Service Contract 3436/R0-Copernicus/EEA.59142

CLC+ components conceptual work

Task 4 Report: Change concepts

Version I1.0

17.12.2023

EIONET Action Group EAGLE Land Monitoring in Europe

Title:

Task 4 Report: Change concepts

Place and date:

Niederanven, 17.12.2023

Contact information:

Stefan Kleeschulte

space4environment sàrl

48, rue Gabriel Lippmann

L-6947 Niederanven

Tel: +352 26 71 41 35

Fax: +352 26 71 45 54

kleeschulte@space4environment.com

http://www.space4environment.com

Table of Contents

1 SCOPE AND CONTEXT

The objective of this negotiated procedure 3436/R0-Copernicus/EEA.59142 was to continue support to EEA with respect to the further development of the EAGLE concept and its application in the context of the CLC+ product suite. The work was broken down into seven individual tasks:

> Task 1 addresses the update of the relevant EAGLE documentation and making them available to service providers as well as users to support the implementation of CLC+ Core, respectively the ingestion of data into the database.
>
> Task 2 concerns the update of the EAGLE web presentation and the integration of the web pages into the new Copernicus web portal.
>
> Task 3 reviews and updates the bar-coding concept in light of the lessons learned during the ingestions of data into CLC+ Core. On the other hand, the task shall help to simplify and streamline the bar-coding of CLMS products and other frequently used feature classes by providing a proposal for a standard bar-coding of these data.
>
> **Task 4 begins to address the development of an EAGLE concept for the characterisation and handling of change data in the CLC+ Core database as well as a critical review on the extraction of change data from the database.**
>
> Task 5 is about providing support to the organisation of Copernicus related meetings.
>
> Task 6 addresses the scope of using AI and ML technologies (and other commercial EO-based analytics) for the gap filling in CLC+ instances.
>
> Task 7 is about providing and “EAGLE view” and support to the ISO standardisation group.

This report summarises the discussions, achievements, conclusions and recommendations of Task 4 at the end of November 2023.

2 ORGANISATION & BACKGROUND

This task was the continuation of a long line of development activity focused on the EAGLE data model and the CLC+ product suite funded by the EEA as part of the Copernicus Land Monitoring Service (CLMS). However, it also represented a considerable departure from previous activity and broke new ground in terms of the concepts, approaches, and challenges it addressed. In this regard the tasked relied heavily on desk studies drawing on previous experiences and discussion sessions between experts within the team.

2.1 Team

The team was mainly made up of EAGLE experts, but also drew on the developers of the CLC+ Core system at Cloudflight.

Task lead	Geoff Smith, Specto Natura
Task partners	Gerard Hazeu (WENR), Barbara Kosztra (Lechner), Gergely Maucha (Lechner), Mirko Gregor (s4e), Stefan Kleeschulte (s4e), Stephan Arnold, Emanuele Mancuso (UMA), Filipe Marcelino (DGT), Mario Caetano (DGT), Christoph Perger (Spatial Focus), Julian Delgado (CNIG), Marian Vittek (WENR), Thomas Mathis (Cloudflight), Johannes Vass (Cloudflight)

2.2 Background

The reliable detection and characterisation of meaningful changes – both in terms of land cover and land use – is the key driver for environmental monitoring. In this respect it was one of the original drivers behind the development of the EAGLE approach, however, to date the EAGLE developments had focused on approaches to describe status, the situation at one point in time, rather than specifically address change between two or more points in time.

The EAGLE approach aimed to account for the variations in nomenclature between different mapping products and thus provide a rich, consistent and flexible description of the landscape. Although these various products are distributed over time, as well as space, the EAGLE data model did not contain all the necessary elements or functionality to capture the linkages which are required to represent that change.

The EAGLE approach is now embedded in the CLC+ product suite. As it becomes fully functional and operational it should give users the opportunities to go beyond conventional change mapping between land cover and land use types held as fixed labels and to look at the more subtle differences related to environmental processes, condition, quality and potentially provide early warning of important changes ahead.

For example, a block of forest with a ‘forest’ land cover class could have one or more functions reflected in land use classes which it may move between depending on ownership or management, etc. For instance, ‘natural’ forest could become ‘forestry’ land use if a license for felling is obtained, or it could become ‘conservation’ land use if the area is designated for protection under national or international regulations. Furthermore, an area with declining tree cover density may remain part of the ‘forest’ land cover and ‘forestry’ land use classes but knowledge of the changing EAGLE characteristic elements could indicate disease infestation, management activity (thinning) or some other process.

The EAGLE data model and CLC+ Core approach also provides the potential for the development of tools and rule sets to improve the detection of change and validate it in a localised landscape context – across different spatial and thematic scales. For instance, an area repeatedly classified as arable due to the presence of bare ground may in fact be new urban if the HRL Imperviousness has identified a high soil sealing, thus preventing a change from being potentially missed.

2.3 Objective

Given the situation at the time of writing described above and the complexities of real change, dynamism and their uncertainty, this task considered the broad needs and challenges when attempting to store and produce meaningful change data within CLC+ Core infrastructure.

At a basic level the task objects had to consider three approaches to the capture and generation of “change” information:

- “input” oriented, i.e., develop a concept for the ingestion of existing change data sets produced by conventional methods such as those used in the CLC and local component updates.
- “internal” oriented, i.e., addressing issues of spatial, thematic, and temporal consistency across the time series of information from different source data sets already held within the CLC+ Core.
- “output” oriented, i.e., creating meaningful change information (real rather than technical or methodological changes) from the layers held in CLC+ Core.

The task at all times considered the existing and ongoing developments of the CLC+ product suite when making suggestions and recommendation. It identified, documented and took into consideration any implications or consequences for the CLC+ Core database architecture and design, and use of the EAGLE-ontology. The task provided a road map for the implementation of best practice examples and recommendations.

2.4 Inputs

A broad range of inputs were identified from actual change datasets already available to documents related to production, evaluation and onward use of these products. These potential inputs are summarised below with notes on dependencies and availability, etc.:

- External change datasets available for CLMS (e.g., CLC, HRL, priority area monitoring products etc.)
- External change dataset from non-CLMS sources
- Feedback documents on CLMS change products
- Production reports (by original service provider)
- External validation reports (by independent service providers)
- Member state reports
- Stabilized EAGLE bar coding for CLC+ (Task 3 of this contract)
- Finalized CLC+ Legacy definition and production workflow based on SP contracts (consultancy contract expected later in the year)
- Concept of backwards compatibility applied for ecosystem accounting / CLC traditional – accounting layers
- Outputs of SC 59032 task 16 on comparison CLC+ Legacy with traditional CLC2018 and CLC+ Legacy (change) data – End of April

3 CHANGE MAPPING

To mis-quote Albert Einstein, 'The measure of intelligence is the ability to [map] change'. Our ability to map environments reliably, repeatedly and objectively already stretches our understanding of landscape features, their abstraction in mapping products and the capabilities of EO to capture this information. Attempting to map environmental change in a similar way is at least an order of magnitude more complex and pushes at the inside of the envelope of our intelligence and perceptions.

3.1 Introduction

There are many national, European and international policies which target the environment and aim for comprehensive, ambitious, and long-term protection of nature. They hope to reverse the degradation of ecosystems and the alarming loss of natural capital that undermines our wellbeing and prosperity. These policies require appropriate change information for their development, implementation, and monitoring.

Off the shelf change information has an important role to play in environmental monitoring because it is much easier to understand and use in reporting. If users were to take two or more separate status products and develop the change information independently the results would just be “differences”, containing a lot of noise (false or technical changes). Working from agreed change products is therefore preferred and also improves consistency between analysis and reporting activity further downstream. Lots of effort is therefore put into producing change products even though the area actually mapped as change may be quite limited.

The production of reliable and repeatable change information is always challenging as it relies on the specification, accuracy and consistency of two or more measurements which must then be compared in the context of the true amount of change. There always needs to be an understanding of the limitations of the input data, the real-world processes and interrelations being mapped, and the fact that sometimes things identified as change will just be errors. For example, landscape changes over the repeat frequencies of the CLMS products are relatively small in extent and require high levels of accuracy in the individual datasets for the changes to be mapped reliably. Significant landscape changes such as felling within rotational forestry, the impact of wildfires and urbanisation may be relatively easy to map even at continental scales. However, over short time periods and for the more gradual changes (e.g., forest regrowth, land abandonment, habitat condition change) their detection may be swamped by the uncertainties of using certain methods.

This is a slightly philosophical point and may be controversial, but users tend to want to do the least work. They therefore expect accurate change information tailored to their requirements, but one size rarely fits all. This results in lots of bespoke change layers that are not comparable or consistent amongst themselves or with their original status layers, and / or spoiled by noise (false changes).

3.2 Approaches

In EO-based land monitoring the change mapping can be based on either the original EO data, the thematic maps produced from the EO data or a combination of both.

The simplest change map is just an indicator of change / no-change, but to be useful it is important to understand the starting and finishing classes (e.g., grassland to urban).

3.2.1 Change between status layers

The most obvious and conventional approach is to take two status maps at different times and intersect them to identify the differences (“post classification comparison method”). Although it is possible to create some form of change technically this way (Figure 1), the important questions are around whether these change results are realistic, reliable and repeatable at future timesteps.

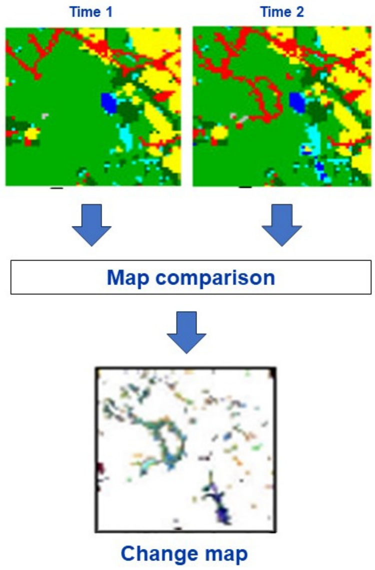

Figure 1. Generation of a change map by the comparison of two status maps.

The difference of two status maps usually includes many "technical changes" beyond the real change features present. Technical changes may be the result of possible differences of:

- Semantic content of status data (class definitions, instructions for visual interpretation, etc.).
- Scale (Minimum Mapping Unit, Minimum Mapping Width, raster spatial resolution).
- Classification methodology and production chain.
- Possible differences of input data used (e.g. EO sensor, spectral bands, spatial resolution, in-situ data, etc.).

The above factors result in uncertainty of the classification of the status maps. When they are compared many small mismatches are likely to appear representing spurious changes caused by errors or misalignment of boundaries. By experience, even if the validation results of the status maps indicate very high overall accuracy, validation result for change maps is always significantly worse.

For the elimination or at least reduction of these technical changes (noise) in production of change maps there are two main strategies:

- Manual, or semi-automatic elimination of technical (false) changes from difference maps.
- Direct mapping of the changes in the first place (see section 3.2.3).

3.2.2 Changes between multi-date EO data

Changes can be mapped directly from the source EO data by measuring the changing response of surfaces over time. The response between two times can be compared to extract some kind of change vector (Figure 2), but the magnitude and direction of the vectors are often difficult to understand in terms of useful real-world changes.

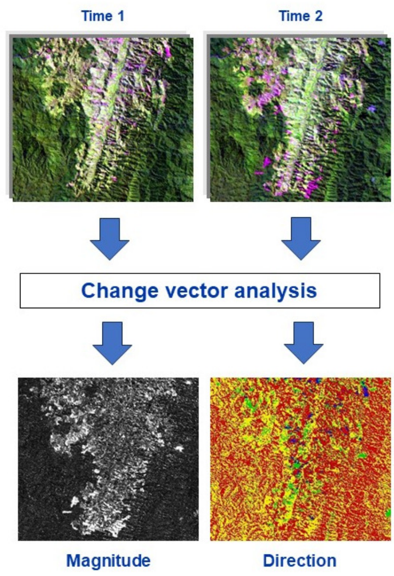

Figure 2. Change vector analysis between two EO datasets.

EO data from two dates can be combined and classified as a whole, resulting in a collection of classes, which represent both the unchanged and changed surfaces (Figure 3). This requires all surface types and the change combinations between them to be identified and understood. These approaches may seem reasonable in principle, but often natural variation in a particular class can be erroneous detected as change (e.g., arable land with rotating crop types or forests with climate dependent phenology), seasonal and climatic variations will have impacts, and gradual changes may involve a potentially infinite number of intermediate states.

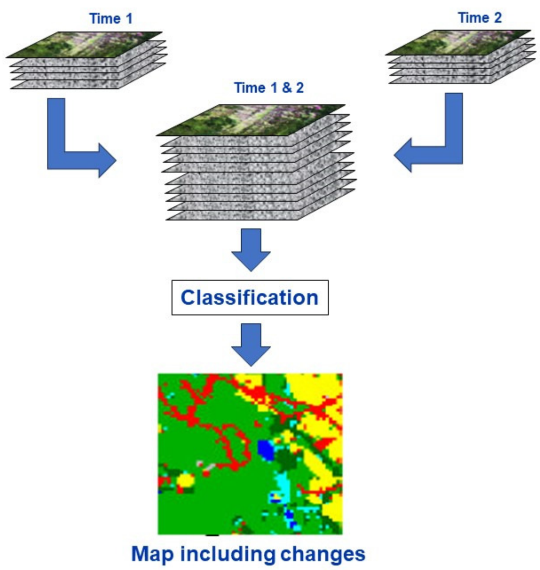

Figure 3. Mapping changes and unchanged areas by a multi-temporal classification.

3.2.3 Change only updating against EO data

The most successful approach to mapping change has been to identify changes directly (using EO data) relative to an existing high quality stock map with the resulting changes used to update to the stock map to the current situation (Figure 4). This is the suggested method for the CLC updates and usually involves overlaying the boundaries of the previous status map on both the previous and current images. In this way both technical (missing from the previous map) and real (apparent in the current images) changes can be identified and mapped.

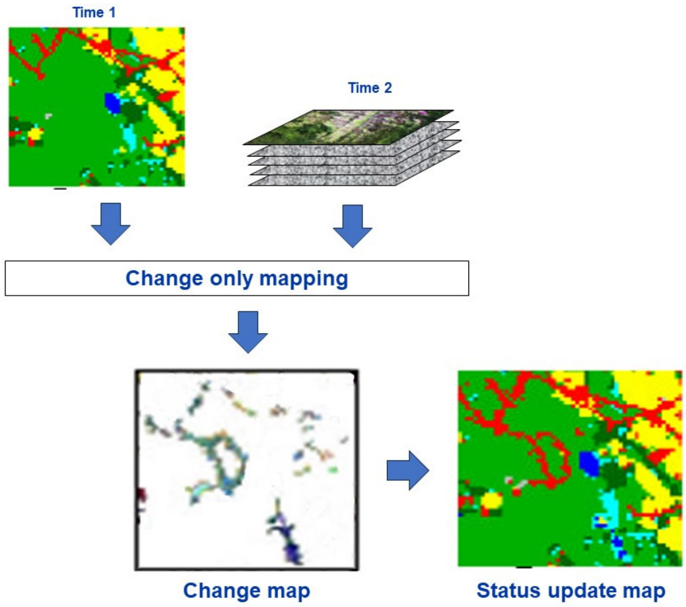

Figure 4. Change only update by comparing an existing stock map to EO data

A variant of change only updating is an equivalent 'backdating' approach, which starts with creating with the current status layer and performs a subsequent (direct) 'backward' mapping of the changes. The previous status map is then produced by a GIS operation.

