uses or measures, which therefore are not tackled by the 30-year-old CLC...
nomenclature. Some of these new practices can fit into existing CLC classes, some would require new classes or new ways of defining the classes to be mapped.

- Landscape elements connected to renewable energy production, such as wind farms or solar panel parks, geothermal power plants can be fit into class 121 (industry and commerce), but due to geometric requirements of CLC (25 ha MMU and 100 m MMW) the scattered spatial pattern of wind turbines with the service roads in between or the pipelines of geothermal plants can be mapped only by either significantly exaggerating the class area or omitting most of the features.

- Similar small surface features characterize fracking (hydraulic fractioning for shale gas extraction) activity. This technology is becoming more widely practiced also in Europe due to development of mining technology and general rise of energy prices. The technology’s impact on environment is of much concern, its extent and depth still not fully explored by researchers. Individual fracking sites are too small and spatially too scattered to be mapped by CLC, still the information on their presence would be needed to include in a general-purpose environmental database.

- Hunger for energy, as well as geopolitical considerations to be energy self-sustainable in Europe are leading to increase of land occupancy of energy / biofuel plantations. Short-cutting-cycle tree / shrub plantations, energy grass, and arable crops for bio-fuel or bio-ethanol production belong to this group. First two cases are hard to fit into CLC classes: short rotation cycle in case of tree plantations and amount of biomass in case of energy grasses make them atypical variants of relevant CLC classes (311 and 231, respectively). Arable biofuel crops can be easily fitted into class 211, but purpose of production is then not identifiable any more, although this information is of much importance when monitoring competition of alimentary crops and energy crops for agricultural land.

- Earlier beginning of snow melt and higher average temperature (indicators for climate change) in the Alpine territories have a shortening effect on the skiing season. In order to keep tourism industry going or also extend its intensity, artificial snow making (use of snow cannons) has become a more and more widespread practice to prolong the skiing season. Visible signs of this practice in landscape are water storage pools, leveling of ski slopes topography and forest aisles. A longer snow cover and use of chemicals in artificial snow (to increase melting point temperature) severely affects natural vegetation cover and shortens the vegetation growing period. In CLC nomenclature ski slopes are not mapped as part of sports facilities (142), but their environmental impact is reasonable enough to map artificial snow areas as part of 142 and to introduce them as particularity in 142 class definition.

- Restoration ecology, which emerged as a separate field in ecology in the 1980s, is the practice of renewing and restoring degraded, damaged, or destroyed ecosystems and habitats by active human intervention and action. In the first CLC inventories only isolated examples of this practice were found, but during the last decades the practice of reconstructing habitat has become more widespread all over Europe. The first step of this process often appears similar to regular construction activities (removal of topsoil, earthworks, piled up construction material) and is mapped as construction site (133) in CLC, then as change from construction site to any natural vegetation (e.g. inland wetland – 411) during change mapping. Although covered by the same CLC class (133) the construction of a factory has very different environmental impact in the long term than the construction of a wetland, which is of special interest when calculating indicators and land accounts from change maps. Therefore it would be beneficial to separate these two construction types, depending on expected outcome.
The dynamic of the appearance of these features is a useful indicator of environmental changes and society’s reaction on them, which implies that they have to be somehow tackled in a general European data model.

### 3.2.3. Status aspect

CLC nomenclature includes a number of classes that characterize temporal status of landscape elements: an intermediate stage during a one-directional change process (like natural succession), or regularly altering phenomena (like tide).

- **Class transitional woodland-shrub (324)** maps areas that are at any stage in the process of forest development, including areas with very much varied land cover (bare soil, sparse vegetation, grass, shrubs, trees) and without separating natural (succession) or anthropogenic processes (afforestation / reforestation). To handle this heterogeneity, a number of countries applied level-4 subclasses of 324 for national purposes to separate the abovementioned aspects. On the other hand, the land cover or appearance (shrub cover) are often identical to stable, climax-stage vegetation types mapped by CLC classes sclerophyllous vegetation (323) and moors and heathland (322), which makes these classes ambiguous, difficult to separate and not straightforward to use in land monitoring analyses. Stable situations in landscape should not be put together with transitional classes even though the momentary stage is comparable.

- **Other classes** aim to capture short-term phenomena that are temporal characteristics of the area. Class burnt areas (334) does not give information on what type of vegetation was affected by fire, on the other hand it is inconsistent by including burnt forests, but not burnt fruit orchards and including recent burns, but not still visible older ones. Temporal aspects of water are even less expressed in CLC classes. The only water-related temporal phenomenon handled by CLC is the tide, by mapping intertidal flats (423) as a separate class, expressing by definition the alternation of water and bare sand/mud. For most water surfaces CLC applies the majority rule, attributing an area to rivers (511) or lakes (512) if it is most of the time (within a year) covered by water. Periodic, seasonal or episodic changes in water cover / regime and the land covers (e.g. gravel beds, grass) alternating with water therefore remain unmapped in CLC. Due to mixed LC / LU character of CLC, it is even harder to map areas where not only LC, but also LU changes together with water cover extent. Temporal lakes with periodic or episodic water regime, that are used for agricultural production when water withdraws / evaporates has to be mapped either as lake or as agriculture, there is no way to represent all important pieces of information on landscape.

