**Service Contract 3436/R0-Copernicus/EEA.59142**

**CLC+ components conceptual work**

**Task 1 Report**

**Version 1.0**

**10.08.2023**

|  |  |
| --- | --- |
|  |  |
|  |  |



Title:

Task 1 report – documentation update

Place and date:

Niederanven, 10.08.2023

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 objectives 4](#_Toc142564424)

[2 Task 1 – documentation update 5](#_Toc142564425)

[2.1 Objective 5](#_Toc142564426)

[2.2 Results 5](#_Toc142564427)

[2.2.1 Update of the EAGLE matrix and the related documentation 5](#_Toc142564428)

[2.2.2 Update of the EAGLE UML model 8](#_Toc142564429)

[2.2.3 Commenting and updating of the CLC+ Core user guidelines 10](#_Toc142564430)

[2.3 Deliverables 10](#_Toc142564431)

# Scope and objectives

The objective of this negotiated procedure was to continue the support to EEA with respect to the further development of the EAGLE concept and its application in the context of the CLC+ product suite.

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 finally addresses 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 task 1 achievements.

# Task 1 – documentation update

|  |  |
| --- | --- |
| Task lead  | Stephan Arnold  |
| Task partners  | Lechner, s4e, Specto Natura, M. Bock, UMA, Spatial Focus, Cloudflight, CNIG  |

## Objective

The aim of task 1 is to update the documentation of the EAGLE concept to make them available for access by interested stakeholders. The task has three sub-tasks:

1. update of the EAGLE matrix and the related documentation;
2. update of the EAGLE UML model;
3. commenting and updating of the CLC+ Core user guidelines

## Results

### Update of the EAGLE matrix and the related documentation

#### Introduction

Within this task, the EAGLE matrix version 3.1.2 has been reviewed and revised. The result is a new matrix version 3.2. The changes between old and new version are new elements, renamed elements, and elements foreseen to be replaced.

For easier navigation through the new matrix version in detail, please refer to the first table sheet “READ ME” in the excel file deliverable of the matrix. Most changes have been done in the Land Characteristics Block (LCH), only few on LCC and LUA.

#### Revision of Matrix Elements

The following segments have been added, respectively revised:

* LCC Poaceae and LCC Cyperaceae [new]

The LCC “Grasses, Sedges, Rushes, Cereals” (LCC level 4) appeared to be not distinctive enough for the VLCC grassland Layers (CLMS product). Therefore is has been subdivided (on level 5) into LCC “Poaceae, Grasses, Cereals” and LCC “Cyperaceae, Sedges, Rushes”.

* LUA Land Areas Not In Other Economic Use & LUA Water Areas Not In Other Economic Use [replaced]

“LUA Land Areas Not In Other Economic Use” and “LUA Water Areas Not In Other Economic Use” are foreseen to be replaced, and only “LUA Areas Not In Other Economic Use” shall be kept, because from strict LUA perspective it does not matter if on land or water, which is expressed by LCCs anyway.

* LCH Crop Types [modified, extended] and Cultivation Purpose [extended]

Under the segment “Agricultural Land Management”, the main focus in the new matrix version lays on the crop type list. The reason for its revision was the inconsistent mixture of subdividing criteria, like crop form, cultivation purpose or climate zone. Some crops where listed multiple times, depending on how or for which purpose they are harvested. The new order of crop type tries to focus on the form of the crop and part of the plant which is of harvesting interest, in the best case also considering botanical aspects. The previous partly existing cultivation purpose has been eliminated from the crop type list, and was moved and completed instead in its own matrix segment, independently from (and now able to combine with) the crop type list.

Crop Type “Harvested field” has been moved to the segment "Growth Stage". Crop Type “Meadow, Pasture” has been renamed to “Meadow grasses and herbs”, because the old name reflected more a management form than a crop type.

To some extend, a pragmatic view still remains in a way that some crops are not listed under strict botanical aspects but under common sense of crop categories.

On the extraction site of CLC+ Core it will be possible to e.g. query all kinds of energy crops (after they have been ingested as such), without the need to browse through the (old) crop type list and manually hand pick all crops that can be consider as energy crops. This segment also opens up the applicability for other thematic domains like e.g. climate mitigation and biodiversity preservation with the element "functional crops".

* LCH Crop Season [new]

In reaction to the new CLMS product HRL VLCC, crop season (winter, summer, year-round) became important, which now has its own short segment within the matrix.

