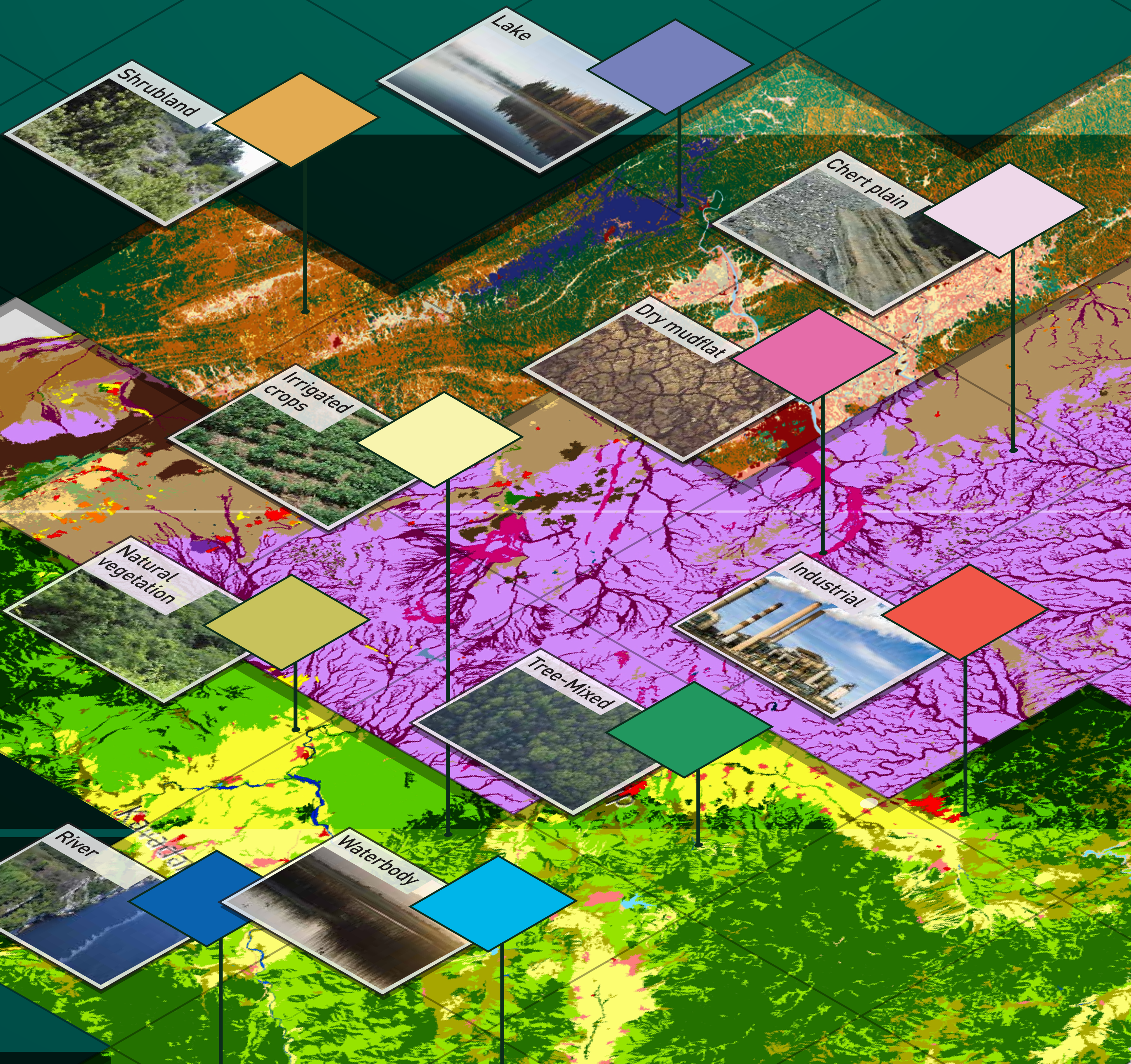
Food and Agriculture Organization of the United Nations

UNIVERSITY OF Southampton

STIIMA

Register implementation for land cover legends



Register implementation for land cover legends

Charles Douglas O'Brien (IDON Technologies Inc), Antonio Di Gregorio (FAO),

Fatima Mushtaq (FAO), Matieu Henry (FAO), Elisee Tchana (FAO), Nicola Mosca (CNR),

Palma Blonda (CNR), Chris Hill (UoS), John Latham (UoS/BSI) & Douglas Muchoney (FAO).