User critiques show a need for these temporal phenomena expressed in form of an attribute of status attached to LC / LU unit.

### 3.2.4. Scale aspect

If the rule of a geometric minimum mapping unit is put aside, a number of CLC classes are applied on different scales within the same classification system. Especially the complex and mixed classes like 241, 242, 243, 244 (complex cultivation patterns) by definition need to have a minimum extent of landscape type to be identified as such. Other classes like 31X (forests) can be identified already based on a much smaller extent, it is however difficult to later separate mixed forests on a stand level (which would be considered mixed also on a much smaller scale) and those mapped as mixed only as a result of CLC’s large MMU (which are made up of <25 ha patches of homogeneous broad-leaved and coniferous patches).
3.2.5. **Land use / Land cover conflicts in change mapping**

Insufficient separation of LU and LC in CLC nomenclature leads to conflicts of understanding especially during change mapping in cases when LC changes, but LU does not, or the other way around. For example:

- A dumpsite that has been shut down and covered by soil and/or grass, but its only function (LU) is still dump storage (being monitored because of dangerous materials) is still dumpsite as LU, but LC has changed to grass. After a few years/decades when no more harmful material/gases are released it is turned into a sport and recreation area (new LU), without the LC being changed.

- A large fenced area covered mostly by natural vegetation and previously used as buffer zone of a military airport is stopped to be used for military purposes. There is no change in LC, still according to CLC rules it should be mapped as change from airport (124) to moors and heathland (322), which is otherwise considered to be a logically excluded (“impossible”) change type.

3.2.6. **Semantic inconsistencies, overlaps and gaps in class definitions**

Uncertainties of understanding, interpreting and separating CLC classes in a number of cases originate from semantic conflicts inherent to the nomenclature’s class descriptions.

**Grasslands (231/321)**

Most users find it difficult to discriminate between class pastures (231) and natural grasslands (321), both during photointerpretation and in modeling. The latter highlights that current class definitions, which mostly focus on botanical composition and less on origin (primary/secondary grasslands) and management/LU practices are ambiguous and not detailed enough. A need for thematically more detailed mapping in some countries is also expressed (by e.g. Austria, Romania), leading to applications of level 4 and 5 subdivision of grassland classes in national CLC nomenclatures (e.g. Hungary, Ireland). A gap in class definitions is also identified. Ruderal, low-quality grass-covered areas (e.g. those on city edges, in between construction areas waiting to be used) are not covered by any of the existing classes. In practice they are usually mapped as 231, but this is normally out of necessity.

**Shrubland (322 / 323 / 324)**

The second group of classes that proved to cause difficulties due to improper and overlaps in definitions is that of shrubs (322, 323, 324). Naturally, this is mostly a problem in areas where any two of them occur in close geographical vicinity. Also, a class seems to be missing for climax stage shrubby vegetation in continental climate Central-Europe (e.g. low shrubs on karstic areas).

**Differentiation between 322 / 333 / 412**

Overlap between definitions of 322 and 333 has been identified when border-matching CLC maps of Norway, Finland and Sweden (more detailed description in task 2.4 report). A revision of these class definitions, in consultation with representatives of Nordic countries would be necessary. Also, the differentiation between 321 / 322 and 412 can often only be made reliably if there is knowledge of the peat depth as the same vegetation of 321 and 322 can occur without significant peat.

**Differentiation based on water salt content**

Information on water salt content (added e.g. as an attribute to CLC polygons) would help to close gaps of definitions that currently leads to non-consistent mapping of inland (non-coastal) salt lakes, salt marshes and salines. For example, if in a coastal location salines are mapped as 422, first two digits referring to coastal character, while inland salines are mapped as mineral extraction sites (131). Salt marshes are also defined by their coastal character, besides that by floristic composition resulting from salt water influence. Coastal wetlands of Estonia however, do not show this floristic composition, simply because salt content of the Baltic sea is too low for such habitats to develop, therefore are hard to fit into existing definitions.

**Built-up areas (111 / 112)**
Separation of urban classes 111 and 112 is not always straightforward. Soil sealing is not easy to judge, while in class definitions that is the key of separation. (The class differentiation could be improved by the use of the imperviousness HR layer results combined with LU information.)

Natural mosaics

While a number of classes in CLC try to tackle areas with mixed agricultural use (in different scales: among and within parcels), there is no way to deal with areas that are mosaics of well defined homogenous patches of different semi-natural classes. For example, in the Balkan region as well as in the Mediterranean, huge areas are covered by a mosaic of sclerophyllous shrub, bare rock, trees and herbaceous vegetation, which do not fit into any single class definition. Description by percentages of LC components would be the solution for a proper representation of these landscapes.

Overlaps and gaps due insufficient description of LU information in class definitions

Class definitions often fail to describe LU aspects of the class, defining them mostly by LC parameters and the presence of functional landscape elements. This leads to difficulties when trying to identify a class by modeling it with information elements taken from class definitions. For example, a village area with 50% tree cover and 50% soil sealing can correctly be mapped as forest (>30% crown cover) or as urban fabric (>30% built-up) following the CLC definitions. Information on LU, i.e. residential as the key defining element for separation is missing from class descriptions.