* Crop Rotation Element [new]

Under this segment, the elementary stages of crop rotation have been added, whether a crop is grown as main crop, secondary crop, side crop or cover crop

* Cultivation Strategy [new]

Under this segment information can be stored if agricultural parcels are managed and cultivated with conventional farming methods or organic (eco-)farming

* LCH Grazing [moved, extended], Grazing Routine [new], Lifestock Species Type [new]

Also under the segment “Agricultural Land Management” the element “grazing” has been extended with a more thorough listing of different grazing cycles and routines. It was done so, because such information is relevant for the assessment of grassland biodiversity and the appearance of herbaceous species depending on the grazing activities. In the old version, grazing was rooted under “cultivation measures”, but has now been removed there and received its own segment, because it has a different character compared to all other only human-based activities like fertilization, irrigation etc.

Grazing routine captures the time intervals (day, night, day & night) that the animals spend on the pasture parcel

The new list of most common Lifestock Species Types can carries information, what kind of animals are grazing the land. As they have different ways to feed, and have different output of animal manure excrements, lifestock has also an effect on biodiversity of grassland types.

* LCH Forest Tree Species [new]

New list of the most common forest tree species, under the existing segment of “Forest Land Management”.

* LCH Ground Water [new]

New entry for groundwater information under the matrix segment “Water Characteristics”.

* Mining Product Type [extended], Mining End Use Potential [replaced], Fossil Fuel Type [replaced]

The Mining Product Type now captures the main mining product types and fossil fuel types, which in older version has been handled with a link to INSPIRE codelists. The latter however turned out to be too broad and complicated, and not easily applicable.

* LCH Climate Zones [new]

New entries under the segment “Geographic Characteristics” with the main four global climate zones.

* LCH Terrain Characteristics [new]

A new segment for terrain parameters has been opened under “Geographic Characteristics”. When implemented, it is now possible to ingest data on elevation, slope and aspect into the Core system. It will be useful to assess probabilities of different habitat types depending on terrain information, or also extract habitat information based parametric terrain data in combination with wetness information and vegetation types.

Terrain Curvature is another descriptive element for inclined terrain worth to consider. However, it has not yet been included in this new version 3.2 of matrix, and can be addressed in a later matrix version if necessary.

* LCH General Parameters

So far, the content names of the parameters have only 1 line to be bar coded. In the case that another matrix field is needed to enter a parameter VALUE, I have added for each general parameter (geometric, temporal) in the lower section of the matrix a line pair "From" & "To". With these fields it is possible to enter or query a range (min / max) of values. If a value is not a range but a fixed figure, the identical value is entered in the "From" and the "To" field.

Together will the colleagues from Cloudflight it will be discussed to what extend it is feasible and necessary to also include the “from..” and “to..” fields in the platform.

* General issues

The EAGLE matrix elements have three identifiers: the URI (Uniform Resource Identifier), the Label Code and the Label Name. When starting to build the CLC+ Core platform based on the EAGLE ontology, it was decided that the URI should be the stable identifier which shall not change. Label codes and label names, however, can be changed when useful or necessary, without jeopardizing the ontology of the running system (regarding ingestions and extractions in use). The user of the platform normally only sees the label code and label name, while the URI is used as unique identifier by the software system behind the CLC+ Core platform.

During the matrix revision process it became evident, that we need a technical approach for changing or disappearing URIs in the system. It is therefore recommended to decide on a procedure, how to conceptually and technically also handle changes in the URIs, and establish semantic mapping mechanisms between old and new URIs.

* Element Label Names [changed]

Label names of matrix elements have been shortened or clarified. This modification has no negative effect on the running CLC+ Core system, only names will appear differently in the user interface.

* Element Label Codes [changed]

The adding of new elements comes along with the need to also adapt the label codes of existing (essentially unchanged elements), whenever these new elements make it necessary to create more numeric space between existing label codes for the new ones.

The adding, moving or replacing of matrix elements goes together with the need to also adapt the label codes of existing (essentially unchanged elements), whenever these new elements make it necessary to create more numeric space between existing label codes for the new ones.