Published by

Food and Agriculture Organization of the United Nations

and

University of Southampton

and

Istituto di Sistemi e Tecnologie Industriali Intelligenti per il Manifatturiero Avanzato, Consiglio Nazionale delle Ricerche

Required citation:
O' Brien C.D., Di Gregorio, A., Mushtaq, F., Henry, M., Tchana, E., Mosca, N., Blonda, P., Hill, C., Latham, J. and Muchoney, D. 2021. *Register implementation for land cover legends*. Rome, FAO and UoS and STIIMA-CNR. https://doi.org/10.4060/cb5130en

ISBN 978-92-5-134560-3 [FAO]

Contents

Acknowledgements vii

Abbreviations and acronyms viii

Executive summary ix

1. Introduction 1

2. Registration approach 2

3. Standardized register structure 4

3.1. Standardized register structure 4

3.2. Unified Modeling Language (UML) classes with attributes 9

3.2.1. Classification register 9

3.2.2. Classification item class – contents of register items 11

3.2.3. Classification legend registered items 12

3.2.4. Reference and reference source 14

4. Table based implementation 16

4.1. Representation in tables 16

4.1.1. Register metadata 16

4.4.2. Register contents 16

5. Conclusion 26

6. References 27

7. Appendix A Representation of legend elements 28

7.1. CORINE class 213 - rice fields 28

7.2. CORINE class 312 - coniferous forests 29

7.3. CORINE class 111- continuous urban fabric 29

8. Appendix B Comparison metrics 30

8.1. Correspondence 31

8.2. Extensiveness 33

8.3. Integration of correspondence and extensiveness scores 33

9. Appendix C Translation guideline 36

9.1. Importance of Land Cover Legend (LCL) translation 36

9.2. Advantages of ISO standard – Land Cover Meta Language (LCML) 37

9.3. Basic guidelines for legend translation process 38

9.4. Steps for a standardized translation and outputs presentation 39

9.4.1. Preparation of legend file using LCCS3 tool 39

9.4.2. Generation of a Unified Modeling Language (UML) 40
9.4.3. Preparation of legend file using Enterprise Architect software 41

Figures

Figure 3-1 Main parts of a register 5
Figure 3-2 Classifiers and rules and their associated descriptions 5
Figure 3-3 Source reference 6
Figure 3-4 Modification of registered Item 6
Figure 3-5 Full register schema 7
Figure 3-6 Simplified register schema 8
Figure 3-7: Classification register attribute - Register metadata 9
Figure 3-8 Classification item class – contents of registers 11
Figure 3-9 Classification legend item 13
Figure 3-10 Reference to source 14
Figure 7-1 Class 213 - rice fields 28
Figure 7-2 Class 312 - coniferous forests 29
Figure 7-3 Class 111- continuous urban fabric 29
Figure 8-1 Correspondence and extensiveness metrics 34
Figure 9-1 Description of a LCL class using LCCS 36
Figure 9-2 Schema of the translated class 38
Figure 9-3 Translation of a legend into the LCCS v.3 tool 39
Figure 9-4 Validation process (LCCSv3 tool) 40
Figure 9-5 List of translated classes 40
Figure 9-6 Example of UML like schema of a translated LCL class 41
Figure 9-7 .xsd file 42
Figure 9-8 .xsd file into Enterprise Architect software 42

Tables

Table 4-1 FAO_ClassificationRegister 17
Table 4-2 FAO_ClassificationItemClass – Datasets 18
Table 4-3 FAO_ClassificationItemClass – legends schemas 20
Table 4-4 FAO_ClassificationItemClass – Classes_Primary 21
Table 4-5 FAO_ClassificationItemClass – Alternate Languages 23
Table 4-6 FAO – Reference 24
Table 4-7 FAO – Reference source 25
Table 8-1 Sample table comparing vegetated elements. 32
Table 8-2 Sample of basic LCML elements extensiveness value 33
Table 8-3 Example of output from the similarity assessment (in %) 35

Acknowledgements

The Land Cover Legend Registry (LCLR) is the result of an initiative to take a first step towards an ISO based international registry for Land Cover Legend (LCL). This open-source web-catalogue is the first world's largest repository of harmonized national, regional, and global LCL with detailed information of LCL class and is linked with original LC datasets. The following paragraphs attempt to acknowledge everyone who supported and contributed to this document.

This publication of the LCLR was made possible through the work of Food and Agriculture Organization of the United Nations (FAO) with the close collaboration of the University of Southampton (UoS) and Istituto di Sistemi e Tecnologie Industriali Intelligenti per Il Manifatturiero Avanzato (STIIMA), Consiglio Nazionale delle Ricerche(CNR).