3.2.7. Country-specific discrepancies in interpretation of CLC classes

Due to national situations and geographical variations, some CLC classes are understood and applied in a heterogeneous way across Europe, being known and accepted by users of CLC data. Most significant and best known of these are forest classes (31x). To keep consistency with national data, the application of national forest definitions different from CLC class definition is accepted by EEA in CLC mapping. In Norway and Iceland forested areas with crown cover density as low as 10% (instead of the prescribed 30%) are mapped as 31x. Also tree height of 2 meters is used in definition of forests in Iceland.

Users are less aware of the discrepancy among countries in applying classes 231 and 321. Those countries where natural grasslands are abundant (typically Nordic countries), tend to be more strict on applying class 231, intending to make sure that what is mapped as 231 is really used for agricultural purpose. Countries with more anthropogenic landscape dominated by agriculture (typically Central-Europe) were more careful on using class 321, trying to make sure that what is assigned to this class are really natural, high-ecological value grasslands. Some countries use a floristic definition of grassland types (e.g. the UK Broad Habitats) compared to the standard CLC definition where 231 is used for all grasslands with apparent intensive use (presence of field boundaries).

3.2.8. Sectoral land information missing from CLC

In general, consistent information on intensity of use (management practices, like irrigation, grazing, fertilization, forest cutting cycles) is missing from (most) CLC classes.

Irrigation: Irrigation might be an indication of use intensity and as such, key information for differentiating between specific CLC classes (211/212, 231/321). On the other hand, irrigation of areas other than arable land is not mapped by CLC, while from general environmental information point and for national purposes this information would be a great importance. In some Mediterranean countries, information on irrigation of fruit plantations and vineyards is missed by users (solved by introducing level-4 subclasses for national uses). Many users would prefer to have information of irrigation added as an attribute to CLC polygons, not as a (much more restrictive) class.

Grazing: Grazing, besides being also an intensity indicator and CLC key class selection information, can also deliver valuable information on human activity on non-agricultural landscapes. In Iceland for example no
clear border lines can be drawn between agricultural areas and areas of (semi)-natural vegetation. Grazing is not limited to fenced parcels but comprises most of the vegetated areas in the country. In the UK grazing occurs on unenclosed uplands (321 / 322) and stocking densities increase with improvements and the appearance of obvious field boundaries toward pasture (231), but this is a continuum and there are no discrete divisions. Therefore information on grazing should not only be attached to class 2xx polygons, but should be treated as a separate attribute, which can be applied to other classes, too.

Besides management practices, information on specific aspects of habitats, such as botanical composition, disturbance and origin, is found to be needed by users.

Grasslands: There is an expressed user need for the more sophisticated separation between various grassland types, not only regarding management practices, but also origin (primary / secondary, sown / self-sown), floral composition, slope type, wetness are characteristics that can be considered to be tackled as subtypes. These contributing factors were suggested as components of grassland HRL, but have yet to be implemented.

Wetlands: Wetland differentiation is found to be too simple. There are many other types of wetlands than can be adequately described by inland marshes (411) or peat bogs (412).

Water bodies: The need for subdivision of lakes (512) to at least natural and artificial lakes is often raised by national users.

3.2.9. Summary of CLC analysis
Analyzing the CLC experience as potential candidate for a pan-European land monitoring data model/nomenclature revealed the following most important shortcomings of the CLC nomenclature:

- No equal representation of landscapes in different bio-geographical regions.
- Classes representing mixture of land cover and land use aspects
- The temporal aspect (transitional phenomena, status) is handled only selectively
- Missing thematic content (types of grasslands and wetlands, cultivation practices among others).
- Inflexible classification system, missing option of attribution of spatial units or handling of new appearing landscape phenomena.

In the next step of collecting criteria, these shortcomings are listed among issues to be tackled by a new land monitoring data model.

4. Collection of criteria for the new data model
This chapter contains the key part of the task 4.3. From the listed user needs, drivers and requirements mentioned above the next logical step is to deduce the criteria for a data model.

4.1. Structuring the criteria
The collection of criteria can be grouped according to different topics such as thematic fields of work, involved stakeholders and users, or geographical regions and climate zones. Further, the criteria for such a data model can be categorized by thematic content itself on one side and by conceptual aspects on the other side. In the following sections a numbered list shows the criteria that have been collected with some further short explanation where necessary.
4.2. **Criteria in detail: The new data model should...**

4.2.1. Conceptual aspects criteria

1. Be **object-oriented** and descriptive in terms of landscape abstraction and follow the paradigm shift from classification to characterization of landscape. (Still a certain level of taxonomy will be necessary for identification of objects and characters).

2. Enable **separation of LC and LU** information through separate capture and storage of this kind of information. Objects that indicate both LC and LU within themselves may need special handling.

   2.1. The LC part should aim at being **mutually exclusive and exhaustive**.

   2.2. The LU part should also aim at being capable of storing **any kind of land use** information.

   2.3. Be ready to identify the distinction between intended land use on one hand and (as it is recorded in cadastral data) and “real” recent land use (as it can be found in-situ or extracted from remote sensing data).