3.2.4 Changes and trends in time series

Changes are always interpreted between two certain reference dates, but most user needs require the analysis of a time-series to be able to derive and assess longer trends. Consistent time-series require harmonisation between stock and change data as well as between change data of different periods.

The CLC accounting method was introduced at EEA primarily to create a statistical harmonized version of CLC time-series and ecosystem assessments. Similar methodology was used lately to harmonize other CLMS time-series as well, starting with HRL Imperviousness and Forest data. The sole aim of this method is to create for statistical purposes a harmonized time-series of status layers with the combination of latest status and dedicated change data, where the difference of any of status layers corresponds to meaningful changes. More detailed description of the accounting method is shown in section 6.2.2.

4 Existing Change Products

Currently only status information can be ingested into the CLC+ Core platform, however there are a number of pan-European, priority area monitoring and national change datasets (e.g., from HRLs or Urban Atlas) and multi-temporal layers that may wish to be ingested in the future. It is therefore important to review these products as many users and organisations already incorporate them into their established procedures and business system where they work fine, more or less, against their requirements.

A selection of the existing change and multi-temporal information products were assessed below. In the next chapter such products will be evaluated for ingestion within the context of the current EAGLE elements and the production version of CLC+ Core. Which products could be ingested if the necessary mechanisms were available? In the final chapter a list of potential problems and suggested strategies to minimise them will be provided. Do we need to develop different approaches to change mapping, and will they be reliable? For instance, can we use other data help within the CLC+ Core to trap problems?

4.1 Copernicus Land Monitoring Service

The CLMS portfolio contains a number of change products which have been produced as part of the regular status layer updates. Selected products have been reviewed via their technical reports published on the CLMS website.

The CLMS products, including change layers are verified and validated within a number of activities at MS and European levels. CLC is verified by EEA's CLC Technical Team, while other products are verified by ETC/DI as well as by the member states as part of their contributions to the CLMS / CLC production activity. The products are then validated independently by pan-European service contracts.

4.1.1 CORINE Land Cover

CORINE Land Cover (CLC) is the oldest and most used dataset in CLMS portfolio. CLC is produced by national teams in the participating countries (EEA39 = EEA member and collaborating countries and the UK) and is coordinated by EEA. EEA's European Topic Centre CLC Technical Team (CLCTT) provides technical assistance (guidelines, software tool, training, consultation, verification etc.) to the national teams. CLC changes are verified by CLCTT and validated by independent service providers (contracted by EEA). Verification is part of the production chain, gives only qualitative evaluation and has a corrective purpose which is fulfilled by several feedback loops between MS and CLCTT. Meanwhile validation intends to assess quality also in a quantitative manner, to inform users after the production has finished.

The CLC product consists of two main outputs:

1. Status layers representing the land cover - in the form of CLC class code assigned to each polygon - in the reference year of inventory, e.g., CLC2018.
2. Change layers representing the real land cover changes – in the form of a CLC class code pair for the two reference years, assigned to each polygon, e.g., CLC-Change$_{2012\text{-}2018}$.

The change polygons should:

- have a minimum size of 5 ha and minimum width of 100 m
- describe a real evolution process in terms of CLC classes[1] between year$_{old}$ and year$_{new}$

[1] There are real land cover changes that do not represent a CLC change (e.g. appearance of newly planted young trees on a formerly bare-soiled clearcut area, both coded as 324 – Transitional woodland and shrub).

- be visually detectable on satellite images (supported by in-situ data, if needed)
- be mapped regardless of whether connected to an existing polygon or be island-like.

The large difference between the Minimum Mapping Unit (MMU) of the status layers – 25 ha, and the change layers – 5 ha, requires specific considerations when selecting the change mapping methodology.

Since the 2006 inventory, '*Change mapping first'* strategy is the only approach accepted for CLC update. The CLC-Change$_{old-new}$ are mapped (delineated and coded) directly, based on a visual comparison of satellite images of the two reference years against CLC$_{old}$ outlines. Change polygon outlines should be based on (and wherever possible, match) the CLC$_{old}$ outlines. In parallel, revision of CLC$_{old}$ is done for eliminating any mistakes. Revised CLC$_{old}$ and CLC-Change$_{old-new}$ are then combined to create CLC$_{new}$. in a GIS operation, which includes the generalization of features smaller than the 25 ha MMU.

The choice of '*Change mapping first' means:*

- only real changes are mapped, each polygon is evaluated by a photo interpreter,
- all changes > 5 ha are mapped regardless of their position (see Figure 5),
- change code pairs represent real change processes as visible and understandable based on visual comparison of satellite images (Figure 6).

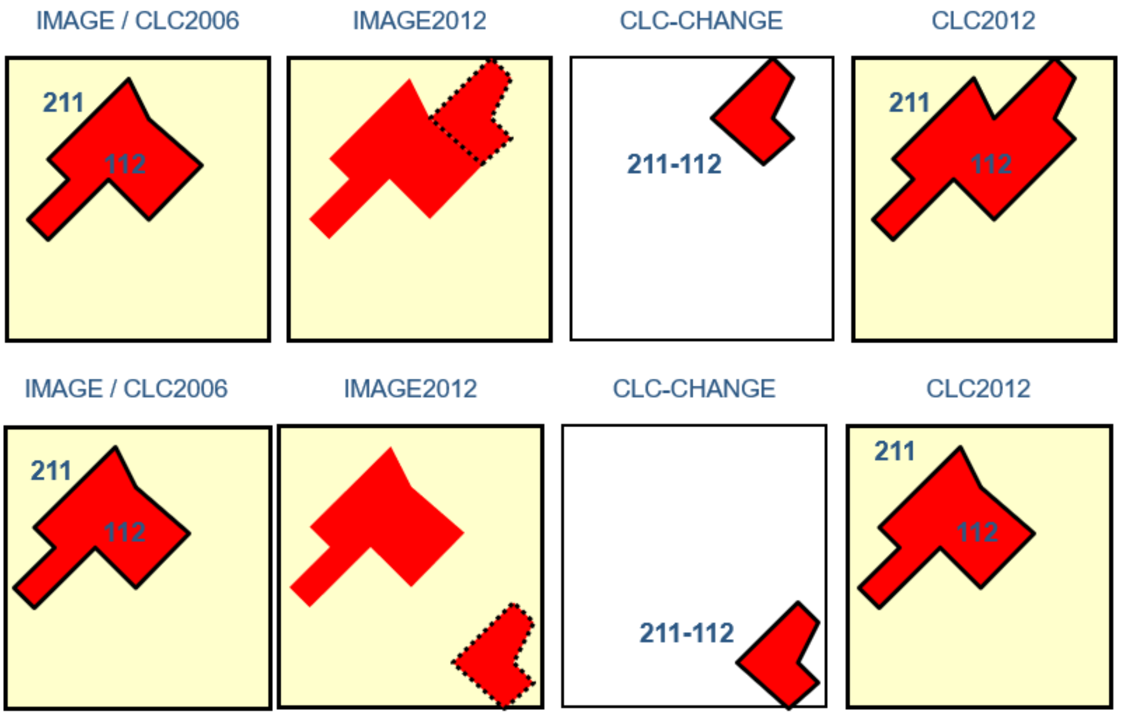

Figure 5. Changes mapped regardless of their position: In 'Change mapping first' strategy a new settlement (112) > 5 ha is mapped is as part of the change dataset (third column) no matter if it is an increase of an existing settlement (upper row) or as an isolated new patch (lower row). The isolated patch would be omitted by 'CLC updated first' strategy intersecting the two status layers (equivalent to first and last figure in lower row).

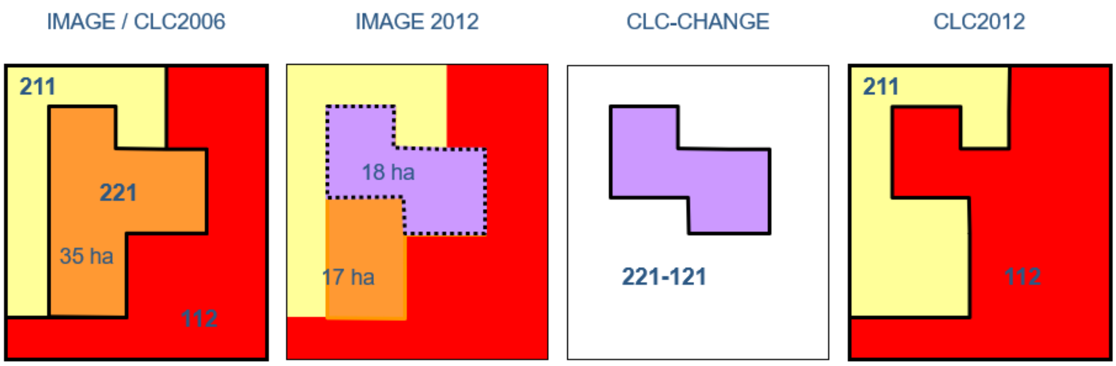

Figure 6. Real changes mapped: In this example, part of a vineyard (221) is occupied by new industry (121). In 'Change mapping first' strategy a change polygon: 221-121 is delineated. As the remaining vineyard and the new industry are both < 25 ha they are generalized into arable land (211) and urban fabric (112).

A backdating variant of this approach is generally used by countries that switch to bottom-up approaches (creating CLC information by combining existing national data).

Semi-automated procedures (as used by many change products) are not straightforward for CLC because of its:

- high thematic resolution (44 classes),
- nomenclature representing much other information besides pure land cover, such as land use, crop types, land management, habitats, status and geographical phenomena.

Identification of CLC classes thus requires either:

- multiple, thematically matching, good quality, frequently updated ancillary input data,
- OR the use of the human mind's ability for abstraction, recognition of shapes, spatial and temporal patterns, context, local background knowledge etc.

This is even more true for the mapping of changes, where not only the above-mentioned capacities are needed, but also a knowledge of what types of changes are expected or possible over the period and at the given location. A simple land cover change showing '*disappearance of tree cover*' can mean a lot of different processes, all depicted by different CLC class code pairs. Decisions between these options relatively simple using visual photointerpretation, but often not so simple when using other methods. For these reasons, CLC change mapping in practice is done fully or partly using visual photointerpretation, also in countries otherwise applying semi-automated/bottom-up methods for creating the status layers.

The *Manual of CLC Changes*[2] guidelines document provides the description of CLC change mapping concept, describes the most frequent change types, with illustrations of good examples and most typical mistakes.

[2] https://land.copernicus.eu/en/technical-library/manual-of-corine-land-cover-changes/@@download/file

4.1.2 High Resolution Layer Forests

The HRL Forests product in made up of a number of sub products which include changes.

The Dominant Leaf Type Change (DLTC) product is a pixel-based 20 m layer covering the time period of 2012 (+/- 1 year) to 2015 (+/- 1 year). It is derived by dedicated GIS operations of the primary status layers Tree Cover Density and Dominant Leaf Type for both time steps. It includes 14 thematic classes including 10 change classes. The product presents a 1 ha boundary filter and changes in the tree cover extent and leaf type are indicated whilst considering a MMU of 1 ha. In the screened validation report, no precise information can be found about how the change layer DLTC was created.

The Tree Cover Density Change (TCDC) product is a pixel-based 100m layer based on the aggregated TCD status layers 2012 and 2015. It summarizes the extent and magnitude of tree cover density increases and decreases over time. The product maps the level of tree cover density changes in a range of -30 to -100% and +30 to +100%, has no MMU (minimum number of pixels to form a patch) and a minimum mapping width of 100 m. The thematic accuracy is determined by the accuracy of the source Tree Cover Density 2012/2015 in 20m spatial resolution. No more than 15% error and preferably less than 10% should be present for both omission and commission errors for the different classes.

The Tree Cover Change Mask 2015-2018 product Provides at pan-European level in the spatial resolution of 20 m information on changes in four thematic classes (unchanged areas with no tree cover / new tree cover / loss of tree cover / unchanged areas with tree cover) between the 2012 and 2015 reference years.

The results at European level for the HRL TCDC show that the accuracy only meets the target accuracy (90%) for both user and producer accuracies of the unchanged area class and for the producer accuracy for both Increased and Decreased classes (few omission errors regarding the change tree cover density).

The main findings and recommendations for the HRL TCDC product can be summarised as follows:

- only meets the minimum accuracy requirement for increased and decreased changes with respect to omission errors.
- shows very high amount of commission errors which lead to unreliable changes.
- updates should include the reprocessing of the change layers to ensure more temporally and spatially consistent changes.
- results provided at bio-geographical regions and country / group of country level, should provide a sound basis for further improving the product.
- the efficiency of the stratification was also demonstrated for the TCDC product resulting in narrow confidence intervals.
- a recommendation for the future TCDC product would be to include information on the time frame of the change.

4.1.3 High Resolution Layer Imperviousness Degree

The IMD layer captures the spatial distribution of artificially sealed areas, including the level of sealing of the soil per area unit for a particular reference year. The level of sealed soil (imperviousness degree 1- 100%) is produced using an automatic algorithm based on calibrated normalised difference vegetation index (NDVI). Imperviousness density change layer are produced between reference years (e.g. IMC2018 is 2015 to 2018) and based on the already existing imperviousness product for that previous reference year.

To derive the Imperviousness change 2006-2009, 2009-2012 and 2006-2012 layers the respective IMD status layers, which were either newly produced or re-processed, are subtracted from each other after considering a rule based adaptation of the historical layers. The classified change is derived by aggregating the IMD change values in specified change classes. The final result are raster datasets of imperviousness degree change including change values from -100 to 100 and raster datasets of classified imperviousness change including defined classes as unchanged areas, new cover, loss of cover, and imperviousness degree increase and decrease. All layers are delivered with a spatial resolution of 20 x 20m.

The product representing imperviousness change 2015-2018 provides a continuation of the legacy of 20 m IMC/IMCC products generated in the previous HRL implementations (2006-2009-2012-2015). However, the increase in resolution from 20 m to 10 m between IMD 2015 and IMD 2018, the effect of changes due to spectral differences is amplified for changes within the sealing area extent. This means – as in the current case a direct subtraction of both layers was conducted after the layers have been adjusted to the same resolution to derive changes – that many areas within the sealing outline will show a certain amount of change which will not be actual (real) changes of sealed surface, but changes resulting from the improved resolution (technical changes).