International experts including Amit Ghosh, Antonio Martucci, Bouajila Essifi, Federica Chiozza, Francesco Tubiello, Gianluca Franceschini, Giulia Conchedda, Inge Jonckheere, Karl Morteo, Livia Peiser, Michael Tefera, Michela Marinelli, Patrizia Monteduro, Rashed Jalal, and Remi Dannunzio, are thanked and highly appreciated for their valuable feedback and review to the LCLR document. Their contribution and support have been vital for the success of this publication.

Abbreviations and acronyms

CLC	CORINE Land Cover
ISO	International Organization for Standardization
STIIMA	Istituto di Sistemi e Tecnologie Industriali Intelligenti per Il Manifatturiero Avanzato
CNR	Consiglio Nazionale delle Ricerche
FAO	Food and Agriculture Organization of the United Nations
LC	Land Cover
LCL	Land Cover Legend
LCCS	Land Cover Classification System
LCML	Land Cover Meta Language
LU	Land Use
NLCD	National Land Cover Dataset
SDG	Sustainable Development Goal
SEEA	System of Environmental-Economic Accounting
UoS	University of Southampton
UN	United Nations
UNGGIM	United Nations Global Geospatial Information Management
USGS	United States Geological Survey
UML	Unified Modeling Language

Executive summary

Land cover assessment and monitoring of its dynamics are essential requirements for the sustainable management of natural resources, environmental protection, food security, humanitarian programmes as well as core data for monitoring and modelling.

Land Cover (LC) data are therefore fundamental in fulfilling the mandates of many United Nations (UN), international and national institutions and programmes. Despite the recognition of such importance, current users of LC data still lack access to sufficient reliable or comparable baseline LC data. These data are essential to tackle the increasing concerns in regard to food security, environmental degradation, and climate change. Critically, maintaining and restoring land resources plays a vital task in tackling climate change, securing biodiversity and maintaining crucial ecosystem services, while ensuring resilient livelihoods and food security.

Moreover, the fourteen 'Global Fundamental Geospatial Data Themes' as developed and adopted by the Committee of Experts of the United Nations, under the auspices of the United Nations Global Geospatial Information Management (UN-GGIM) have been extensively implemented by the global geospatial community and beyond, with "Land Cover and now a new and evolving standard and metalanguage on Land Use" becoming part of this process.

To further support their implementation, ISO/TC 211 recognizing the urgent need for an agreed upon and improved Land Cover and Land Use meta-language by further evolving the ISO 19144-2:2012 "Geographic information - Classification systems -- Part 2: Land Cover Meta Language (LCML)" standard and a developing a new meta language standard for Land Use, ISO 19144-3, and ISO 19144-4 on Registers.

Importantly, LC information has been playing an important role in the monitoring and attainment of the achievement of Sustainable Development Goals (SDGs) and their related targets. Ensuring where possible that all countries have LC maps that are suitable for and consistent with the requirements for SDG monitoring: alignment with global Land Cover Classification System (LCCS) and have sufficient spatial detail and are interoperable and the System of Environmental-Economic Accounting (SEEA) aligned.

The Land Cover Classification System (LCCS) was developed to respond to the need for consistent and reliable assessment of LC resources through the use of standards, definitions, classifiers, methods, approaches, semantic interoperability and preparation of interoperable, scalable and interchangeable LC products at various levels. FAO methodology was perpetuated in many areas of the world at global, regional, and country levels and in many fields of study.

The objective of this era of work was to:

- establish a global collaboration for developing a fully harmonized approach;

- to make the required reliable and comparable land cover and land cover change data accessible to local, national and international initiatives;
- to support the production and dissemination of reliable, consistent and updated land cover information;
- to strengthen the capacity of stakeholders in developing countries.

The overall objective of LCML has always been to ensure the semantic interoperability of LC classifications, i.e. the ability of systems to exchange LC data with unambiguous and shared meaning.

In particular, LCML has acted as a method to bring the LC community together to create a common understanding of LC nomenclatures with the aim of producing global, regional, and national datasets capable of being reconciled at different scales and different levels of details, in different geographic locations. More Importantly LCML has been used to standardize the attribute terminology rather than the final classes. It was used to create a set of standard diagnostic attributes (LCML basic objects, their properties and characteristics) to describe different LC features.