3. Be **scale independent** as far as possible and open for data storage on any level of detail / spatial resolution.

4. Be a vehicle for **semantic translation** between different existing nomenclatures and classification systems, taking into account **existing standards** and be capable of **decomposing** or semantically mapping their classes. Some prominent standards/classification systems to be considered are:

   1.1. CLC

   1.2. LUCAS

   1.3. INSPIRE: Take the thematic concept of abstractive description at least for data specs of Land Cover (LC) and Land Use (LU) into consideration. Be aware of cross connection to further neighboring INSPIRE themes such as Buildings (BU), Hydrology (HY), Habitats (HB).

   1.4. LCML (get in contact with ISO / CEN people). LCML is used as a tool of description, it is not a nomenclature of some particular institution (besides FAO as main driver of standard)

   1.5. LPIS (at least in line with structure, also uses LCML and the classification rules of LCCS)

5. Allow two parallel streams of data encompassing a **bottom-up and top down approach** equally: Enable data users to follow the bottom-up approach of integrating national data into European initiatives & make European data also usable and useful for national purposes and applications.

6. Be **INSPIRE compliant** regarding technical specifications (besides thematic specifications) for data modeling (UML).

7. Be capable of modeling both

   7.1. **area features** as well as

   7.2. **point features**. (Integration of line features still under discussion, they can be also seen as long and stretched area features).
8. Have sufficient **thematic depth** (detailed subdivision) and the thematic **width** (thematic completeness) in the model to be capable of modeling all kinds of LC/LU relevant landscape objects and landscape types with the necessary detailed attribution.

9. Be capable of storing **parameterized information** such as soil sealing degree or crown cover density as attributes.

10. Be **operationally proofed** as a working concept.

11. Be applicable on both **national and European level**.

12. Be **extendible and flexible** enough for the integration of **new classes** or model **elements** that might gain importance at a later stage. Consider new LC / LU classes (like habitats / ecosystem types) that can express a more **complex landscape situation** and also open a broader applicability of the model in various fields of work.

13. Be ready to **populate** the spatial reference units of a system (polygons, points) with additional attributes.

14. Be capable of storing information on landscape **independently** from the data capture methods and from any **source of information** (be it remote sensing methods as well as in-situ data or other auxiliary data).

15. Follow in principle the **UML application schema** and modeling method as applied in ISO standard LCML or INSPIRE.

16. Have an **easy to understand** structure and a simplified version usable for non-UML experts and have an alternative documentary or modeling explanation like the EAGLE matrix besides the UML-chart.

17. Tackle following temporal aspects:

   17.1. Be capable to create **time series** data and **change maps**, ideally both in line with pan-European common reference years but also continuously and independent from them.

   17.2. Model **repetitive temporal aspects** like seasonal or daily alterations.

18. Be ready for mapping with **small minimum mapping units** (MMUs) based on very high spatial resolution imagery (~1m).

19. Ensure statistical consistency for given spatial units across nations and borders, e.g. NUTS regions.

20. Try to guarantee **backwards compatibility** in describing and modeling both old existing and new LC/LU information or classifications.

**4.2.2. Thematic content criteria**

21. Store **soil sealing degree**, which is important for hydrological modeling (surface water run-off, filling of ground water aquifers), urban climate, monitoring of urban sprawl, population density, etc.

22. Store **crown cover density**, which is important for forest monitoring, biomass calculation, habitat properties, carbon sequestration, etc.

23. Integrate **habitats and ecosystem types**, offer enough details to map characteristics of habitats that are needed for their monitoring.

24. The possibility to include **landscape features** like hedgerows or riparian zones, which usually are linear or elongate elements.

25. Handle indications on **natural hazards and risk zones** in general such as geological risk zones, flood risk, avalanche risk, etc.
26. Contain model elements to describe aquatic vegetation like e.g. floating forbs or algae. This is especially important for wetland habitat characterization and monitoring of its ecosystem function.

27. Handle information on the crop type. This is important for agricultural statistics on national and European levels. A commonly used data model and survey method can contribute to a better harmonization between national and European statistics.

28. Store information on irrigation methods. What kind of irrigation method is applied (none/rain fed, sprinkler, channels, flooding, carrousel), also the source of water (rain fall, ground water, river, reservoir).

29. Handle information on the functional status of a land unit like under construction, destroyed, abandoned, and on the physical status, such as burnt, wind-broken, insect damaged area etc.

30. Besides land use categories give further detailed information on the land management form. Especially in the field of agriculture relevant land management characteristics are grazing, mowing, fertilizing, irrigation, ploughing, shrub clearance etc.

31. Data on forest management shall give information on logging forms such as clear cuts, selective logging, short term rotation, long term rotation.

32. Spatial patterns are important to know when describing the diversity and complexity of a certain landscape type, like plantation patterns, stone walls, homogenous / heterogeneous / mosaic / mixed phenomena.

33. Be open for storing information on soil type. It can help to better model habitats (close connection between pedosphere and biosphere) or productivity of agricultural areas.

34. Information about snow cover can give a good indication on climatic changes in vertical (alpine) or horizontal (geographical distribution) dimension, and the duration of vegetation period can be monitored (also connected to artificial snow on ski pistes). Further it benefits the calculation of avalanche risk, melting water run-off.

35. Decide if generic information on ground water should be somehow included in LC modeling, like the height of the ground water level, or also seasonal variability of ground water level.