The Degree of Imperviousness Change - support layer (IMCS) 2015-2018 is a 20 m raster dataset showing the technical change in imperviousness between 2015 and 2018 reference years. The product shows decreased and increased imperviousness density due to technical change (as opposed to real change), or soil sealing and is based primarily on the analysis of NDVI (Normalized Difference Vegetation Index). The additional change-support layer contains the information on all changes caused by technical improvements and thus ensures consistency between the historical and current products. Hence, the IMCS can be used to calculate an IMD 2018 layer with 20m resolution that is consistent with the IMD 2015.

The results for the thematic accuracy of the changes fell short of the target requirement of 90 % at the pan-European level and for the majority of the regionalisations. Nevertheless, the change accuracy results are much improved compared to the previous assessment performed as part of the 2012 production.

Following are the recommendations common for all product time steps noted in both reports (2015 and 2018) unless indicated:

- No details on the input image data.
- Possible underlying landscape level effect on the results.
- IMD products 2006, 2009, 2012 and 2015 are underestimates relative to the actual imperviousness values.
- There has been an improvement in the performance of the IMD2018 layer.

4.1.4 Priority Area Monitoring Urban Atlas

The PAM Urban Atlas (UA) provides reliable, inter-comparable, high-resolution land use and land cover data for 788 Functional Urban Area (FUA) for the 2018 reference year. It has a MMU 0.25 ha in urban areas and 1 ha in rural areas with some exceptions. The nomenclature includes 17 urban classes and 10 rural classes. The UA Change layer provides land cover and land use change data between the current and previous reference periods. The MMU varies from 0.1 ha to 0.25 ha depending on the classes involved and there is a MMW of 10 m. The product is available as vector data.

Change detection was performed with the VHR ortho-rectified optical satellite imagery for the two reference years as the reference. A manual methodological approach based on the visual comparison between the images from two different reference years applied, but with some automatic change detection approaches (image-to-image or map-to-image). The change layer shall only contain real changes between two reference years, and technical changes were excluded. Technical changes were only corrected in the immediate surroundings of the real

LU/LC change area. The actual changes mapped depended on the level of the hierarchical nomenclature and the context. The resulting change layer is therefore not consistent with the two status layers, even after revisions of the previous reference years status layer.

The overall accuracy for changes between 2012-2018 was calculated on level 1 of the nomenclature, except for the urban class where also changes at finer levels are counted as change. However, most of the changes occur within classes of level 1 and the reported accuracies were in the mid-90% fulfilling the requirements in the specification. There were no recommendations to improve the change mapping.

4.1.5 Priority Area Monitoring Natura 2000

The PAM N2K status products for 2006 and 2012 and the N2K 2006-2012 change product represent a limited number of selected N2K sites with a 2km buffer (approximately 630,000 km^2). Land cover / land use information is extracted from VHR satellite data and other available data to provide 55 thematic classes following a nomenclature based on the MAES typology of ecosystems (Level 1 to Level 4) and CORINE Land Cover. The MMU is 0.5 ha and the MMW is 10m. Change mapping was carried out by visual interpretation of 2012 or 2018 vector data and satellite imagery plus ancillary data of the timeframe 2006 or 2012 accordingly, and subsequent direct delineation of change polygons. The thematic accuracy assessment was conducted in a two-stage process, i.e. a blind interpretation followed by plausibility analysis performed on all samples in disagreement with the production data.

The main issues are related to thematic accuracy were the overall accuracy with a range of 76 – 92% (conditionally accepted), correctness of object delineation (not accepted) and the polygon delineation detail (conditionally accepted) are not meeting the requirements.

Several recommendations are given:

- the accompanying scene name information did not include the source,
- auxiliary data sources may have sometimes been used in the production, but this was not always entirely clear.

4.1.6 Priority Area Monitoring Riparian Zone

The PAM Riparian Zones product classification is based on the MAES nomenclature which was revised multiple times during the four specific contracts. The last revision of the nomenclature came up with 4 levels and up to 56 thematic classes with a MMU of 0.5 ha and 10 m of MFW. The aim of the last revision was to harmonize the Priority Area Monitoring (formally Local Component) products (mainly Riparian Zones, Natura 2000, Urban Atlas and Coastal Zones products). The methodology unclear, > 80 % thematic accuracy for change layer 2012-2018

Change was recorded as “CHANGECODE, Change Class code of MAES Level 4 for LC/LU 2018 & 2012, String, Length 11” E.g., 1111_1111 to 10000_10000.

The following definitions and rules for Land Cover Change (LCC) mapping are based on the LCC rules of CORINE Land Cover (CLC) change detection methodology. The given rules were adopted and expanded to the Natura 2000 specifications and requirements and are now implemented for the Riparian Zone change mapping. Therefore, the change mapping of the RZ dataset is fully aligned with the CLMS N2K dataset.

For the Riparian Zones, change mapping is carried out by visual interpretation of 2012 LC/LU vector data and satellite imagery of the timeframe 2012 and 2018 and subsequent direct delineation of change polygons. The final vector data file contains the complete LC/LU status for both timeframes (status layers) and the changes (change layer). Although the basis of identification of changes is the interpretation of detectable land cover differences on satellite images from 2012 and 2018, the support of ancillary data has proved to be very useful. Plausibility checks are performed to detect unlikely or impossible changes.

A key message from this work was that the production must maintain specifications between reference years.

4.2 National experience

A number of countries have detailed change data, but they are not necessarily available. Others have addressed the challenges of mapping reliable change information and developed useful strategies. This section tries to capture knowledge and lessons learnt.

4.2.1 The Netherlands

Generally speaking, in The Netherlands there are two different approaches used to generate land cover / use change datasets, i.e., **mapping changes** and **generating changes**.

The **mapping change** approach uses the previous version of the land cover/use dataset (t-1) and adds the observed changes to generate the new land cover/use dataset (t). This corresponds to the approach described in section 3.2.3 above.

The **generating change** approach uses independent land cover/use status layers for each time step and tries to derive changes afterwards by subtracting the datasets at t and t-1. Often the subtraction results in a combination of real and false changes. The elimination of false changes is often based on a combination of expert rules and thematic aggregation of land cover/use classes into monitoring classes (agricultural land, greenhouses, orchards, forest, water, urban areas, infrastructure and nature). This corresponds to the approach described in section 3.2.1 above.

The **mapping changes** approach was applied in previous versions of Landelijk Grondgebruiksbestand Nederland (LGN) of Wageningen Environmental Research (WENR), i.e., LGN3, LGN4, LGN5, LGN6 and LGN7, CORINE Land Cover for The Netherlands and also in the Bestand BodemGebruik (BBG) and partly in the upcoming new BBG (NBBG) of Statistical Office of The Netherlands (CBS). In the case of LGN, changes are only mapped at a higher thematic aggregation (monitoring) level and marked as 1 in a separate change layer. The changes can be generated by subtracting both status layers and/or a separate change layer is existing. All changes are real changes, and no false or methodological changes are present. There is no trend break!

In recent years the mapping of land cover/use for LGN has switched to the mapping land cover/use status layers and **generating changes** out of both status layers by subtracting them (from LGN2018 onwards). The more automated production, the higher quality of underlaying datasets and limiting manual visual detection of change were incentives for this switch in the approach to map land cover/use and its changes.

Although the production methodology between LGN2018, 2019, 2020, 2021 and 2022 did not vary much, still a lot of false changes were present when combining two consecutive status LGN layers. The generation of false changes is caused by 1. changes in the underlaying source data, 2. small differences in the production methodology of the independent LGN layers and/or 3. remote sensing classifications of nature areas.

The solutions implemented for eliminating false LGN changes generated by subtracting two independently produced LGN status layers are the following:

1. Filtering false changes.
2. Replacing the LGN_t class by LGN_{t-1} class for areas with false change.
3. Aggregation 51 LGN classes into 8 monitoring classes.

The filtering / selection of false changes (step 1) is an approximation that highly depends on expert knowledge of both datasets and the Dutch landscape. False changes are filtered / selected by size, shape and /or due to methodological differences (i.e., changing class definitions or underlying datasets). Step 2 is implemented to generated two status layers for t and t-1 that after subtracting come up with only the “real” changes. By aggregating (step 3) land cover/use into only 8 classes the number and surface area that is affected by LC/LU changes is decreasing. Changes within, for example urban areas, agricultural areas or nature areas are not any longer seen as changes. Often these changes are temporary (changes between agricultural crops) or due to methodological issues (changes from forest into high shrub/bush vegetation).

4.2.2 Hungary

The development of National High-resolution Layers (NHRL), inspired by CLMS HRLs, was started as in-house research and development activity in the Lechner Knowledge Centre – LTK (and its predecessors) in reaction to emerging need for detailed national land cover data with a yearly update cycle. Implementation was made possible with the consistent provision of Sentinel optical and radar imagery from year 2016 onwards and the special position of the LTK. The institution, being the public administration body for land surveying and GIS, has access to a number of otherwise not freely accessible national databases (however, sometimes with limited usage possibilities). Taking advantage of this opportunity, with the accumulated professional knowledge of remote sensing and data integration methods allowed the creation of initial in-house NHRL layers. The primary basis of the classification are seasonal time-series of Sentinel satellite imagery, but a high variety of available in-situ datasets are integrated by a Random Forest classification algorithm. As production is mostly remote sensing based, extension back in time is possible as well, with certain limitations. The database is a 10m raster data in GeoTiff format.

The nomenclature of NHRL is still under refinement and dependent on the availability of actual in-situ data layers. The main aim is to reach a robust land cover classification over years with gradual refinement of categories. The current nomenclature has 11 classes in the public version, and 18 classes in an expert layer.

Raw NHRL layers are created by individual classification procedures, consequently the time-series of layers is not fully consistent. Due to natural uncertainty of the classification procedures the difference of NHRL layers include many technical (false) changes beyond real land cover changes. Therefore, after the creation of raw (and publicly freely available) NHRL layers a longer visual inspection process is performed; differences of the actual layer compared to the NHRL2016 (first layer, considered as reference) are checked and corrected manually if necessary. Visual inspections and manual corrections are performed down to a certain minimum area by experienced photo interpreters. Final harmonized NHRL layers are created as the combination of NHRL2016 and the manually revised difference layer.

4.2.3 Germany

In Germany, there are two main fundamental cartographic products: ATKIS, the topographic information system, and ALKIS, the cadastral data sets. They both are based on the same feature type catalogue but have different minimum mapping units (ATKIS is coarser than ALKIS), and some differences in the mandatory extent of mapped feature types and attributes (ATKIS is more differentiated than ALKIS). The reason for that is rooted in the historical purpose and development, ATKIS for mapping and orientation applications, ALKIS for documentation and management of land ownership. Both data sets contain information on land use and/or land cover.

The topographic data ATKIS from the German land surveying authorities is used to derive CLC updates, in combination with analysing EO imagery, with the outcome of a digital map in vector format. The cadastral data is used to build up the timeline of the official areal statistics in Germany, resulting in statistical tables. ATKIS generalises small and narrow land units to points and lines without an area size, however ALKIS covers the entire territory with polygon features and therefore true area calculation can be easily done.

Both these bottom-up approaches run into problems when generating robust real change information, because changes in the data bases were not differentiated between real and pseudo-changes. So, when using ATKIS data to derive CLC, and also to generate statistical timelines based on ALKIS, both approaches needed a method to somehow make quality checks on the encountered changes. In the case of ATKIS with remote sensing data, in the case of ALKIS with plausibility routines of eligible change rates. Real change appears after construction sites, or alterations in land use forms from arable land re-naturalised protected riparian biotopes. Pseudo changes appear due to differently interpreted mapping rules, methodological or systematic revisions, e.g., when switching the label of grass-covered tracks between agricultural land parcels from “transportation network” to “arable land” or applying a different minimum mapping unit.

Under the influence of the EAGLE concept, and supported by German EAGLE representatives, two new nomenclatures have been designed, one for pure land cover and a second for pure land use information. The traditional catalogues did not provide complete and separated information to generate full coverage maps / statistics for the themes land cover and land use, in such a level of details as required e.g., for pan-European land monitoring initiatives.

One important detail within the design of these new feature type catalogues are the two new mandatory attributes to flag the character of the changes in the database for every edited unit in an object-based manner. These are “result of review” and “date of review”. The date of review is self-explanatory. The result of review expresses the reason for the change in the database in the form of four code list values:

- error correction,
- current status confirmed,
- new object,
- geometric change (in area size of existing object).

The new nomenclature with the change attributes will be activated by the end of 2023 when switching to 2024. With this new information available, pseudo changes (error corrections) and real changes (new object / change in object area size) can precisely be quantified and differentiated between them.

While the land use change information is basically still done by manual editing (together with the change attribution), the land cover product and its changes come with the commonly known issues of algorithm-dependent accuracies.

Database of land parcels with temporal life cycles. Changes include the retirement of an old object which remains in the database and the creation of a new object. The user can then extract a snapshot of the whole dataset every year or just the changes (using begin and end of life cycle attributes). It took decades to produce this type of dataset. Any changes to mapping rules now can have big impacts on the statistics, but no change logs are kept recording why the changes happened (change reason attribute now added, e.g., real or tech changes as is used in CLC).

4.2.4 Spain

SIOSE – Land Cover and Land Use information system for Spain describes the land surface by a complex data model based in objects (as classes of LC and LU) and parameters associated (as attributes for LC)[3]. The characteristics of the SIOSE data model are quite similar to the EAGLE philosophy but essentially with reduced information content:

- The unique working unit is the polygon, geometry entity of the model.
- Associated with the polygon, the thematic information on land cover and land use is defined. There is a catalogue of types of land covers and land uses.
- Within each homogeneous polygon, it is possible to identify different land covers and land uses, measurable by a percentage cover (%).
- Land cover can be characterized by attributes. There is a catalogue of attributes assignable to classes.

Two particular experiences of dealing with changes in SIOSE:

1) SIOSE updating. To generate new versions based in changes monitoring.
2) Use of SIOSE changes for CLC 2012 and CLC 2018 production.

SIOSE updating

Depending on SIOSE approach, the changes concept is different:

- SIOSE traditional, scale 1:25.000 and versions 2005-2009-2011-2014.
- High Resolution SIOSE, scale 1:5.000 and versions 2014-2017 (and 2020 under production).

SIOSE updating was firstly considered to focus on changes monitoring. With the objective to detect and delineate separately the changes between versions using mainly photointerpretation tasks. SIOSE changes were later integrated on past version (t_1) to get the new version (t_2).

Due to the nature of SIOSE data model (multiple dimensions of information) this method required an advanced photointerpretation and management of geometries and thematic content altogether. The traditional method to identify that a class 'A' has changed to class 'B', similar than CLC, could not be followed because the model allows to have many classes in t_1 that may be changed in t_2.

The typology of changes accepted only certain, either only thematic, sometimes thematic and geometric and even sometimes the changes affected to many polygons in t1 that motivated the concept of 'polygon affected by a change' that needed to be mandatory revised the thematic content, but they were not changed themselves. It was not possible to determine the change flows, only to know the gains and losses for each class in the complete patch of change.