* Replaced matrix elements

When a matrix element has been replaced, it is necessary to elaborate a way how do handle such changes inside the CLC+ Core platform. In all cases, where a matrix element has been replaced, other remaining matrix elements should take over the bar code from the previous (and now disappeared) elements. This is either a 1:1 or a 1:n transition. This may come along with a bar code on a higher hierarchical (parent) level as an easier solution, or in a different segment of the matrix, distributed to a number of other elements.

The ChangeMode Matrix Table (part of deliverables) shows more details.

* LC / LC / Gen assignment of matrix elements

According and in preparation to the ideas formulated further down in the chapter of EAGLE UML model, to distribute all LCH elements to either the theme land cover, land use, or general aspects, each matrix element was flagged respectively. In the matrix column “thematic relevance” each matrix element is marked as “LC”, “LU” or “Gen”. Based on this flagging, LCH elements can later be assigned to the corresponding LCC and LUA model classes in the EAGLE UML data model.

#### Modification of EAGLE element codes (URI, Label Name, Label Code)

The content revision of the EAGLE matrix comes with also modifications in the element coding. The matrix uses three kinds of codes: “Label name” (the textual name of the EAGLE element), “Label Code” (indicates the hierarchical level of EAGLE elements and parent/child relation among them), and “URI” (Uniform Resource Identifier). While the first two codes are for the convenience of users, the URI code is used for machine reading processes.

The pattern of element codes remains the same as in previous matrix. “Label name” is only textual, “Label Code” a combination of matrix block abbreviation (LCC-, LUA-, LCH-) plus the hierarchical position within the element´s parent matrix block (e.g. 1\_1\_3), URI is a combination of the EIONET root and the notation, which is based on and similar to the Label name.

It was discussed, to change the pattern of URI from basically also textual to random alphanumeric codes, but this ideas has been overthrown, to stick with the similarity connection to the label name of the EAGLE element.

### Update of the EAGLE UML model

During the project meeting in November 2022, a number of required (and nice-to-have) modifications to the UML data model have been discussed. Some of which require a considerable amount of effort. The project team has agreed to tackle the following points immediately, whereas others will be kept back for discussion and implementation at a later stage. These are the steps that have been performed already, and the intermediate results are available:

* Update information on the required UML model changes, based on the EAGLE Matrix 3.1.2 (predecessor of the currently developed 3.2). The Excel spreadsheet ([link](https://spatialfocus-my.sharepoint.com/%3Ax%3A/g/personal/christoph_perger_spatial-focus_net/EULXEmuC8hZCg0pToPaZjaUBPtRoR0FKjYt2pGZLwaT7wA?e=NCqKiB)) contains all the required information to update the Enterprise Architect UML data model, however the implementation has not started yet since the Matrix update 3.2 was not finalized, as well as other architectural considerations have not reached an agreement yet.
	+ ***Conclusion****: the implementation of this point (based on version 3.2 of the matrix) is foreseen in the context of the current contract.*
* Proposed the use of codelists (e.g. from the EIONET [EAGLE data dictionary](https://dd.eionet.europa.eu/vocabulary/landcover/eagle/view)) for all look-up attribute values instead of using custom enumerations and codelists; drop enumerations (closed list) in favour of codelists (open and extensible).
	+ ***Conclusion****: this point will be implemented together with the bullet point above after the decision how to handle codelists and data dictionaries.*
* Proposed harmonization steps between the UML model and the matrix have been proposed but not decided on. In short these would be the required structural renaming / regrouping steps:
	+ Land Cover Components -> LC classes in the UML model
	+ Land Use Attributes / Components -> LU classes in the UML model
	+ Land Characteristics -> Attributes of the model, that cannot be assigned to either LU or LC
	+ Land (Cover) Characteristics -> Attributes of the LC classes
	+ Land Use Characteristics -> Attributes of the LU classes (TBD)

Two explaining notes regarding this harmonization between matrix and model:

* + In general, all Land Characteristics (LCHs) from the EAGLE matrix (where they are listed side by side with LCCs and LUAs) need to be attached as UML attributes to concrete UML classes in the model at a specific hierarchical level. Placed at that certain position in the UML model, the LCH inherit the carried information downwards to all subordinated subsequent levels of UML classes. Once a new branched UML class structure for Land Use is opened in the model (as proposed above), it needs to be decided where to place and how to distribute the LCH among the LCC branches and the LUA branches. Therefore, the LCH elements in the matrix are pre-tagged as a) being related to the theme Land Cover or b) to Land Use or c) being general Characteristics that are assigned to the Land Unit class in the UML model (above the point where LCC and LUA branch out).
	+ Renaming, re-arrangements and introduction of new matrix elements also need to be adapted in the UML model. Besides, they are content-wise modifications and are different from the above addressed structural modifications in the UML model.
	+ ***Conclusion****: this issue was put on hold during the Vienna meeting.*
* Draft implementation (prototypical) of a potential UML model extension to integration the land use aspect into the data model. LUAs would be modelled as classes (similar to LC components) and related characteristics as attributes.