Traditionally, a LC class was defined by LCML object(s) «biotic» or «abiotic», organized in horizontal or vertical patterns and other temporal relationships. LCML objects, customized by adding additional attributes of Land Use (LU) and LC characteristics (climate, landform, topography, etc..). Thus, a LC class was defined by the object(s) that it contained, and their spatial and temporal relationships. LCML described in a Unified Modeling Language (UML) diagram. All rules and conditions of the meta-language were installed in a user-friendly software (LCCS3) that allowed the creation, edit and export of LCML-based legends. National and regional LC databases produced in FAO were integrated/ adopt into the LCML standard easily. The separation and removal of the LU attributes will now make the LC classification more pure, in regards to its attributes, while evolving, in parallel, a commensurate LU metalanguage 19144-3.

From this period of application, it was determined that there are many different Land Cover Legend (LCL) and LC classification schemes used to classify LC. However, in many practical situations, the rigidity / vagueness of traditional LC classifications based on names and relative text description limited the interoperability between LCL and contributed to the inconsistency of LC information at regional and global levels. Therefore, it was also determined that there is a need to develop a platform, with updated LC datasets, along with harmonized LCL to address regional and global and monitoring requirements and a variety of environmental issues.

This report now presents an overview of the structure of a set of registers and an associated registry for the description of the LCL used by FAO and associated other international organizations. This set of registers is described both in an abstract manner, in alignment with the ISO standards, for the registration of geographic information and in alignment with ISO 19144-1 and also as a set of flat tables that may be implemented in a spread sheet program of a relational database.

The high-level design is described as a UML schema. However, the implementation is in terms of simple tables where several of the UML classes may be represented in a single table. The legends and their classes semantic content and schemas are described with an object-oriented procedure using the LCML as defined in the joint FAO / ISO TC211 standard (ISO 19144-2).

Three distinct types of information are registered. These are:

1. A list of Land Cover (LC) datasets held by FAO, or associated organizations.
2. A list of the classifications system legends (or schema) that are used within these LC datasets; and
3. A list of classification classes used within these legends.

This development aims at:

i) contributing to research and education on LC representation,

ii) supporting the use of geospatial information in support to eradicate poverty (SDG 1), and end hunger and all forms of malnutrition (SDG 2), good health and well-being (SDG 3), clean water and sanitation (SDG 6), sustainable cities and communities (SDG 11), climate action (SDG 13), and life on land (SDG 15) through FAO Hand in Hand initiative and

iii) to contribute to the ISO TC211 for the future development of the proposed standard 19144-4 classification systems.

This document presents the standardized register structure of LCL registry, description of attributes used in classification register, table-based implementation for populating the LC datasets/ LC legends /LCL classes data, representation of LCL elements, comparison metrics for LCL classes and guideline for translating LCL into LCCS.

1. Introduction

The representation of LC, and the associated representation of Land Use (LU), is an important source of information for decision makers, planners, researchers, and private interests about terrestrial ecosystems. The United States Geological Survey (USGS) identifies the four important application areas as being (USGS, 2011):

- **Land management** – LC data are crucial to the development of improved land management practices and sustainable land use practices;
- **Habitat loss** – knowledge of where tracts of intact habitat occur is key to targeting the most effective conservation measures;
- **Climate** - information about LC is critical to understand climate change, as a driver for climate modelling and where to focus climate mitigation and conservation actions;
- **Biodiversity** – data on terrestrial ecosystems is key to helping land conservation decision makers better match biodiversity goals to land protection programs and activities.

Many nations have similar requirements and the United Nations (UN) compiles global data sets and assists developing and other nations in producing LC and LU data for their national and local areas. Creating global LC data is very difficult because data are gathered for different purposes through different survey and acquisition projects, and various nations and institutions use different classification schemes to represent these data. There are, accordingly, many different legends used to classify data. However, the integration of data is required to address regional and worldwide requirements.

FAO, together with ISO technical committee 211 on geographic information / geomatics, has developed a series of standards for the management of LC and related classification systems. The standard ISO 19144-2:2012 (ISO, 2009b) specifies a LCML that allows different LCCS to be described in an ordered and standard process. It provides a common reference structure for the comparison and integration of data for any generic LCCS (Gregorio & Jansen, 2000). The standard ISO 19144-1:2009 (ISO, 2009a) establishes the structure of a geographic information classification system, together with the mechanism for defining and registering the classifiers for such a system.

This document describes the structure of a set of registers and an associated registry for the description of the LCL used by FAO and associated other organizations. It is based on the registry concepts identified in ISO 19144-1 and makes use of the descriptive metalanguage described in ISO 19144-2 (Appendix B). This document also provides comparison metrics for assessing LC similarity (Appendix B) and concise guideline to support translations of conventional legends into the object oriented LCML (ISO standard 19144-2) using FAO LCCSv3 toolbox (Appendix C).