Thie monitoring changes method was more complicated than expected, and more than traditional CLC experiences. And it did not offer a precise and classical sense of a changes (i.e. class A to class B) because it is impossible to know the flow of changes among classes. Only it is possible to know in groups of classes (i.e. classes A, B, C to classes A, D). Statistics of classes were usually exploited by grouping areas as administrative units, like municipalities, provinces, etc. but not actually the change area itself. This fact resulted in most of the users preferring use directly the status version of SIOSE.

This method for changes monitoring was carried out between versions 2005, 2009, 2011 and 2014. Since 2014 the new high resolution SIOSE was established, and changes monitoring approach was modified and based in integration of data and not by photointerpretation. Now each status version is generated in isolation, without changes tracking. Only the modifications made manually in t1 are respected and taken as other data source to be integrated in t2. This method is more coherent with data sources used, such as cadastre, LPIS, environmental

[3] https://www.siose.es/web/guest/presentacion

inventories, etc. that are actually also generated by status versions or continuous updates without databases of changes. This method is applied for changes between versions 2014, 2017, 2020.

Use of SIOSE for CLC production

To produce CLC changes firstly required the transformation SIOSE data changes into a CLC nomenclature following certain criteria. It was recommended to complete the thematic conversion first before geometric adaptation to CLC requirements, because SIOSE contained richer thematic information and some changes in SIOSE did not appear as change in CLC. This step reduced the number of changed areas.

To do the thematic conversion, the output 133 classes (original CLC ones and combination of them needed also to reproduce CLC) were decomposed in an Excel sheet that contained the criteria based in SIOSE classes and other aspects that each class must meet. Subsequently, through a program in FME and following the already mentioned criteria, each of the polygons is assigned to one of the 133 existing classes. Performing this process with two different thematic content for changes in t_1 and t_2 and providing approximate CLC change polygons.

The next step was the geometric adaptation, where SIOSE polygons changes coded as extended CLC needed to be joined, dissolved, and even extended in order to get appropriate CLC changes polygons. This process had an initial automatic phase, that could be successful around 50% of cases, and additional manual revision to finalise and get right CLC changes polygons. Final codification of 44 CLC classes were assigned during these automatic processes and manual revision. This methodology was applied to the CLC campaign of CHA06-12 and CHA12-18, versions generated from pure bottom-up approach based in SIOSE.

4.2.5 Portugal

In Portugal, the National Land Use and Land Cover Map - Carta de Uso e Ocupação do Solo (COS) is the LULC cartography with the longest time-series and highest thematic resolution. It is produced and published by DGT. The COS series has currently five reference dates 1995, 2007, 2010, 2015 and 2018. It is available as vector data with 1ha minimum mapping unit, and the current hierarchical nomenclature has four levels and a total of 83 classes.

The production methodology is adapted from the CLC "change mapping first" approach (section 4.1.1) and is based on the detection and interpretation of changes equal or bigger than 0.5 ha, through a comparison of previous and recent orthophotos, with a spatial resolution better than 50 cm. The change layer allows for the distinction between what is revision / technical change and what is real change. If not for the change layer, all the revisions in the base map during production would be considered as change, and therefore, result in the introduction of false changes in the map. Although the change layer is an important tool during production it is not used to derive change statistics. All statistics are created from status layers, and it is important to establish that before the production of statistics from status layer an exercise of harmonization with all versions of COS is performed to remove false changes created from revisions. This process is called 'Downdating'.

To conclude, change statistics are derived from status layers rather than the change layer, change layers are created only as a method of production, and therefore, changes only exist internally and are not made public.

4.3 Global

This text focus on some global initiatives and specifically how they assess the changes between releases of a certain product at different time-steps.

4.3.1 Copernicus Global Land Cover

The Copernicus Global Land Cover product[4] has a specific approach to provide the change information. The latest release, version 3, of the CGLC product provides information at 100m in 23 discrete classes for the following years 2015, 2016,2017, 2018, and 2019 in a standardized way allowing the calculation of changes.

The annual CGLS products are generated using an 'epoch', where an epoch consists of 3 years of input data (one year before and one year after the reference year). The process includes two main steps, a first iteration to generate a first Near-Real-Time (NRT) version, using images from the previous year of the reference year (reference year -1) and a second step, introducing additional inputs (reference year -2) to obtain a consolidated (CONSO) version. The aim of the second iteration is to validate and consolidate the detected changes from the NRT map, specific technical details on the process are described in the product user manual[5]. The CGLC is not making available change layers as such, but just a "change confidence layer" showing the certainty that a change occurred compared to the previous year (only delivered for maps produced in CONSO or NRT mode. The Change Confidence layer provides the user with a quality layer regarding the change detection of the current mapped year to the previous mapped year. The confidence mask for the break detection has been created based on the amount of evidence there is for the break detection. The users are free to create the change layer by their own using the 3-level confidence mask for all CONSO and NRT LC maps.

On the validation side, from the Product User Manual, is mentioned "Land cover change between 2015 and 2018 were assessed at change and no change level to gain understanding of the consistency and differences of the annual maps. The overall accuracy is 99.6% of the change/no-change map. Here, the no-change class is mapped with very high accuracy, while change class is more likely to be committed than omitted (land cover change commission error 45.6%, omission error 36%)." It is also remarked that "Users should use the annual land cover maps with confidence but should be careful and critical when doing detailed land cover change analysis since uncertainties (i.e. land change commission and omission errors) and related limitations vary for different world regions."

4.3.2 Global forest watch

The Global Forest Watch[6] provides a forest cover map, plus loss, and gain for the period from 2000 to 2012 at a spatial resolution of 30 m, with loss allocated annually.

Forest loss was defined as a stand-replacement disturbance or the complete removal of tree cover canopy at the Landsat pixel scale. Forest gain was defined as the inverse of loss, or the establishment of tree canopy from a non-forest state.

The use of this change information requires cautions as remarked in the usage notes[7]. "Due to variation in research methodology and date of content, tree cover, loss, and gain data sets cannot be compared accurately against each other. Accordingly, "net" loss cannot be calculated by subtracting figures for tree cover gain from tree cover loss, and current (post-2000) tree cover cannot be determined by subtracting figures for annual tree cover loss from year 2000 tree cover."

---

[4] https://land.copernicus.eu/global/products/lc

[5] https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_PUM_LC100m-V3_I3.4.pdf

[6] https://www.globalforestwatch.org

[7] https://storage.googleapis.com/earthenginepartners-hansen/GFC-2022-v1.10/download.html

In the same note, on the validation side is mentioned “The authors evaluated the overall prevalence of false positives (commission errors) in this data at 13%, and the prevalence of false negatives (omission errors) at 12%, though the accuracy varies by biome and thus may be higher or lower in any particular location. The model often misses disturbances in smallholder landscapes, resulting in lower accuracy of the data in sub-Saharan Africa, where this type of disturbance is more common. The authors are 75 percent confident that the loss occurred within the stated year, and 97 percent confident that it occurred within a year before or after. Users of the data can smooth out such uncertainty by examining the average over multiple years.”

Despite the increase technology and improvements in producing data at higher granularity and frequency, the assessment of change is still a major challenge. The indication given by the documentation reviewed promote the use of the estimate changes with certain caution and with a margin of interpretation left on the user’s judgement, a flexible option rather than a close product.

4.4 Summary

The message or information content of a change product is the difference between two points in time and in the case of land monitoring it could be the difference between two status layers that represent real changes on the ground. It does not produce a wall-to-wall map, but lots of white areas and then information at the specific locations of change.

This review has shown that there are many different approaches that are used for change mapping and that these specific implementations and final outputs can be highly variable. The main division seems to be between subtractive and comparative, change only mapping, approaches. In reality, all viable processes to derive sufficiently reliable changes are more complex than just a simple subtraction. The comparative approaches may appear to be more time consuming, but subtractive approaches will always need lots of post processing. The resulting change layers will still sometimes be questionable and / or thematically / spatially less detailed as the status layers.

The specification of the properties being mapped also have an impact. If we are looking for a change in a simple numeric parameter (e.g., increase in soil sealing, synonymous with urbanisation) in a pixel then we just need to measure the difference between to cover values. This is a purer statistical approach. However, when comparing thematic maps with complex geometries it is far more challenging because then we have to compare not only single numeric values, but also thematic compositions which can contain various parameters amongst more abstract class descriptions.

It is also known that there will be a range of change quality based on land cover types being used, the stability of the geometries and the resulting object sizes. Most approaches are likely to form small sliver objects due to geometry changes and that these small objects are mostly invalid.

When dealing with change products, the issues of the status layers (specification, scale methods etc.) and change approach are often baked into end products usually without the required support information (e.g., whether changes are real or technical). Once the eggs are cracked into the omelette, we can’t extract the yolks and the whites. This then means the users or developers are expected to disentangle the true situation later from what could be an ambiguous situation as the changes cannot be reverse engineered to get two original status layers.

If we produce a change product from CLC+ Core we will also be baking in certain constraints and conditions (concepts, original purpose, MMUs, regionalisations, rules, etc.). It is important whether you are asking a question of a processed product or of the raw data which went into that product.

What is the end use of the information? We may want either statistics, topographic spatial data or land cover information. For example, the indicator of settlement and transport area increase in Germany is based on cadastral parcel data. The parcels have land use attributes, and their areas are put into a tabular form at municipality level and the increase or decrease is the change from the previous year. This is only a mathematical operation without much reference to the actual change processes.

If the change product is fit for purpose, then fine, but if not maybe it is best to go back to the raw data. CLC+ Core offers this opportunity, to build your own specific questions.

5 INGESTION OF CHANGE LAYERS

As a way of beginning to explore the issues of change mapping with CLC+ Core and the EAGLE data model the process of ingesting existing change layers was considered. This section therefore explores how change layers might be ingested, what strategies would be used and what extensions might be needed in the EAGLE data model.

5.1 Introduction

Given the current status of the EAGLE data model it was always envisioned that revisions and extensions to the list of EAGLE elements would be required to allow the ingestion of the change and multi-temporal information layers into CLC+ Core. The EAGLE data model is getting well established therefore major adaptations and new elements for change recording should be avoided. The work therefore focused on how we might describe changes based on existing / ”finalised” EAGLE concept from earlier tasks with the minimum of alteration. The open questions for this activity were therefore:

- How to ingest conventional change layers?
- How to mark / ingest the “from” and “to” components of change as EAGLE elements?
- How to ingest temporal information i.e., the “same” datasets with different reference years?
- How to define the link between the “same” datasets? Ingestion of the “same” datasets will be straightforward, but how to extract them as changes?
- How to take into consideration minimum area, type of changes, format (discrete / continuous classes / values) etc?
- How should any piece of land (polygon or pixel) be characterized by a *validFrom* and *validTo* date? To extract a status instance a valid date range will be needed.
- How to use / extend the metadata sections in the EAGLE data model in the barcode?

5.2 Approaches

It should be noted from outset that the current version of the EAGLE data model was designed for the description and characterisation of status information only. Some elements of the data model are related to dynamism, such as agricultural practices, but they do not capture the time aspects of change that are required.

By definition a change is a measure of temporal difference between two states, so **the main problems when using CLC+ Core to ingest a change layer is not to capture the two states, but to link them and determine the time difference between them**.

A good example as a starting point for ingestion into CLC+ Core is to consider how the changes are stored and reported is the CLC Change layer, that is derived as part of the CLC change only update process and contains both technical and real changes. It is attributed with one change code (X), a combination of the two states (e.g., 311-324, deciduous woodland changing to transitional woodland and scrub) and separate codes for the before and after status (before = 311, after = 324), plus an indicator of whether it is a technical or real change etc. The temporal difference is encapsulated in the product name and can be made more specific in the product level meta data.

The before and after status values can be bar coded as normal within the EAGLE data model and there are established and agreed bar codings for all the CLC classes. However, they cannot both be ingested at the same time without generating a mixed and likely confusing bar code.

The EAGLE data model already handles an object’s temporal validity in the same way as the UK’s OS MasterMap, with *beginlifespan* and *endlifespan* attributes, but this just gives the validity of one of the status codes. It cannot handle pure change information as these would need to be two or more separate objects / ingestions with beginlifespan and endlifespan attributes to capture all information. In effect it would have to be two separate status descriptions with some form of link rather than a single change description.

So, as things stand there are no EAGLE elements relate to actual changes and any temporal information relates to a period of stability in the status layers rather than a period of change. Therefore, how are we going to describe input classes from the change products for ingestions into CLC+ Core?

5.2.1 Adapt the EAGLE data model elements

The first and most obvious option would be to add additional elements to the EAGLE data model to record change. Then a single ingestion would identify the change aspects from the change layer and record all the necessary change information in CLC+ Core. The new elements would therefore be related to the status changes that we would expect when mapping change classes and some additional meta data to hold the details of the time differences.

However, there is an added layer of complexity as the CLC+ Core ingestion process requires the **input change class to be barcoded into EAGLE elements**, so the new elements must address element change, not class change. The barcoding of the change class would have **to determine which EAGLE elements have changed and potentially use a mixture of 'status' and 'change' elements**. For example, a change from forest to felled will be a change in the land cover elements, but the land use elements may remain the same.

This would be a **highly diverse set of new change elements that would hugely increase the size and complexity of the EAGLE data mode**l. As it currently stands there are up to 33 land cover components, 102 land use attributes and 624 land characteristics in the EAGLE data model. As change between elements will only be possible within the three categories it could be necessary to include up to 1056, 10302 and 388752 new change elements respectively. Obviously not all of these change elements are physically or practically possible, but it will still be a huge number of additional elements which would make use of model **unworkable**.

5.2.2 Adapt the EAGLE data model for gains and losses

A variation on the above would be to record gains and losses per element in the EAGLE data model. For example, the felling of a forested area might result in a loss of woody vegetation and gain in bare soil, but the forestry land use would remain the same.

Currently, EAGLE elements are accompanied by a barcode which determines their relationship to other elements, so an additional value would need to be added to show the gain or loss of the element due to the change. **These ingestions would also get very complex as each change class will need to be barcoded into a set of gains or losses**.

Maybe we also need to consider the **impacts of cover / density changes** within the CLC+ Core 1 ha grid cell. This will be difficult as it depends on the application if the impact of a change is high or low. Also, at the pixel level of the original data it is known that the density layers are not representative, and they will need to be aggregate spatially and thresholds applied before they are ingested. These rules/thresholds may not be the same for different type of thematic content.

So, in conclusion it looks like this approach would also be unworkable due to complexity of recording gains and losses for ingestions and then interpreting them for extractions.

5.2.3 Adapt the EAGLE data model for two ingestions

As the ingestion of change in CLC+ Core is in effect capturing two status situations then a more appropriate solution may be to undertake **two separate ingestions that can be linked in some way rather than a single ingestion**. One ingestion for the previous situation and one for the current situation with an additional step / process to identify the ingestions as part of a single change item. This would require the addition of a small number of elements to the EAGLE data model so that the two ingestions are linked and additional functionality to the ingestion interface of CLC+ Core.