At the November 2022 project meeting another improvement for the UML model was discussed, which could be seen as the “missing link” between the matrix and the model. Originally the matrix was designed as a tool to describe nomenclature classes, whereas the model was planned (and used in practical test cases) to describe concrete datasets and its features (incl. object geometries). In the weeks after the project meeting a high-level draft was created ([link](https://spatialfocus-my.sharepoint.com/%3Ap%3A/g/personal/christoph_perger_spatial-focus_net/Eamng-Rrb6BGoLE2XcnMA7YB4RMBJxVhKVuu1GzQzO8r4A?e=phpC89)), on how nomenclatures could be integrated into the UML data model. The PowerPoint slides explain the necessary modifications in order to describe nomenclature classes by setting up relationships between land cover components (LCCs), followed by two concrete examples.

* + ***Conclusion****: this point cannot be achieved / implemented with the available resources as it would require more conceptual changes to the EAGLE approach. Nonetheless, a draft / prototype development could be envisaged as preparation for implementation at a later stage.*

### Commenting and updating of the CLC+ Core user guidelines

During the Eionet training sessions on the use of the CLC+ Core database (i.e. ingestion and extraction of data sets) is clear that the CLC+ Core user guidelines need to be improved in their clarity and by providing better examples.

In the task 22 of the ETC Copernicus service contract 59032 the ETC partners tested the data ingestion process in CLC+ Core. The exercise included the following steps:

* Bar-coding of land use layers form 5 countries;
* Ingestion of those layers into CLC+ Core;
* Testing the extraction of the same layers from the database again;
* Development of a LULUCF instance (not part of the update of the CLC+ Core guidelines).

Based on the experiences from the first three steps above two types of lessons learned were obtained:

* Using the bar-coding approach to characterise land use databases from 5 different countries and the issues from having to consider national specificities or slight differences in the national definitions of similar classes;
* Practical experience from different partners try to ingest the land use data sets into CLC+ Core, using the existing guidelines to support the process.

The experiences from the second group of lessons learned was used to propose concrete text edits (mainly related to practical examples and improved descriptions of individual steps) as well as making suggestions to the general processing / handling of the data (to be handled by the consortium responsible for implementing the database).

## Deliverables

* D1-1a: Updated EAGLE matrix (Excel format) –

<https://projects.eionet.europa.eu/eagle/library/service-contract-no-3506-r0-copernca-eea.59142/task-1-documentation-update/delivery_2/updated-eagle-matrix-excel-format-v3.2>

D1-1b: Updated EAGLE matrix (CSV file ready for upload in EIONET Data Dictionary) - <https://projects.eionet.europa.eu/eagle/library/service-contract-no-3506-r0-copernca-eea.59142/task-1-documentation-update/delivery_2/d1-1b-updated-eagle-matrix-csv-file-ready-upload-eionet-data-dictionary>

D1-2: Updated EAGLE UML model (EA and PDF-Format) - *postponed*

* D1-3: Updated Content Explanatory Documentation of the EAGLE Concept (HTML and PDF format, including the definitions of the matrix / model elements). - <https://projects.eionet.europa.eu/eagle/library/service-contract-no-3506-r0-copernca-eea.59142/task-1-documentation-update/delivery_2/d1-3-updated-content-explanatory-documentation-eagle-concept>
* D1-4: Mapping table between old and updated EAGLE matrix versions - https://projects.eionet.europa.eu/eagle/library/service-contract-no-3506-r0-copernca-eea.59142/task-1-documentation-update/delivery\_2/d1-4-mapping-table-between-old-and-updated-eagle-matrix-versions
* D1-5: Update of the CLC+ Core User Guidelines – <https://projects.eionet.europa.eu/eagle/library/service-contract-no-3506-r0-copernca-eea.59142/task-1-documentation-update/delivery_2/summary-recommendations-update-clc-core-guidelines>