36. Avoid the expression “other” in the name of classes or model elements, but instead describe as precise as possible the content of the category. However, “other” categories are necessary when the ordering system (classification or characterization) does not cover any possible relevant property to be found in real landscape.

37. Be capable of modeling forest data according to the FAO forest classification (Eurostat requirement).

38. Handle upcoming renewable energy utilities like windmills, solar panels or geothermal installations.

39. Find a way to handle greenhouses both as buildings and as agricultural production facilities.

In addition to the textual bullet points, Table 1 shows a list on all collected criteria to bring them into relation with existing nomenclature and the EAGLE data model proposal. Criteria are tagged as “conceptual aspect” (ca) or “thematic content” (tc). The list of criteria does not claim completeness and is open to be extended if needed.

A cross table helps to evaluate some of the existing classification systems. Also the proposed draft EAGLE model has been included in the evaluation cross table, for indicating to what extent it would fulfill the criteria. With reference to the numbering above the listed criteria are checked against the properties of certain selected LC / LU classification schemas and nomenclatures and evaluated whether they are represented, partly/potentially represented or not at all represented by the given system.
Table 1: Criteria cross table assessing the capability of CLC, LUCAS, LCML, GlobCover, EAGLE data model and SuperCORINE (see chapter 6.2) to tackle recent and upcoming requirements by fulfilling the listed criteria.

<table>
<thead>
<tr>
<th>Nomenclatures</th>
<th>category</th>
<th>CLC</th>
<th>LUCAS</th>
<th>LCML</th>
<th>GlobCover</th>
<th>EAGLE</th>
<th>SuperCLC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>no.</strong></td>
<td><strong>Criteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>object-orientation</td>
<td>ca</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>separation of LC/LU</td>
<td>ca</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2a</td>
<td>mutually exclusive &amp; complete LC</td>
<td>ca</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2b</td>
<td>any kind of LU</td>
<td>ca</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2c</td>
<td>distinct intended LU versus real LU</td>
<td>ca</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>scale independent</td>
<td>ca</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>semantic translation</td>
<td>ca</td>
<td>N/A</td>
<td>N/A</td>
<td>1?</td>
<td>N/A</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>integrate bottom-up / top down</td>
<td>ca</td>
<td>1</td>
<td>1</td>
<td>2?</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>INSPIRE LC compliance</td>
<td>ca</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7a</td>
<td>area compatible</td>
<td>ca</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7b</td>
<td>point compatible</td>
<td>ca</td>
<td>0</td>
<td>2</td>
<td>?</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>thematic width and depth</td>
<td>ca</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2?</td>
</tr>
<tr>
<td>9</td>
<td>parameterization possible</td>
<td>ca</td>
<td>0</td>
<td>0</td>
<td>1?</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>operationally proved</td>
<td>ca</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>11a</td>
<td>applicable: national level</td>
<td>ca</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>11b</td>
<td>applicable: European level</td>
<td>ca</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>extendible / flexible for new classes/elements</td>
<td>ca</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>attribution of spatial units</td>
<td>ca</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>independence from source of information</td>
<td>ca</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>UML application schema</td>
<td>ca</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>no.</td>
<td>Criteria</td>
<td>category</td>
<td>CLC</td>
<td>LUCAS</td>
<td>LCML</td>
<td>GlobCover</td>
<td>EAGLE</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------------------------------------</td>
<td>----------</td>
<td>-----</td>
<td>-------</td>
<td>------</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td>16</td>
<td>easy to understand structure</td>
<td>ca</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>17a</td>
<td>time series, change maps</td>
<td>ca</td>
<td>2</td>
<td>0</td>
<td>2?</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>17b</td>
<td>repetitive temporal aspects</td>
<td>ca</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>small mapping units, high spatial resolution</td>
<td>ca</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>statistical consistency</td>
<td>ca</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>backwards compatibility</td>
<td>ca</td>
<td>N/A</td>
<td>N/A</td>
<td>2</td>
<td>N/A</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>soil sealing degree</td>
<td>tc</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>crown coverage</td>
<td>tc</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>habitat / ecosystem types</td>
<td>tc</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>include landscape features</td>
<td>tc</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>natural hazards and risk zones</td>
<td>tc</td>
<td>1</td>
<td>0</td>
<td>1(?)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>aquatic vegetation</td>
<td>tc</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>27</td>
<td>crop types</td>
<td>tc</td>
<td>1</td>
<td>2</td>
<td>?</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>28</td>
<td>irrigation methods</td>
<td>tc</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>29</td>
<td>functional status</td>
<td>tc</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>30</td>
<td>land management</td>
<td>tc</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>31</td>
<td>forest management</td>
<td>tc</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>32</td>
<td>spatial patterns</td>
<td>tc</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>33</td>
<td>soil type/soil organic content</td>
<td>tc</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>snow cover</td>
<td>tc</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>35</td>
<td>ground water</td>
<td>tc</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>36</td>
<td>avoid &quot;other&quot; category names</td>
<td>tc</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>37</td>
<td>FAO forest classification</td>
<td>tc</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>38</td>
<td>Renewable energy utilities</td>
<td>tc</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>39</td>
<td>Greenhouses</td>
<td>tc</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Task 4.3 Enhancing European Land Monitoring System – Criteria Collection for a Future Data Model
5. Outcomes and proposal – the data model

The above presented criteria have been collated against several existing data models and nomenclatures in the previous chapter. As a result it turns out that none of the considered classification systems fulfil these criteria sufficiently, particularly to serve for thematic harmonization. Instead, the EAGLE data model is capable of considering and actively referring to most criteria. The EAGLE data model is introduced briefly in the coming chapters.