It is therefore proposed to adapt elements in the EAGLE data model characteristic block where previous and current state elements could be recorded. If we bar code a change layer we have the two situations, before and after update, combined. The double ingestion would flag up whether it is the previous or current timeline scenario. So, we may need two processes, with different characteristics generated and whether it is a technical or real change. It needs to be clearly understood where the change information is generated, is it itself individual status layers with some change logic applied or is the change data in a single dataset.

The following EAGLE element / bar coding ideas offer different approaches for how to store a change data set in CLC+ Core. The details / context of the change information could be expressed in a bar-coding exercise with the matrix elements in (**Table 1**). The first column shows the element code, second column the element name, third column the hierarchical level within the matrix structure, the last column shows the role of the element (whether it is just a segment heading in the matrix or a “bar-codable” Character).

Table 1. Additional characteristic elements to identify the components of a change layer.

Code	***Label***	***Level***	***Element Type***
TLI	**Timeline Information / Land Change Information**	**0**	Heading
TLI-1	**Timeline Window**	**1**	Heading
TLI-1_1	**Previous Timeline Scenario**	**2**	Character
TLI-1_2	**Current Timeline Scenario**	**2**	Character
TLI-2	**Land Change Type**	**1**	Heading
TLI-2_1	**Real Land Change**	**2**	Character
TLI-2_2	**Technical Land Change**	**2**	Character
TLI-3	**Land Change Extent**	**1**	Heading
TLI-3_1	**Content-wise Change**	**2**	Character
TLI-3_2	**Spatial Extent Change**	**2**	Character
TLI-4	**Land Change Balance**	**1**	Heading
TLI-4_1	**Land Change Gain**	**2**	Character
TLI-4_2	**Land Change Loss**	**2**	Character

When ingesting certain input data into the CLC+ Core, and change information is already available in the source dataset, previous and current state of landscape situation could be expressed with TLI-1_1 and TLI-1_2. Further, when a differentiation is given between technical changes (due to systematic issues, without true changes in reality) and real changes in landscape, such scenarios could be capture with those two elements TLI-2_1 and TLI-2_2. Additionally, it could be expressed whether the difference between two timeline scenarios are of content-wise nature (without geometric change), or of spatial extent change (while content stays the same), or both of those two kinds. When such information is not yet clear during ingestion of input data, these TLI elements still could be used to mark changes after a particular extraction ruleset which results in a future change product. The abbreviation TLI for “Timeline Information” is not what comes first to mind when naming “land change data” (LCD), but LC is usually occupied with the term “Land Cover”, therefore here a different name to not cause confusion with any land cover related model elements.

So, if it is assumed to run two ingestions ‘passes’ for each change product, then on the first pass, which describes previous class, TLI-1_1 is set to 1, on the second pass, which describes the current class, TLI-1_2 is set to one. This will result in two separate ingestions, but they will need to be linked in the database and catalogues as a single item. This will require some additional database management functionality.

5.3 Issues

The above descriptions show that it is not currently possible to ingest change products into CLC+ Core and additional development work to the EAGLE data model and the CLC+ Core system will be required. The most viable option (section 5.2.3) has been identified, but a number of issues still remain to be address at the technical and organisational levels.

Firstly, what is the value of holding the change layers as EAGLE elements? They will result in a highly complicated set of elements in CLC+ Core and the extraction of information will be very challenging to formulate. If change is required as an output of CLC+ Core, then why not go back to the original change product or devise an extraction based the original status layers?

What is the future of individual change products within the CLMS? Will they continue to be produced alongside the status layers when they are updated? A case could be made for populating CLC+ Core with status information only and then letting extraction generate the change layers (see section 7).

There are some known issues with the standard terms within the EAGLE data model which may become confusing when being related to the ingestion change information. The team needs to standardize terminology across all elements (e.g., ‘from – to’ should be ‘previous – current’, etc.) when dealing with parameter ‘ranges’ etc. so that confusing terms don’t become established.

There will be issues with the scale and extent of changes when they are ingested into the CLC+ Core with a 1 ha grid cell size, especially if small changes cross grid cell boundaries. There may also be issues with the disappearance of features between the first and second ingestions and how that is identified as change to the EAGLE elements.

Given that the CLC+ Core will contain a wealth of other geospatial information there may be the potential to validate the ingested changes with existing status and ancillary layers. However, it is likely that as the breadth of information being ingested into CLC+ Core expands there will be contradictions between the various elements, e.g., due to differences in thematic scale (nomenclature), spatial scale (minimum mapping unit) or temporal scale (different reference periods). It will be more prominent problem when different national datasets that have been ingested are used to validate changes or derive EU harmonised extracted products.

5.4 Way forward

It is clear from the above that the ingestion of change layers into CLC+ Core will be challenging and that the benefits of the long-term storage and use of this information within the system need to be evaluated. It would appear that change product ingestion may carry with it more complications than benefits.

The most viable way forward given our understanding of the EAGLE approach and the system capabilities has been identified. Further development will require work on extending the EAGLE data model and adding functionality to CLC+ Core. Once in place, ingestion rulesets will need to be stress-tested, applied to real data with the minimum set of requirements, and the resulting element populations compared to identify any issues. These issues will need to be reviewed in the context of how change information within the CLC+ Core would be used and the most appropriate way forward, such as refinement of rulesets or the suggestion for prioritisation of data sets or addition of warning flags, proposed.

6 INTERNAL CONSISTENCY

Most of the work of this task has focused on ingestions and extractions as these are the parts of the CLC+ Core system that are exposed to the user and potentially of most obvious use. However, internal consistency, the harmonisation of time series data, may also be seen as a function appropriate for CLC+ Core, but again it is likely to add further complexity and there will be issues around storing multiple versions of the same input data.

6.1 Introduction

Experience from the development of the time series products with EO data, such as the updates of the HRLs and the production of CLC with variable input data, has shown that consistency of feature identification and property estimation can be a problem, especially when estimating change. Various solutions have been proposed to address this issue, but it continues to remain a problem requiring considerable resources. The production process may need to include additional steps to correct the previous iteration or the whole time series of input data may need to be reprocessed each time there is an update to maintain consistency.

This type of reprocessing may result in multiple versions of the same product being available at the same time or combinations of status and change layers that are not aligned.

Given the storage of multi-temporal version of status layers in CLC+ Core the establishment of internal consistency between them may be seen as a key function of the system. The CLC+ concept will potentially offer new opportunities to build consistent time series or identify issues during production which can be corrected efficiently.

Several environmental indicators are requiring credible time-series providing information about the evolution of land cover and land use in time.

6.2 Examples

To be able to create a consistent time-series – as well as to filter out possible bias in derived indicators due to the presence of "technical changes", harmonization methodologies were developed and applied / tested for some of CLMS products.

Lots of work has been done within the CLMS on harmonisation of layers and during verification tasks due to the operational and sustainable nature of the service. Also, with the emergence of the Sentinel era the repeat cycles have been shortened and any uncertainty or miscalibration has been amplified.

6.2.1 CLC Update Production

As was noted above the recommended CLC production uses a 'change mapping first' based on the previous status product and visual interpretation of satellite imagery. This approach aims to map real change, representing the change process on the ground, and also technical changes (errors in previous product that were missed during the previous update). This is achieved by interpreting change based on a comparison of multi-date satellite imagery with direct delineation of change polygons relative to the previous status map. This produces a CLC $\text{Change}_{\text{old-new}}$ layer with an MMU of 5 ha. The process also produces a revised version of the previous status product in which the technical changes are corrected. The current status product is then produced by combining the CLC change polygons and the revised polygons from the previous product and is represented by the following equation for the 2018 update:

$$\text{CLC2018} = \text{CLC2012}_{\text{revised}} + \text{CLC-Changes}_{2012\text{–}2018}$$

As the CLC status layers are mapped at a 25 ha MMU and the change layer is mapped at 5 ha MMU after intersection and unification, any small polygons are generalised with their neighbours according to a priority table.

So, the CLC production creates a change layer and an additional version of the previous status layer which will need to be handled in some way.

6.2.2 CLC Accounting Layers

CLC-based indicators are usually calculated as the combination of certain CLC classes or CLC change types (e.g. land take indicator). To be able to filter out possible bias caused by CLC methodology and its changes, it is important to understand the (country-specific) evolution of CLC classes and possible bias within these. To detect and illustrate this, CLC time-series statistics were created on a country basis, showing the differences of specific CLC classes over reference dates. While the changes of some CLC classes seems to be as expected, some CLC classes may show critical breaks (sudden increases or decreases inexplicable by known real-life processes).

Possible reasons causing inconsistencies in time-series are as follows:

- While the fundamental mapping and update concept of CLC is unchanged since the 2000-2006 change mapping and update, some countries have switched from using visual photointerpretation to bottom-up approaches that may consist of integration of national data and semi-automated image analysis.
- Continuous revision of status layers based on better input imagery or availability of national ancillary data (e.g. the accessibility of LPIS data leads to identification of more young permanent crop plantations).
- The CLC update methodology does not include the revision of full time-series, concentrates only on the revision of past status and to the correct delineation of changes.

The solution for the harmonization of CLC time-series is applicable for the European CLC mosaics from CLC2000 onwards. It is based on the idea to combine CLC status and change information to create a homogenous quality time series of CLC / CLC-change layers for accounting purposes fulfilling the relation:

CLC change = CLC accounting new status – CLC accounting old status.

Additional aims were:

- Add more detail to the latest CLC status layer (CLC2018) from previous CLC Changes information and use this "adjusted" layer as a reference.
- Create previous CLC status layers by "backdating" of the reference, realized as subtracting CLC Changes based information for CLC2018.

Based on the above principles, the working steps of the creation of CLC accounting layers is as follows:

1) Include <u>formation information</u> from CLC-change layers into current CLC2018 status by creating CLC2018 accounting layer.
    a. Overwrite CLC2018 with code_2006 from CLC-change 2000-2006. Intermediate result: A1_CLC2018.
    b. Overwrite A1_CLC2018 with code_2012 from CLC-change 2006-2012. Intermediate result: A2_CLC2018.
    c. Overwrite A2_CLC2018 with code_2018 from CLC-change 2012-2018. Result: CLC2018 accounting layer.
2) Create CLC2012 accounting by including consumption information (code 2012 from CLC-change 2012-2018) into CLC2018 accounting layer. Result: CLC2012 accounting layer.

3) Create CLC2006 accounting by including consumption information (code 2006 from CLC-change 2006-2012) into CLC2012 accounting layer. Result: CLC2006 accounting layer.
4) Create CLC2000 accounting by including consumption information (code_2000 from CLC-change 2000-2006) into CLC2006 accounting layer. Result: CLC2000 accounting layer.

Harmonization leads to an increased comparability in CLC time series statistics as many effects causing apparent false change signals are filtered out. The resulting data layers are known as accounting data layers and are used presently by EEA's Land and Ecosystem Accounting system[8].

There are still many remaining known conceptual and practical issues, which were introduced with the accounting methodology:

- The generalization level of CLC status layers (25 ha MMU) and CLC change layers (5 ha MMU) is different, correspondingly these datasets are statistically not fully comparable.
- As a consequence of the harmonization methodology, some features below MMU (i.e., smaller than 25 ha) appear in accounting CLC status layers. This does not mean that the overall spatial resolution of the CLC status layer was increased; high spatial resolution features appear occasionally, bound to the location of CLC changes. The presence of features below MMU in a status layer contradicts the original CLC rules and biases statistical characteristics of resulting CLC data.
- Although the "change mapping first" approach assures a high reliability of CLC change features, no cross-harmonization was ensured between CLC change inventories. This may lead to contradictions in CLC evolution processes captured by distinct CLC change datasets and finally may appear in a form of unrealistic features in CLC accounting data.
- Omitted changes cause that false features are backwards "burnt into" previous accounting layers. E.g. if a change from arable to new forest plantation (211-324) is not noticed (omitted) in one inventory, but the evolution from new afforestation to grown-up forest (234-311) is noticed in the next inventory, the 324 polygon will be presented in all previous layers. The extent of this issue is however not known and is being currently investigated.
- While the revisions were included in the new CLC layer, the pan-European mosaics of previous reference dates were not consequently updated with revised CLC layers in CLC2000 and CLC2006 inventories → inconsistency between „new" and „old" status. Example: National CLC2006 was created via revision and update of national CLC2000. This leads to harmony between national CLC2000 and CLC2006 status, but the revisions in national CLC does not appear in European CLC2000 mosaic, leading to local inconsistencies of European mosaic of CLC2000 and CLC2006 layers.
- Revisions were not applied to all previous status layers → inconsistency between „old" and „older" status. Example: CLC2018 was created via revision and update of CLC2012, these revisions were included to European CLC2012, but not harmonized with CLC2006 and CLC2000 leading to but local inconsistencies with previous (CLC2006, CLC2000) status.

[8] https://www.eea.europa.eu/data-and-maps/data/corine-land-cover-accounting-layers/clc-accounting-layers

- Revisions were not applied to previous CLC changes → inconsistency between revised "old" status and previous change layer.

6.2.3 HRL reprocessing

The HRL teams have experienced harmonization issues, such as imperviousness, where uncertainty within and between the individual status layers were causing spurious changes to appear. Water surfaces are expected to be relatively easy, although they may have seasonal extent changes which are real changes. However, tree cover is already problematic and an SLA task in 2023 proposed to discover possibilities of time-series harmonization.

Imperviousness accounting

On-going work performed in the frame of ETC/DI SC59425 "Support to the development of indicators and assessment of impacts of land use on the landscape scale" upon request from DG REGIO includes "Task 1 –Analysis of usability of Imperviousness vs. CLC+ backbone data for mapping sealed areas". Results include a harmonized time-series of soil sealing created on the basis of combining CLC+ Backbone status with Imperviousness Classified Change (IMCC) data.

The HRL Imperviousness (IMD) continuous datasets capture the spatial distribution of artificially sealed areas, including the level of sealing of the soil per area unit. The level of sealed soil (imperviousness degree 1-100%) is produced using an automatic algorithm based on calibrated normalised difference vegetation index (NDVI).

The 2018 upgrade to a 10 m spatial resolution from 20 m has unfortunately resulted in a break in the IMD based areal statistics. While the former 20 m resolution IMD time-series (2006-2009-2012-2015) has been successfully harmonized and have shown credible evolution in sealed cover, the upgraded IMD2018 data detect significantly more sealed structures than before, thus the amount of sealed cover is showing an unrealistic growth in the 2015-2018 period.

CLC+ Backbone 2018 categorical data provide a wall-to wall land cover coverage of Europe in 10 m resolution containing altogether 11 land cover classes. The semantic content of class 1 (Sealed areas) is by definition very similar to the semantic content of the IMD data.

QA/QC analysis results and additional comparisons of CLC+ Backbone 2018 and IMD 2018 data indicated, that CLC+ Backbone 2018 class 1 "Sealed" provides slightly more consistent estimation of real sealing, than IMD 2018, although derived statistics and appearing features are very similar.

Additionally, "new cover" and "loss of cover" classes of Imperviousness Classified Change (IMCC) data have shown acceptable to good quality by Member State and ETC verification results.

This led to the idea of applying a CLC-accounting-like harmonization methodology by combining CLC+ Backbone 2018 status with IMCC change data to create a harmonized time-series of soil sealing data for the reference years of 206-2009-2012-2015 and 2018.

As experience of visual verifications showed, IMCC 2015-2018 data includes significantly more noise-like technical changes, than IMCC data for previous periods, therefore a complex algorithm was developed to filter out suspected technical changes for the latest period.

Results of harmonization include:

- 10 m resolution binary sealed/unsealed time-series for EEA39
- Aggregated sealing density time-series in 100 m and in 1 km resolution.

An EU reference sealing level database was created by visual interpretation, estimating sealing densities for more than nine hundred 1km x1km sample areas. This database allowed the comparison of EO based sealing level estimations with the point sampling-based estimations in a direct and detailed way. Based on this comparison, statistical bias of CLC+ Backbone / HRL

Imperviousness based soil sealing level estimation was corrected for the aggregated 1km sealing level data.

<u>Harmonization of HRL Forest time-series</u>

CLC-accounting-like methodology was also applied to create a harmonized HRL tree cover mask for the reference years of 2012-2015 and 2018 by combining the latest (2018) status with change data available for previous periods. This more experimental activity was performed and supported with limited resources in frames of two different contracts:

1. Under AP2021 Task 1.2.4.2 "Forest assessment, supported by Copernicus and contributing to FISE".
2. In the frame of SC59425 "Support to the development of indicators and assessment of impacts of land use on the landscape scale" Task5: "Urban green infrastructure".

The first harmonization effort has resulted 20 m resolution time-series by exploring two options:

- Creating harmonized tree cover presence maps based on the combination of a tree cover mask based on DLT2018 and previous Tree Cover Classified Change (TCCM) data.
- Creating harmonized dominant leaf type (DLT) maps based on the combination of a tree cover mask based on DLT2018 and previous Dominant Leaf Type Change (DLTC_1518) and TCCM_1215 data.

As DLTC layer was not available for the 2012-2015 period, CLC2012 status and CLC 2012-2018 change layers were included as input to a complex ruleset to create harmonized DLT2012 status.

Results of the harmonization activity were 20 m resolution harmonized raster layers of:

- Tree Cover Mask accounting layers: TCM2018, TCM2015, TCM2012
- Dominant Leaf Type accounting layers: DLT2018, DLT2015, DLT2012

The second harmonization effort was performed by combining CLC+ Backbone tree cover status (sum of CLC+ Backbone classes 2, 3 and 4) and HRL TCCM change data, resulting 10m resolution harmonized Tree Cover Mask accounting layers.

Harmonization efforts on forest time-series are done mostly by GIS combination of available raster layers. Changes in specifications of HRL Forest layers, missing DLTC 2012-2015 data and known issues especially concerning TCCM 2012-2015 data presuppose significant quality issues still present in harmonized datasets. Limited resources available for harmonization activity has not allowed detailed validation of harmonized results yet.

6.3 Issues

As the CLC+ Core becomes populated with status there may be many potential opportunities to harmonise results and attempt to produce harmonised time series of information. However, there will be many issues to address as well as those already noted above from similar activities.

As a first step it is critical to limit activities to the layers which have been produced to the same specifications understand the, e.g. tree cover from same year from different maps.

In particular the HRLs offer some unique challenges for change mapping as they may not only change to something else but also features will disappear as they will not be present in due to the product specification. Additionally different temporal frequencies of production may miss some changes.

6.4 Way forward

For the internal consistency processing there are still many unknowns. The experience from the accounting layers and the imperviousness and forest harmonisation is extremely useful in understanding the requirements. Also, beyond Copernicus some image datasets and analytics layers have been reprocessed one or more times. The key questions is how would we implement similar processes and capture the results within the CLC+ Core?

At a higher level, should the layers within CLC+ Core to be reprocessed to make them internally consistent before or after ingestion? Then have both the original and reprocessed versions of a layer exposed to CLC+ Core users and how do we identify / link them with the EAGLE data elements?

There will be a need to develop a concept for maximising time-series consistency for other inputs to the CLC+ Core such as LUA / LCH layers, although currently major change products are land cover related and little is known regarding ingestion of land use change products. There is no information about LUA time series yet (except CLC based).

These layers are derived from highly heterogeneous and numerous sets of sources. The CLC+ Core approach will offer a range of options to cross check these inputs against historic and contextual information already held in CLC+ Core. It is likely that this work will propose a generic set of tools and best practices which may need to be tailored for different inputs and / or an approach which harmonises a time series once it has been fully ingested into CLC+ Core.

It is known that the original status layer should not then be used for comparison with the change layers after internal consistency processing, but who will be informing the users to tell them which layers are valid? There needs to be strategies set up so that methodologies for status layer production are consistent and then it is decided if the changes will be good enough.

Who will have permission to alter layers within CLC+ Core; administrators, data owners or everyone?

7 EXTRACTION OF CHANGES

Although there are many issues around CLC+ Core and change, the extraction of changes from the status layers in CLC+ Core should be the main issue moving forward. It goes to the heart of why the CLC+ Core was implemented and how it will address user requirements either as specific instances, standard products or the ability to answer questions.

7.1 Introduction

Change products capture a start and end state separated by a clearly defined timestep. For example, the CLC Changes$_{2012\text{-}2018}$ is a commonly used change product created during the 2018 update of CLC.

As noted earlier the EAGLE data model was only designed and used up until now to describe status information and thus the CLC+ Core is essentially a store of thematic status information from a broad range of sources. The status information is encoded as a collection of EAGLE elements for each class in each of the input datasets. By default, the CLC+ Core also holds some temporal information as each input dataset will record status for a specific period of time and thus has a reference year (or period) added to its metadata when it is ingested. So, in principle it should be possible to produce change information from CLC+ Core if suitable extraction rules can be established to access multiple status layers and make the temporal link between their reference years.

It has been suggested that if all the status layers were loaded into CLC+ Core, then they could be used to generate change that may be more reliable and consistent than existing products by cross comparison of different datasets. The basic comparison may seem obvious conceptually, but a change process cannot simply subtract status layers as this will incorporate, real changes, technical changes and perpetuate, or even amplify, any errors. Also, unrealistic proportions of an area might change if the specifications of the input status layers have changed between epochs. This is one of the main reasons why creating bespoke change products seems preferable at the moment and why so much effort is put in to mapping changes which may only represent a few percent of the total area being considered.

Also, the extraction of a change product based on a pair of input layers with the same specification which had been ingested into CLC+ Core maybe be ‘relatively’ straightforward, but when input layers with different specification are combined to generate changes then it is far more challenging. It will most likely be difficult to understand / explain the majority of the changes identified. Issues of thematic changes and property changes need to be considered. In the European landscape the resulting changes are dominated by small subtle changes which are mainly errors.

A key question is, therefore, what should the CLC+ Core offer in terms of extracting change for the layers ingested. It could be tools for users to generate their own change products from scratch. It could be pre-loaded extraction rule sets which come with a quality stamp. It could be that the change layers are prepared by experts from CLC+ Core and loaded back into CLC+ Core. Also, the derived products could represent changes in EAGLE elements or changes between classes in an agreed nomenclature (e.g. for a CLC+ instance). Maybe CLC+ could provide confidence in the changes ingested or extracted.

So, the main aim of this section is to see how we get information on change out of CLC+ Core by performing extractions on status layers with essentially the same specification. Secondly, we will consider the impact of the integration of different source data is potentially a model of how you might extract changes from CLC+ Core long term.

7.2 Useful examples

To guide the development of extraction concepts a number of different approaches to change detection, the existing change products, the status layer update and harmonisation have been considered in the previous sections. Below are couple of illuminating examples for extraction of changes with EAGLE elements.

7.2.1 Spain

In Spain within the SIOSE system the parameters (elements) could be compared at different times, but the actual change was difficult to interpret. There are many elements available at each time point and some or all elements may be changing and in different ways. The complexity of the changes amongst the elements must be considered in detail. By comparison CLC is quite simple as there is only one dimension (the thematic class) that is changing. The element changes were too complicated, and the users preferred to stick to the status layers of the classes. Recently, simplified geometries were created to help with the temporal comparisons. It was concluded that a complex data model results in even more complex changes.

The automated aspects of the work were very basic. It was found to be more appropriate to simplify the SIOSE elements to CLC classes and then compared these to identify changes. It was concluded that it was best to classify thematically to reduce the complexity. However, it still required lots of manual checking as changes were very complicated, with an approximate 50 / 50 split between automated and manual processing.

This experience suggests that we may need to simplify change extraction concepts to deliver something reliable / acceptable. It is likely the optimal change products will be some kind of compromise between a change / no change map, the 442 potential change classes of CLC and an even greater number of changes between the EAGLE elements.

7.2.2 Germany

In the Germany changes cadastral data modelling have included a new attribute to mark model changes e.g., technical changes, mapping rules, generalization approach etc. It has been suggested that if MS load their data into CLC+ Core they should include an attribute which identifies such technical changes on an object level. This is a big ask, but in the long term this is the ideal process for bottom-up approaches. MS should develop their approaches to deliver this type of approach which has benefits at the MS and EU levels. This national procedure could take quite a while to implement, as in Germany it took years of discussion. On a pan-EU level we need this, and it is best if it comes from bottom level. Spain has also mentioned this. It would cost some money but would need to justify it for our MS decision makers. It is somewhat a chicken and egg problem as the EEA need to set a requirement or there is no progress. The work of the EAGLE Group should be proposing what the MS should do.

An example where this approach would be of benefit is the indicator of urban sprawl from Germany. It is based on land use statistics, but they have issues burnt in which caused unreal results. Then it is hard work to disentangle.

7.3 Possible approaches

The approaches to generating change information from CLC+ Core will rely on the extraction rules. Given the setup of CLC+ Core it is important to understand whether the characteristics of changes to be mapped can be encoded from EAGLE elements, and then can the system be searched for these characteristics. For instance, to map the loss of biotic surfaces and the appearance of artificial, and the change from agriculture and forestry use to urban, the changes will be a related to a number of EAGLE elements which must be combined in two or more rules. An increased number of changes and the greater complexity of changes will require access to more elements and the creation of more rules. By exploring the development of rulesets to generate changes, the complexities and challenges of generating standard products and subsequently the CLC+ Instances will be appreciated.

7.3.1 Inputs – Status and changes

It has been noted a number of times above, but it is worth repeating that the CLC+ Core holds the original status, and in the future potentially change, layers coded as EAGLE elements. Any development of change rulesets and approaches to producing change information must bear this in mind. Unless the change required is in the EAGLE elements alone, a multi-step approach will be required which would deliver intermediate outputs requiring further processing. The layers and elements selected will control the viability and reliability of the changes produced and therefore the input data and their ingestions will have a large impact on the outputs.

When we first envisaged (and as we still visualized) CLC + Core it would only be one set of EAGLE elements but populated by a range of different input datasets. Thus, there would effectively only be one layer or one set of EAGLE elements within the CLC+ Core, possibly with different temporal versions. However, the ability to access separately ingested layers giving the same EAGLE elements we must considered, along with their impact on the complexity and variability of the information available for extractions. For instance, different version of the same change layer could be generated by the same ruleset but drawing elements from different input layers with their own unique specifications.

7.3.2 Output options

The thematic extent of reporting is important. Conventionally, change maps have reported the difference between two status layers using the nomenclature of the status layers as the starting point. Should the change data be targeted at specific changes rather than the standard approach where all possible changes are considered in the output. It would be impossible to get a single change output from CLC+ Core with all information to support all of the potential end uses. Any one size fits all approach would probably resulting in the conflating of lots of uncertainties and require manual editing. It could be possible to design the 'most used' change instances (i.e. LULUCF, SDG, etc.) in a standard way by a core team, and then after being made available for general users avoiding misalignments / mistakes.

It may be better to think of particular reporting questions, e.g., forest and agriculture domains. Develop a ruleset to produce a map, or even a statistical table for a particular requirement. For instance, has the proportion of irrigated land increased or decreased over a period of time, could be a relatively simple ruleset if the appropriate layers, which contain irrigation information, have been ingested. However, it should be made clear that wall-to-wall extractions of change at the pan-European level may not be possible or very difficult if there is too much heterogeneity between the input status layers at the sub-European level.

The types of questions to be asked of CLC+ Core need to be considered in detail. What forms should the change mapping take? Change - no change, change in EAGLE elements, changes related to process (e.g., urbanisation), link elements to ‘change classes’. The answer could be a table rather than a map.

Given the detail in the EAGLE data model it may be best to deal with the EAGLE elements individually. To combine them into something like change classes is likely to be very difficult. Elements can change, but there may not be any real change at the grid cell level. Grid cell may change in reality, but the elements may stay the same.

If change classes at the thematic level of, for instance CLC classes, are required, it might be easier if we were just to continue to compare the classes within the original input datasets. Or you could create extractions for specific years, e.g., 2018 and 2022, and then compare the classes in a separate process outside of the extraction rule sets. This could return the three layers described above, change / no change, time 1 class and time 2 class.

An option is to consider change processes, such as urbanisation etc., and generate them automatically. These results could then be provided as layers to MS who are updating CLC.

All CLC+ Instances are likely to be accompanied by change information and the need for a consistent time series. These are likely to be highly complex in terms of EAGLE elements and might be better suited to a post extraction comparison process.

Could we generate a bar code output per grid cell showing positive, negative and no change per element? Difference in per cell bar codes, based on coverages.

7.3.3 Building rules – current

After considering some of the options from a conceptual point of view this work will now look at the practical development of change rules in CLC+ Core by looking at the capabilities of the current system and thus identify limitations.

When we consider building an extraction rule, we first need to know what is available in CLC+ Core as any change mapping stands and fall on the input data already there CLC+ Core. Data in CLC+ Core is accessed via the originally ingested layers therefore it is vital to know the specification / boundary conditions / quality of every dataset which is going to be used in the extraction. Extractions will depend on the data ingested and the ingestion rules. In CLC+ Core we will actually be comparing EAGLE elements, therefore it is also vital to understand the ingestions that were used to load the dataset into CLC+ Core.

Datasets are selected for inclusion in the extraction rules and then EAGLE elements are selected from them to build the actual rules to create the output classes. The creation of extraction rules first requires the selection of ingested datasets / layers, and they will have their specific characteristics and limitations. The rule developer needs to determine from which datasets the EAGLE elements are picked and with what priority by the order they are listed. Multiple datasets can be requested in the selection and then they are examined in order of priority. If an element is not found in the first dataset, then the extraction rule moves on to next. However, once an element is found then all other datasets are ignored. The priority order of the selection of layers in the ruleset is very important.

The query itself is based on the EAGLE elements used in the formulation of the ingestion rules for the specific input dataset. The specification of the input layers will control the EAGLE elements present in the CLC+ Core version of the layer. We will need to compare elements at t_0 and t_1 to determine change. All rules are a comparison of the coverage of an EAGLE element against a threshold. The operator and value in the rule builder comprise the coverage rule.