5.1. Structure and content of the data model/matrix

The EAGLE concept comprises of two forms of representation: the data model (UML chart) and the matrix (Excel cross table).

The content of the matrix is parted into three blocks:
I. Land cover components (LCC) [e.g. buildings, sealed surfaces, trees, herbaceous plants, water bodies];
II. Land use attributes (LUA) [e.g. agriculture, forestry, residential, mining];
III. Further characteristics (CH) [e.g. cultivation measures, biophysical parameters, ecosystem types, status, spatial pattern, temporal pattern].

Both the matrix and the data model have the identical thematic content. The matrix contains all model elements (as table columns) in a hierarchically grouped order. The data model with its UML chart brings these model elements (LCCs, LUAs, CHs) into relation with each other. The compilation of the three descriptive matrix blocks (LCC, LUA and CH) is based on the reflection of existing land cover and land use specifications (i.e. INSPIRE, CLC, LUCAS, LCCS-FAO and national nomenclatures).

Figure 1 shows the simplified structure of the object-oriented EAGLE data model, consisting of three main segments of Land Cover Components (LCC) distinguished by their (bio-) physical appearance: Abiotic (non-vegetated) artificial and natural surfaces, Biotic (vegetated) surfaces and Water surfaces. Each segment contains a number of LCCs hierarchically ordered in sub-branches. The LCCs are further described with attached landscape characteristics (CH) that give more specific details about their properties.

Structure of EAGLE data model

Based on the selected LCCs and their attached characteristics, one or many (1...*) LCCs build a Land Cover Unit (LCU) that also can have by itself some specific characteristics. The Land Cover Unit is completed by the additional information of Land Use Attributes coming from the HILUCS (Hierarchical INSPIRE Land Use Classification System) classes, extended by some EAGLE-specific sub-types. Several LCCs can occur inside a

---

Harmonised European Land Monitoring

Task 4.3 Enhancing European Land Monitoring System – Criteria Collection for a Future Data Model
LCU, but cannot overlap, meaning they are mutually exclusive. LUA, however, can occur in an overlapping manner in the same part of land.

A real landscape example to be modeled can be a public open-air bath on a natural lake that is also partly-used as a sand extraction site. It has changing rooms (being renovated and extended) and administrative facilities. The lake is surrounded by a sandy shore and grass areas with groups of shrubs and trees. Opening time is only in summer season.

In terms of abstracted model elements it would be described as follows:

- **LCCs** – buildings, sand, graminaceous plants, woody vegetation, water surface
- **LUA** – recreational/sport & leisure (bathing area and lawn) and mining (sand extraction area)
- **CHs** – temporal duration for recreation purpose: March-October; temporal duration for mining: 12 months; status of buildings: under construction; spatial pattern for grass area: homogenous growth; cultivation measures for grass area: repetitive cutting; salinity of water: fresh water.

### 5.2. How to use the EAGLE concept

Based on a decomposing and object-oriented modeling approach, the EAGLE data model and matrix can be applied in a twofold way to describe and abstract a land surface unit (polygon, grid cell, linear feature or sample point) representing a given landscape situation.

1. It can be used as a tool for semantic comparison and translation between class definitions within one or between several different classification systems. Given abstract class definitions are analyzed and decomposed with the model elements as in a descriptive and diagnostic manner (without looking at real landscape scenario).

2. It can be used for mapping activities to describe a land surface unit (polygon, grid cell, linear feature or sample point) representing a given landscape situation by using the generic descriptive model elements to characterize that single land cover unit. Upon this stored information each land cover unit can then be classified according to the chosen appropriate application purpose and nomenclature.

Figure 2 shows another example of how an agricultural farm infrastructure with associated land is described with a descriptive decomposing method.

As has been indicated above, the data model itself does not represent another classification system, but rather functions as an intermediate descriptive and decomposing tool for the semantic transformation of data and a vehicle to interchange land monitoring relevant information on landscape between different nomenclatures and data sets.
Further details on the content and structure of the EAGLE data model and how to use the EAGLE matrix can be found on the EAGLE website\(^7\). It is important to state here that the EAGLE concept with the model, the matrix and its explanatory documentation are living documents and do not yet claim to be finalized or complete. Furthermore, any suggestion for improvement or modification is welcome and will be taken into consideration.

### 5.3. **Assets and benefits of new concept**

The EAGLE concept clearly separates between land cover and land use information, being scale independent while also tackling the temporal aspects of transient or altering phenomena. It reacts on generic or specific user needs arising from different fields of application or different institutional levels. The proposed data model is object-oriented. This means it can store several attribute details as properties of a certain land surface unit. It is flexible enough to be adapted to specific user needs by integrating new model elements that fit into the skeleton of the model. The modular structure of the data model allows the modification of certain parts of the model without the need to re-organize the entire data model.