One can compare coverages of EAGLE elements against constants (1) or one EAGLE element coverage against another (2). At the moment CLC+ Core rules look as follows:

coverage(LUA-agriculture, 3) > 50 (1)

coverage(LUA-agriculture, 3) > coverage(LUA-forestry, 3) (2)

The rule is defined, and the elements required are compared for the selected input datasets.

Comparing the coverage of the same EAGLE element between different years is currently not implemented. This could be extended to include a time attribute from the metadata, as it is just an enhanced way of selecting the datasets to be used in the rule.

As CLC+ Core is not a true GIS there is no comparator operator where a combination of inputs results in a particular output. For the development of extraction rules it should be possible to select the EAGLE element from only a single input dataset and not from all of them depending on the priority order. We therefore need to build rules which specify which datasets (and therefore times in the case of change) the elements come from so that temporal change can be identified. The comparison of specific EAGLE elements from specific data sets is currently not possible. At the moment the rules only work on the priority ordering of the input classes.

7.3.4 Building rules – changes

The key issues to producing changes will be the functionality exposed to the rulesets and how they can interact with information in CLC+ Core. Whether they pick up change information or rely on the time series of status information in CLC+ Core will be crucial. There must be a clear strategy for moving from changes in EAGLE elements to the change information required by the users.

It seems likely that each possible change will have to be built as a separate rule. If it is assumed that the change extraction will also include the extraction of the CLC classes, then almost 2000 rules would be required for the potential changes in CLC. The number could be reduced if the changes are limited to the most realistic land cover flows.

As can be visualised, these will be extremely complex rules to develop, each based on many EAGLE elements. By analogy with language, we can think that words are being broken down into their letters when they are ingested, but the extraction cannot recover the words as the CLC+ Core grid cell is just a collection of letters. However, from a practical point of view we should stick with the current approach to ingestions, but potential enhancements are listed in section 8.8.

The user has to define specific change extractions linked to their requirements. As described above, the extract rules for producing a conventional change class will need to be made of two basic steps; the identification of change and the conversion of the elements to the thematic class label. These steps can be arranged in different ways. Do we convert (extract) the EAGLE elements into a thematic status class and then do change or do we extract the changes in the EAGLE elements and then create change classes.

The **first approach** would create two extractions representing different times and then compare them. The results of the comparison would be a three-layer group; the area of change, the initial class and the final class. Run two extractions which give the same classes from different (temporal) input layers, although the rulesets might not be the same. To do a change extraction then there must be a common set of elements between the input layers being compared otherwise the development of rules starts to get bizarre very quickly. Humans can do this sort of operation relatively easy because of their understanding complex relationships, such as bushes are just small trees, but we need to be able to code these relationships into a ruleset.

The **second approach** is more complex conceptually as it would need a deep understanding of the behaviour of the EAGLE elements within a change process. Some behaviours are obvious, e.g. clear felling involved a change from woody vegetation to a non-woody land cover, but even they are likely to be complicated by the specification of the original layers and how they were barcoded for the ingestions. It is not realistic for anyone who is not an expert in the target or source nomenclatures and the EAGLE model to be able to undertake such a task.

A further option may be to simply report the **changes in EAGLE elements** as this may be more realistic to achieve. However, would this offer enough information in a concise form to the end users.

The time stamp(s) for the validity of an input dataset is vital. A 'valid from' and 'valid to' attribute would control which data can be used in a comparison of change extraction, but assumptions would need to be made about the relevant reference period. We need to be able to say which datasets are used and incorporate time into the extraction rules.

The foundations for the use of timestamps and temporal processing are present in the CLC+ Core system code base, but they are not being exposed in the user interface of the current application. Therefore, the reference year could be added to the extraction rule with relatively low effort. A rule will still need to be built for each of the EAGLE elements. The value can be another element and therefore there only needs to be the selection of the date / time range to make it work as required.

The design of the CLC+ Core system is prepared for implementing a comparison between the coverage of an EAGLE element between two time points without affecting the core extraction logic. We just need to refine the mechanism by which the coverage function chooses the input classes such that it takes into account the validity time range of the ingestion. That would enable us to support rules which check for an increase of a certain EAGLE element:

coverage(LUA-agriculture, 3, 2018) < coverage(LUA-agriculture, 3, 2022)

This example checks whether there was an increase in agricultural land use between 2018 and 2022. Such rules can still be combined with other rules in arbitrary manner just as the current rules to check for more sophisticated changes.

7.3.5 Summary

As it was shown that the ingestion of a change layer may be possible with minor changes to the EAGLE data model and ingestion process (See section 5), so creating simple change products from CLC+ Core will also need work on the extraction rules.

It is difficult to develop one standard change ruleset, because the queried changes can have manifold aspects over time and there will still be the impacts of the input datasets selected and their ingestion barcodings. There will still be the issue of discriminating real change and technical changes.

If we are comparing classes, then why not compare the original input datasets and we are back to the standard situation. Unless the data products don’t already exist, and they are being compared for the first time. However, for now we suggest that producing change layers should be done through CLC+ Core by extractions then all the other info and layers can be accessible.

7.4 Issues

The three possible approaches for the extraction of change from the CLC+ Core described above have a considerable number of issues associated with them which will need to be addressed. These relate to the concepts of change mapping, the inputs available, the functionality of the platform, the format of the outputs, the access provided to users and the governance of the service.

The CLC+ Core will be populated with data from many sources. If we did ingest all the status layers from the CLMS, MS and other services into CLC+ Core, then there will be a vast amount of complementary, contradictory and duplicate information. If this collection is used to generate change, then this may offer the potential to be more reliable and consistent or poorer quality than the existing layers. Change generation will only be more reliable and consistent if the different status layers are ingested in a harmonised way, i.e. both datasets should be described by EAGLE elements in a consistent way. And both status datasets should have the "same" content. Another point of attention is do we look for changes at class level or change at EAGLE element level?

However, they are likely to be highly heterogeneous in terms of spatial, thematic and temporal details. Also, some historic layers have lots of unknowns. Inconsistencies of MS datasets will become more obvious when ingested into CLC+ Core. Only use the latest for testing to minimize uncertainties.

The accounting layers were the way these issues have been dealt with previously, but that was still challenging event though the CLC status and change layers were relatively consistent. HRLs are better as they have simpler thematic content and the same, or similar spatial resolution. This is an ongoing problematic, one thing is fixed but others are inconsistencies.

In theory a change layer could be produced from CLC+ Core by using very different input data. For example, the CLC2018 could be compared to a MS land cover dataset from 2020. In principle, the forest classes can be compared if their definitions have been bar coded to EAGLE elements correctly. However, there may be other uncertainties which impact the end result. If two extractions are used to produce two forest layers, then the spatial resolution and generalisation of the input layers will have an effect. If the 'forest' elements are compared, then the original class definitions will have an impact.

We have to also be careful that the EAGLE elements used for the ingestions will change over time which could result in technical changes.

We need more information on the ingestion side in a similar way to what is implemented on the extraction side. On the input side the bar codes which express relationships between elements are simple and handle the elements one by one, but don't work for groupings. However, when building an extraction rule, it explicitly links groups of elements with logical relationships. We should consider adding this to the ingestions to be specific for a particular dataset. This is then important for the extraction side when determining change. We are therefore losing information when we ingest. We would need to think about impacts of the changes to the ingestion when deploying extraction rules. Still problems with bar code value 2. It is quite a weak association, but from an area point of view it is highly relevant.

The EAGLE data model was designed for sematic descriptions, but the links between the elements are not clearly made on the ingestion side. When barcoding in the matrix there is no way to make these links. The Excel version is handy but has some short comings. UML would be better, but it is more complicated to work with.

It is clear that there will need to be developments to the functionality of CLC+ Core to allow for the extraction of change information. How should we approach the creation of change rules?

We must remember that extracts are based on EAGLE elements not the original classes. It may be important to look at EAGLE elements and their relationships in the original ingestions, but this is currently impossible and would increase the complexity considerably for the developer and the user who is creating the rule. We can be more precise on the extractions if there is a more specific ingestion description. We need to specify links between elements in the bar coding.

If we do extract changes from CLC+ Core the rule sets will need to be dynamic as each time a new dataset / timestep is delivered it is likely to have specification changes, also products may contain both real and technical changes.

Users will need a deep understanding of the input and ingestions to develop extractions. The simplest approach is to write a rule and not consider source of the elements. But how would the CLC+ Core system perform the sorting and priority setting to decide what is the best source of the element? At the moment this is not implemented and may not be possible.

The extraction rule may draw information from two different and potentially unrelated classes in the input data. For example, a rule to map 'agricultural buildings' in a grid cell may draw from two classes, one defined by agriculture and one defined by buildings, but the buildings may not be related to actual agricultural activity.

There are also issues with transfers between child and parent elements. If one dataset has woody vegetation it must be possible to compare it with trees. Also, bar coding can either be based on the textual definition or on what we expect to appear in the data in reality. In the latter case it becomes very blurry and confusing. If an element or sub element are not part of the ingestions, they cannot be used in the extraction. For example, if looking for changes in bushes, but only woody is encoded then nothing will be mapped. Users must keep in mind the differences between reality, their perception of it and what has been bar coded and ingested. The system must make the link between an ingested child and an extracted parent. And visa versa. Some of this is possible and already implemented.

Time is very critical in the extraction of change but has a number of inconsistent representations in the input data sets and CLC+ Core. Each status layer has a reference year, but in reality, the information will represent a time period over which each status layer is produced. Currently when a layer is ingested the reference year is added to the metadata rather than a range. There is already a time range capability in the metadata. It would be entirely possible to use the previously described mechanism of specifying the year in extraction rules and compare that against the time range instead of the reference year. These differences between reference years and input data windows will have an impact on the resulting changes. This will be amplified by combining data sources with different temporal relevance. Temporal consistency is also vital. For instance, in many countries the LPIS data is updated on a rolling multi-year process, so there is no single snapshot. Also, heterogeneity between different regions, often related to aerial photo coverage update programmes.

Even though working on a 100 m grid in CLC+ Core the spatial resolution of the input data is still critical to which EAGLE elements and proportions are added to the grid cell. For example, this impact was seen when the increase in spatial resolution of HRL Imperviousness layer had an impact on the trends that can be reported as new features and levels of change appear. When updating CLC with Sentinel-2 rather than Landsat the changes from 30 m to 10 m spatial resolution of the input data meant that additional features were mappable, these are differences between the two maps, but they are not actual changes.

We may also have a case of total change within a single grid cell, but the EAGLE elements will remain the same. For example, if a grid cell contained two fields, one of grass which changed to arable and one of arable which changed to grass, then in terms of EAGLE elements nothing has changed.

The results of the extractions would still be differences between datasets and not verified changes unless the EAGLE data has an approved quality and time consistency. Current work on the verification of CLC+ Backbone data and included a request from EEA to assess whether it could be used to map changes. There needs to be an understanding of how much of difference between 2018 and 2021 is change and noise by simple subtraction. It is likely to be mostly noise, the filtering of which takes a lot of effort. Some of the noise caused by spatial fragments and different understanding of classes. There have been some improvements between the dates, but some things are also worse.

7.5 Way forward

From the description of the potential change extraction rule sets and the issues identified above, it is obvious that there is still a lot of work to do both on the understanding and practical application of change mapping with CLC+ Core. Change extractions will be always different, not possible to extract a single change product for all applications.

There should be a discussion with EEA about standard rules sets for more conventional change products, the changes in EAGLE elements and open questioning of the database by users. We will need to formulate some rules, for example SDG indicators might be better delivered through open questions, whereas a CLC changes layer would be a standard product with a fixed ruleset. Should there be some restrictions rather than having every user trying to generate their own CLC changes.

You should use standard rulesets for standard products. There should be official products labelled as such. If the ruleset is fixed, why not have the change layer stored as part of CLC+ Core. Each user has their own space and can publish if wish at various levels (e.g. user, organisation, country, public?). Different levels of access to datasets and rulesets. Extractions with similar names could have very different inputs and rules.

8 SUMMARY CONCLUSION AND RECOMMENDATIONS

This final section of the report will attempt to summarise the findings of this rather complex and challenging piece of work.

As the Magratheans[9] discovered when they devised Deep Thought, a supernatural computer to calculate the answer to the Ultimate Question of Life, the Universe, and Everything, the answer may actually be to better understand the question. So, it is the same with this work which, by exploring how to produce change, has raising questions about our understanding of change, how it can be represented by a data model made up of basic landscape elements and how these elements are ultimately used to provide useful change information.

This report will therefore ask as many, if not more, questions than it answers and provide recommendations for further exploration of the issues by various groups as this area of activity is taken forward. These findings should be considered first by the EEA to guide the evolution of the CLC+ product suite and then the CLC+ Core Service Providers who will need to implement changes to the system. The Member States and other potential users should be consulted further to understand the role CLC+ Core can play in supporting their change detection needs and answering their reporting questions. The whole of the EAGLE Group should be involved to help determine the next steps.

8.1 Overview

It is important to remember that the current EAGLE data model and the CLC+ Core were primarily designed and built to address the requirements for describing and storing status information and not change information.

The EAGLE group envisaged the CLC+ Core as a store of ‘approved’ consistent thematic information at a pan-European level and it was described as an ‘All-in-one data container for environmental land monitoring information according to EAGLE data model’. Although it was to be populated with both European and national datasets for land cover and land use information, it was thought that CLC+ Core would be a single reference sources for EAGLE elements. However, it has now started to be seen as a repository for different versions of similar information at both the European and MS levels which can be accessible by the end user. This apparent drift of focus and the exposure of multiple versions of the EAGLE elements from a heterogeneous collection of input data has resulted in the appearance of a number of issues which affect the complexity and usability of the service.

The results reported here and the issues they raise may help EEA decide whether to continue with the production of individual change products in the future or let a database such as CLC+ Core generate the changes. However, **this will only be possible if a harmonised time series of status layers are ingested that have filtered out any technical changes and uncertainties**. It could also be suggested that change products should be done only for those product lines generated with higher levels of human interaction (i.e., traditional CLC or Urban Atlas) and thus with false / technical changes filtered out automatically. For the remainder of the CLMS products mainly produced by image processing, there is a sense that the changes are not representative and require time series reprocessing. They have accuracies lower than 80%, and changes can have even lower accuracies. For users it is always easier to manage status layers, but it required effort to produce changes.

This task has been challenging as it has forced us to push our understanding of the EAGLE data model and its implementation within CLC+ Core to the limit to address change issues and thus

9 Adams, D., 1979, The Hitchhiker's Guide to the Galaxy. Pan Books. ISBN 0-330-25864-8

it is sometime quite hard to follow the discuss as reference needs to be made to the whole workflow from EO data to the final product. We have tried to include examples, but with limits on what can be done practically with the current CLC+ Core implementation much of the activity has had to be conceptual. Even with relatively simple products such as CLC+ BackBone change mapping is challenging and the sorting of real changes from noise still has to be done manually.

The work has resulted in good discussions and philosophical questions about how the EEA should develop and promote the use of CLC+ Core. Is it a question answering service providing tailored responses specific to a particular user and / or a product producing service to deliver standard change layers? There is a strong need to get a feel for the resource requirements and to understand the effort involved in all these use cases. All change mapping methodologies, whether they be manual interpretation, conventional automated change mapping, or extraction rulesets for CLC+ Core, are not simple.

8.2 Drivers – Why

The success of the CLC+ Core will be determined by whether or not it provides the information required by users in an effective and accessible way. Supporting the production of consistent change information will be one of the key targets for CLC+ Core extractions whether directly or through the generation of consistent status layers that can be compared outside CLC+ Core itself.

It must be understood who will use the service, what will they need to know and when will they make use of it.

This raises the question, why are we ingesting change layers. Isn't the extraction of changes more important? **The intermediate conclusion from this work may be that change layers should not be ingested.**

Why and who will extract change from the CLC+ Core? Should it be generating standard products? Why would users do it themselves? We should ask ourselves if anyone would actually extract changes from CLC+ Core if the ingestions and extraction of layers in itself is already too complicated for most people.

It may be better to understand the drivers behind the reporting requirements and then develop products that address them. Can we shift away from this urge to producing change layers as wall to wall, one size fits all static products? It is not easy to make these products either conventionally or through CLC+ Core extractions. Is there a way to write reports without a change product? For example, comparing status layers in the reporting process to allow context to be included. Reports are the traditional way to map and use changes identification (i.e., by photointerpretation). From now and in the future, most of data will be generated in highly automatic way and in raster format. Reports should be considered as getting changes by status layer comparison. Other discussion would be around where to get the comparisons, directly in CLC+ Core or outside by comparing instances.

However, the production of conventional change layers is likely to still be highly important as they are used to monitor development of the environment over time to assess policies. If CLC+ Core is not making change products in standardised way it may result in a "mess" of change products mixing up technical / real changes and cherry-picking information.

People will probably still want wall-to-wall maps as they will want to know what is happening at specific locations rather than statistical summaries. Users want to know what is changing in a certain period, what are the impacts etc. There could be many questions and many requirements in many formats. It is also important to have the location of change for assessing hot spots of change. Don't necessarily need accurate figures, but actual locations.

8.3 Current limitation

The ingestion and extraction of changes is beyond the capabilities of the CLC+ Core as it is implemented at the moment and further development will be required to support some of the ideas laid out in this report. The EEA are aware of this situation and these results are therefore seen as support to a road map to further development of the service.

Beyond the technical service limitations there are some fundamental issues which underly any change mapping. Change mapping is, and always has been, very tough. By comparing inconsistent / not harmonized status layers we can derive only a mixture of real changes and noise (false / technical changes). Their separation requires considerable (and often unaffordable) effort. It represents an on-going conundrum for the delivery of change products. We want accurate change information that is also consistent with the status layers, but often these two types of accuracy (status and change) are traded-off due to technical and / or resource limitations. Some existing change products are not well explained.

There is both a curse and a benefit of having multiple datasets to call on in CLC+ Core. For example, the loss of forest area could be derived from CLC, CLC+ Backbone, HRLs, national data and all will give different results. The benefit is they generate a range of results and we [as experts!] understand the limitations / uncertainties.

Given the large number layers likely to be available within CLC+ Core there will be a need to trap the use of layers with different specifications for different time points. CLC+ Core should not allow extractions to produce changes between CLC and CLC+ BackBone, without appropriate generalisations and thematic transformations. However, it will be challenging to provide thematic transformations that allow the extraction of meaningful results from two data sets with such different spatial resolutions. It should avoid suggesting that it is possible.

Unfortunately, the producers of land monitoring data don't always see the benefit of statistical snapshot. This means that some of the input status layers to CLC+ Core may be themselves poorly focused in the temporal domain and consequently will produce unreliable change information.

8.4 Philosophy of approaches

In many areas we are seeing paradigm shifts. We need to move away from paper map thinking and offer new ways based on what is available in CLC+ Core. We need to adapt to shifts in geospatial use and new ways of accessing information.

The approach to change mapping should be seen in the context of the evolution of the CLMS product portfolio. The overall philosophy of EEA seems to be shifting towards creating consistent input layers (e.g., CLC + Backbone update has no change component) that can then be used to generate changes from CLC+ Core. However, the intersecting of the current two versions of CLC+ BackBone does not yet support this. What is the role of the change layers moving forward? Do we need change layers to produce an instance? Or are we creating two status instances and then generate the changes between them.

From one perspective, CLC+ Core was not designed as a general storage unit for everyone's data, but as a processing and thematic support engine. The processing engine does the task of extracting change layers, so should not be ingesting change.

From another perspective it might be seen as a harmonised and consistent description and storage of all kinds of data to make processing / extracting of information possible. And in that way, it makes it transparent how a CLC+ Instances are produced.

The philosophy of CLC+ needs to be revisited. CLC+ is a paradigm shift and these exercises will help the EEA understand the new approach. It may be more appropriate for the EEA to see CLC+

Core as an extraction focused service rather than both an ingestion and extraction focused exercise.

Maybe CLC+ Core just generates status layers if the changes are too difficult to produce automatically. For example, you would need access to all of the input layers that went into CLC+ Legacy to generate change, which you would have if it was only derived from CLC+ Core layers. However, this could result in a very complicated rule set for change detection.

The power of CLC+ Core is in the extractions, not the ingestions. **The ingestions should result in a set of qualified, compliant layers with quality flags** rather than try to develop internal consistency.

We need to move away from standard products which aim to support all requirements, such as producing wall-to-wall stand-alone maps. We should move towards a central reliable, consistent information source to which we can ask specific questions. For instance, if a user is assessing water resources and drought status across Europe then they may have a question "Give me all the irrigated areas in Europe?". Look for all cells with the EAGLE element irrigated. Thus, move from a single panoramic product to specific questions on demand. Could this be one answer to why people would use CLC+ Core? This is the real potential of the system.

Two approaches: Ask questions of the underlying data or produce a change layer and ask questions of that. The former is still far in the future.

In theory, we could cross the change check results with the additional information within the CLC+ Core. However, there will be many uncertainties and limitations in the cross-checking datasets themselves which will complicate the comparisons.

The ability to write products back to the CLC+ Core raises further issues. Do you apply a ruleset once and store the results or keep the ruleset available and run the ruleset on the fly? Do we want change as well as status information in CLC+ Core? Should change products only be an instance built from the CLC+ Core? The CLC+ Core would probably get very messy and confusing for users with change and status layers and extraction results being stored together. Extraction rulesets should be available rather than lots of results.

It is a huge advantage that we can disentangle these features by using EAGLE elements and recombine them for a particular use case, but it can become highly complicated quickly.

8.5 Ingestion of change layers

The ingestion of change information is not straightforward. The EEA may see it as an input to CLC+ Core, but maybe it is not a good idea. The EAGLE data model and the CLC+ Core system can be adapted to ingest change, but the resulting information in terms of multi-layered ingestion will be very difficult to interpret, even for expert users. At this point in time, it would be difficult to envision an extraction rule that could exploit the presence of ingested change information within CLC+ Core.

The new CLMS products, e.g., CLC+ Backbone, are not being developed with integral change layer. However, it is important that the updated layers included attributes so that technical and real changes can be identified. Is there a concept to generate / map changes with CLC+ BackBone or is the EEA thinking this will come out of CLC+ Core by extractions? Until now we see a lot of difference between the CLC+ BackBone 2018 and 2021 which are for sure not all changes. These studies are still ongoing.

The creation and ingestion of accounting-type layers (combining status and change information) can serve as input and could fulfil the role of ingesting change information. These would offer some new potential for harmonisation of established concepts which may be migrated to CLC+ Core.

8.6 Internal consistency

Although the development of internal consistency would seem an obvious function of an information storage system such as CLC+ Core, the complexity of trying to do this is extreme. The only functionality currently available to users are the extraction rules and it is difficult to see how internal consistency could be delivered in this way. The accounting layers give an indication of the issues that would be involved when dealing with classes, but it could be an order of magnitude more complex to deal with EAGLE elements.

Developing the ability to reprocess the layers and then write the results back to CLC+ Core means there may be multiple versions of the effectively the same information present and accessible to users likely to create further confusion.

8.7 Extraction of change

The source information for producing change instances, in the form of status layers, is available within CLC+ Core, but we do not have the functionality or methods to express them correctly through a semantic context. Further work is required on CLC+ Core to expose the necessary functionality via the extraction ruleset builder to allow EAGLE elements from different times and datasets to be extracted reliably and compared to thus generate changes.

The extraction of changes is at the moment not the immediate next step for CLC+ Core. First the ingestion of national and European layers, combining the ingested data and extracting new information should be fully tested (develop extraction rules etc). If this is working well and you can extract harmonised and consistent status layers for two or more timesteps you should probably do the change extraction in a GIS. If the different status layers are really produced in a consistent and harmonised way the GIS operations should theoretically result in only real changes, but this may be too optimistic. Without **qualified, compliant, harmonized input status layers with quality flags** the above procedures only theoretically result change data, while in practice a mixed noise and change dataset is created.

If we produce a change product from CLC+ Core we will also be baking in certain constraints and conditions (concepts, original purpose, MMUs, regionalisations, rules, etc.). It is therefore important whether you are asking a question of a processed product or of the raw data which went into that product. What is the end use of the information? If the change product is fit for purpose, then fine, but if not maybe it is best to go back to the raw data. CLC+ Core offers this opportunity, to build extractions for your own specific questions.

8.8 CLC+ Enhancements

The primary enhancements for CLC+ will be the development outlined for the extraction rule builder functionality in CLC+ Core so that tests can be made on the ability to deliver change from the stored status layers. Secondary developments would be around the ingestion of changes if that seen as an important step forward. There is a need to set clear requirements for the developers to explore and which processes, ingestion or extraction of change, are more important.

Although the bar coding of status layers into EAGLE elements for ingestion into CLC+ Core is operational there are still some issues that will then get embedded in the CLC+ Core and meta data is required so that any extractions can be realistic. Change rules will therefore have to pick up or generate a lot of meta data for the users to understand the outputs correctly.

Anything to do with rates will need to have a handle on the actual time of the information collection. Handling time is as important as handing thematic information when dealing with change.

We are suggesting that we need a better understanding or documentation of the input classes through ingestion process. This is an important limitation when users come to develop extractions based around their own concepts of landscape and understanding of bar coding.

On the ingestion side, we may need to do multiple ingestions per dataset to allow for the fixing of the relationships via barcoding. This goes back to the multi-line barcoding examples developed in the past by the EAGLE Group, but these can't be worked with on a machine level. So, a single dataset will be represented by multiple ingested datasets with each a subset of classes to appear in the CLC+ Core.

It is important to understand that there will be limitation on the change mapping we can do if we don't know the links between elements when they are ingested into the CLC+ Core. For example, elements may change, but they may not necessarily represent the change we are looking for. More complex rules on the ingestions side show potential, but they need further exploration and are likely to be complicated to implement.

A manual is needed for the use of status layers and how they can and cannot be combined, especially if national data is combined in an instance such as CLC+ Legacy.

8.9 Governance

The CLC+ Core platform was already quite sophisticated and has become very complicated, even for experts and technician like the EAGLE Group. The EEA will need to develop a strong governance system to control, promote and support its use. If it is completely open, then anything could be uploaded and potentially extracted, and the results added back to CLC+ Core.

What is the CLMS long term strategy for populating CLC+ Core? A critical question for the EEA / CLMS is whether it is willing to fund the creation of the missing data required from the MS. These are typically the land use data, often non-existent or not available. Unless someone finances their creation, there is no way to produce e.g., LULUCF instances, where land use information is crucial.

EEA need to make some decisions. It is too much to expect that each country will have a team of CLC+ Core experts to ingest and extract information. In CLC, the MS were involved directly in producing the change information. In CLC+ Core the restriction of rule writing to the EEA would be very opaque to them. The MS then won't feel involved in the information that is used for decision making. Access to CLC+ Core should not be restricted. MS may then also restrict access to their national data if they have no control over CLC+ Core outputs.

Is it appropriate that any EIONET member can have full access to it, even if their data is inside? The system shouldn't be totally open and needs someone who understands the system to validate. May be only 'authorized technicians' should work with its full capabilities. Every instance with reporting implications (i.e. LULUCF) should be at least invited to be validated to national responsible bodies. This is a tricky matter more rightly dealt with by EEA and EU Copernicus bodies. Everyone should have access to CLC+ Core, but when are products classified as official? Will the MS want to share their data?

Who should be allowed to produce change layers from CLC+ Core? Should there be official layers? Standard rulesets? Who should be allowed to play with this danger toy?

The ingested datasets and extraction rules can be made public and private. It is a political / governance question of who will populate and use CLC+ Core. Only EEA and some EAGLE countries are involved so far but it could become more popular. It is vital then that the change rulesets are properly documented in terms of input data and rule application to add a level of transparency to the resulting change information. Ownership of rulesets and the drive behind it (e.g., EEA, MS, academic, NGO) must be clear. Same level of documentation as the product specification CLMS already uses.

The science community should also have access to CLC+ Core and its functionality. This will likely pump prime development and EEA would benefit from new ideas, but should their experimental results be written back into the CLC+ Core?

8.10 Outlook

A lot of this work has resulted in questions back to the EEA, so there will be a need for a final wrap up meeting with the EEA and potentially a dedicated meeting on change.

When looking to the future of CLC+ Core, it could potentially evolve into a question answering service. However, is that in terms of conventional change products or more tailored user specified results. Both are possible, but they will require different emphasis on the development of the service.

Many of these issues listed will be more tractable when we can start to test ideas, such as a change product, a query, or a SDG indicator in an operational system. Many of the questions around change monitoring with CLC+ Core are too complicated to be addressed fully in a conceptual framework. It is vital that some of the suggested developments of the CLC+ Core are implemented.

When Cloudflight started the CLC+ Core project, it was known that change would be an important issue moving forward, but it was not an initial requirement. Proof of concepts were done at the start of the project, so the required functionality is not fully implemented but some of the groundwork is already in place and could be built upon.

Users now expect human-like interaction, ease of operation and low latency in the response of system. This is seen in the increase of chatbots, intuitive user interfaces that require little or no training and the movement from linear TV to streaming. Users might not wait for new data releases or the production of standard products, they will just want to ask questions in the moment. "Alexa – how much forest has changed in my area?". CLC+ Core offers the potential to deliver this type of service, but it is clear that strong governance will be required.

There are very few datasets that are consistent over the whole of Europe over time to reliably extract change and fewer still are in CLC+ Core. It would be good to show whether the direct questioning approach could provide viable results.

When looking to further develop and test the CLC+ Core, responding to the needs of SNR and the SDGs could be a separate topic as the EAGLE data model is starting to build up these sorts of elements. The EAGLE land characteristics include some condition-oriented elements. E.g., forest damage. They can be populated as shown by the work to Task 6 work.