In principle, the EAGLE concept is designed to be scale-independent, which in consequence means also that it is independent from the data capturing method. However, once the question of source of information is raised, the thematic content of the data model could be categorized referring to the data capture methods by differentiating between remotely sensible information and other information that require further ground truth or in-situ data. This aspect, if necessary or helpful, is beyond scope here but still can be discussed at a later point of time when it comes to the implementation of the concept.

Among the stakeholders who are involved in the land monitoring process or who work with land monitoring data, several institutions on European level can be considered as beneficiaries of the outcomes of this task: EEA, Eurostat, DG ENV, DG CLIMA, DG REGIO, DG AGRI. Also for national and regional land monitoring institutions many aspects covered in this task have the potential to be beneficial for their work and reporting obligations.

For European stakeholders:
- a conceptual framework for data integration and harmonization on pan-European level and for better coordination of data collection, provision of common comparable semantics for all available national and European data.

For Member Countries:
- guidance on how to set up a national land monitoring system and relevant standardized elements.

For service provider:
- get a clear guidance on how data productions should be tailored, what kind of data are required => orientation for product developments.

For Statistical agencies:
- Synergetic requirements and cooperation between land monitoring and statistical community on national level and between national and European level can be channeled through the translator function of the proposed data model.

\(^7\) EAGLE – EIONET Action Group on Land Monitoring in Europe.

6. Follow-up proposals

6.1. Paradigm shift in data modeling

When considering the future capture and management of environmental data, the harmonization of thematic data content and interchange of thematic information between land monitoring initiatives and systems, it is obvious that a paradigm shift is needed on the long-term regarding the data modeling concept. This can be strongly supported by the suggested EAGLE concept.

Figure 3 sketches the outline of the EAGLE concept: The data flow from national and European initiatives fills up a central envelope (data model), where the LC / LU information is stored in a decomposed and object-oriented manner. Information needed by users on different levels is then served from this central envelope. The data model itself does not stand for yet another classification system, but rather functions as an intermediate descriptive and decomposing tool for the semantic transformation of data and a vehicle to interchange land monitoring relevant information on landscape between stakeholders and different nomenclatures.

Integration Scheme of European Land Monitoring Framework

![Diagram of the Integration Scheme of European Land Monitoring Framework](image)

Figure 3: Vision for an integration scheme of a future European Land Monitoring Framework, having the EAGLE concept with data model and matrix as the center piece for integration, storage and exchange of land information.

6.2. SuperCORINE

The above mentioned shortcomings and critical remarks mentioned in chapter 3.2 can be tackled by applying one or more of the following modifications to CLC, contributing to an enhanced CLC concept, a so called „SuperCORINE“:

A. Add further levels as subclasses to CLC classes
B. Add attributes to CLC polygons
C. Improve description of CLC class definitions
Harmonised European Land Monitoring

Summarise step A and B

A short-term mode of applying steps A and B to the existing conventional CORINE Land Cover concept would be to keep the basic form of land cover mapping with CLC polygons, but convert CLC into an object based approach by populating the CLC polygons with additional attributes (e.g. irrigation, mowing activity) or parameter (e.g. soil sealing degree, crown cover density). The necessary information to do so could be partly extracted from HRL or other auxiliary data sources. Also the CLC level 4 and 5 classes that can be found in some national initiatives (e.g. PT, HU) could be used as orientation. However, this approach is limited to only some thematic areas and would not solve the shortcomings of CLC in general. It is assumed that statistical users might benefit from the SuperCORINE approach, while other applications ask for more precise location of LC/LU information and require a finer spatial resolution, which in this case could not be fulfilled by SuperCORINE.

Summarise step C

Step C is closer to operational implementation. Within the frame of a task in ETC SIA 2013 implementation plan, the CLC nomenclature guidelines have been revised by applying principles of the EAGLE concept on the definitions of the CLC classes. That way, some semantic gaps and overlaps have been identified and fixed. Also the logical structure of the definitions have been improved. The future successful and meaningful implementation of CORINE Land Cover inventory is seen to be made possible by shifting towards the SuperCORINE concept that can also be seen as a first step to the implementation of the object-oriented land monitoring in Europe (towards EAGLE).

6.3. Methodological harmonization

After the first common European land monitoring standard of CLC has passed its zenith, it is time to face the next (second) phase of harmonization process in Europe. For many European countries CLC is till the only available land monitoring inventory, created by using traditional CAPI methods with national data only as ancillary input. In parallel, the growing number of national bottom-up approaches with all their own specific properties and methods may lead to the development of many varying workflows and processing chains. If more and more countries join in the bottom-up data production approach without commonly agreed principles, the effect will again be that information interchange is only possible with difficulties between individual data sets. Standardization of workflows and processing steps (generalization) and harmonization of data content is needed among all involved countries, if the overall goal is data interoperability and harmonization on pan-European dimension and on national level accordingly.

The EAGLE data model can help in that context, but needs to be fine-tuned and further tested for semantic translation purpose in form of a feasibility study in real application environment. Further development of tools are necessary to make the data model easy to understand and efficient to work with during operative application.

There is a need to formulate and agree on common minimum voluntary standards of both European and national user needs, and as a result provide a simple core data model that can be easily understood and easily applied for the description of most important categories of LC and LU.

8 In the HELM consortium this approach was discussed under the wording “SuperCORINE”.

9 ETC - European topic center for spatial information analysis
6.4. **Standardization**

On the long term perspective, the EAGLE data model could be expressed as a European extension of the (global) ISO standard 19144-2 (**LCML** – Land Cover Meta Language). In parallel to that it is also worth considering the EAGLE data model as an (European) **CEN standard**. A closer collaboration and exchange of expertise with ISO LCML (incl. FAO representatives) should be envisaged and carried out.

6.5. **Legal framework for implementation**

With a view on forthcoming implementation actions, it is crucial to have a robust legal framework that indicates secured funding, clear responsibilities and competences among the involved institutions and stakeholders (addressing key players such as the EEA, Eurostat and MS). Such legal act on the harmonized land monitoring should also regulate the time line of data production and dissemination processes. Further, the licensing issue must be tackled within such a common legal framework. For the preparation of such a legal framework it is suggested to establish a Land Monitoring Advisory Group comprising representatives of EEA, Eurostat, DGs, MS and non-institutional technical experts.

7. **Conclusion/ Key messages**

a) Upcoming reporting obligations will produce demanding requirements for land monitoring activities on national and European level.

b) A single nomenclature as an overall compromise will not be sufficient, and also not flexible enough, to meet all existing and future information requirements compared to an extendable and adaptable object-oriented data model.

c) Therefore it very important to have semantic translation schemes between classification systems.

d) Consistency and backwards compatibility of timelines is clearly within consideration.

e) The object based approach - with spatial reference features which could be populated with attributes - allows more flexibility and can be adapted and fine-tuned according to recent and upcoming requirements.

f) Among the involved and consulted experts from the HELM consortium the EAGLE concept was very much appreciated regarding its flexibility of populating spatial reference objects.

g) The involvement and integration of EU member states interests and requirements is key to the acceptance, co-financing and successful implementation of any kind of future land monitoring initiatives.
8. References and Literature


9. Acknowledgements

During the work on task 4.3 a number of persons supported the author in collecting input and writing this report with their contributions. Many opinions, ideas and suggestions have been expressed. The authors would like to thank all colleagues from the HELM consortium for their commitment and valuable work contribution and great innovative spirit. The proposed EAGLE data model and concept is developed as the result of joint efforts of the EAGLE group (See Annex).
10. Annex – Who is EAGLE

Forming of the group

The EAGLE group is an open assembly of technical experts from different MS NRCs on land cover. The group works and meets on a voluntary basis, not financed by any external budget, except the home institution seconding the expert to the 2 or 3 annual working group meetings. The meetings themselves were supported by FP7 geoland2 and FP7 HELM which also has incorporated EAGLE as a task. The most active core members of EAGLE group are:

- Austria (Gebhard Banko, UBA; Christoph Perger FHWN)
- Czech Republic (Tomas Soukup, GISAT)
- Finland (Markus Törmä & Elise Järvenpää, SYKE)
- Germany (Stephan Arnold, DESTATIS & Michael Bock, DLR)
- Hungary (Barbara Kosztra & Gergely Maucha, FÖMI)
- The Netherlands (Gerard Hazeu, Alterra)
- Norway (Geir-Harald Strand, NFLI)
- Spain (Nuria Valcarcel & Julian Delgado IGN; Cesar Martinez & Roger Milego, UAB)
- Switzerland (Charlotte Steinmeier, WSL)
- United Kingdom (Geoff Smith, Specto Natura)

Within EAGLE an Editing Committee (B. Kosztra, N. Valcarcel, M. Bock, T. Soukup under lead of S. Arnold) was entrusted to collect the groups conceptual input and develop working material including the data model and its explanatory documentation. The European Topic Centre on Spatial Information and Analysis (ETC SIA) provides secretariat service (Emanuele Mancosu).

Connections with other committees and stakeholders
EAGLE tries to bring together the knowledge from the existing EU- and national land cover and land use classifications and initiatives. EAGLE members are also engaged in a broad range of European Land monitoring related activities:

- Several EAGLE members participate in the INSPIRE process as experts in the Thematic Working Groups for Land Cover (TWG LC) and Land Use (TWG LU), which resulted in a productive interchange of expertise concerning the trans-boundary context and the conceptual data modeling method with UML.
- Consultations with the CLC Technical Team (also active in EAGLE, lead by György Büttner) regarding the data model content and its compatibility to CLC are part of the process.
- Overlapping membership between EAGLE and the FP7 project “Harmonized European Land Monitoring” (HELM), which embraces a broader circle of the European MS land monitoring community, including many NRCs.
- Involvement in the validation of High Resolution Layers (HRLs) concept, which was elaborated under FP7 project geoland2.
- Contacts to Copernicus / GMES Initial Operations (GIO) staff in the field of administration and industry on national and European level are maintained on a regular basis.
- Experiences and best practices from already existing object-oriented and/or national bottom-up approaches or from those being under development have been brought together from predecessor projects. Some examples are: LCCS (FAO)/LCML (ISO Standard 19144-2), SIOSE (Spain), DLM-DE (Germany), LISA (Austria), AR50 (Norway), LCM (UK), SLICES (Finland).