2. Registration approach

FAO maintains, prepares and use a wide range of LC datasets, both globally and developed for specific regions and nations. In addition, it uses data developed by other -organizations or countries using specific legends and systems. Most of FAO LC datasets apply a common taxonomy based on FAO LCCS (ISO, 2009b). LC data from other sources may also be used by FAO; however, these LC datasets may be represented in one of many other well used international classification systems.

The purpose of this document is to define a register of LC legends used by FAO and other organizations together with their representation using the standardized LCML. Legends that have not been represented using LCML may also be included in the register but marked as not yet translated (Appendix C)). This register is deliberately implemented as simple flat tables that can be easily manipulated using common tools such as Microsoft EXCEL, or simple relational database tools. However, the intent is for this register to comply with the register structure defined in ISO 19144-1. To this end, some of the registration classes defined in 19144-1 have been combined into single tables. Since they are represented as flat tables no automatic mechanism is defined to manage this set of registers as a registry. Maintenance must be done manually by editing the tables. However, this does not preclude some future implementation being able to import these flat tables and produce a more manageable database.

The UN, moreover, supports multiple national languages. Also, there is a need for different national sources to manage the data in their own language. There are several different approaches allowed in the ISO standards to handle this requirement. One possible approach is to use the LOCALE mechanism in ISO TC211 which allows a single character string to carry multiple sub-text strings together with language tags. This approach is very useful in a communications system where there is a need to have only one message containing multiple language text. However, managing many languages in a single register is a maintenance problem. Therefore, an approach which uses multiple language tables in a federated register system is proposed. In effect this means that there will be parallel tables for alternate languages.

However, the descriptions in different national languages may get out of step with the primary register(s) maintained in a single language, since the alternate language information will likely be managed by different people. This independent management can be undertaken using the federated register concepts permitted using the standard ISO 19135-1 Geographic Information - Procedures for Item Registration (ISO, 2015). Essentially, this is a master register that contains a description of each registered item with the descriptive text in one language (in this case English), and the technical description using LCML elements. In addition, there may be one or more additional registers (or pages/tables in EXCEL or a database) that link to the same unique registered item ID and contain the descriptive information in other languages. It should be noted that the national language information does not have to be entered at the same time, nor does it need to be inclusive. Only those descriptions deemed necessary by the national

language manager needs to be included. This makes management of multiple languages much simpler.

Like all ISO 19135 based registers, this register contains a certain overhead to allow for rigorous maintenance. Once an item is entered in the register it remains forever. The unique registered item IDs are permanent IDs (PID). If a registered item is edited or replaced (or rarely deleted), it is replaced with a new entry with a new unique registered item ID and the previous item's status is changed from VALID to SUPERCEEDED (or rarely to DELETED). The status may also be identified as PROPOSED to allow for the recording of proposed entries. Since not all proposals are accepted, proposals may also be marked as Proposal Not Accepted. This code list of status states may be customized by FAO register to suit its management processes. The important thing is that elements in the register retain a permanent unique ID and a status.

There is also a logical pointer (or equivalent) established from the superseded entry to the new entry. This means that any external use of the register that points to the permanent unique registered item ID remains valid even if the register is modified and the external national language information sub registers continue to point back to the original permanent unique registered item ID. The system is robust and maintenance of all parts of the system do not have to be undertaken at the same time. In the implementation described below using flat tables the logical pointer is implemented using a secondary key. That is, there is no actual pointer, which would be difficult to maintain. Rather a pointer can be dynamically constructed when needed by a search in the table using a unique label for the registered concept, the validity flag and the date.

3. Standardized register structure

This section describes the register structure as derived from the ISO standards on registration and on classification systems structure.

3.1. Standardized register structure

The set of classifiers that compose a legend may be managed by registration. The approach to registration is defined in two ISO standards,

The standard ISO 19144-1 states:

"A classified object is a subtype of feature attribute (as defined in the standard ISO 19110 Geographic Information — Methodology for feature cataloguing (ISO, 2016)). Like a feature attribute, a classification may be predefined with the definition stored within a dictionary. The dictionary may contain all the classifiers available within the classification scheme. New classifiers may be added to the classification scheme and the definitions of certain classifiers may be modified. An appropriate manner of managing such changes is through the process of registration."

The register schema is derived from the standard ISO 19135-1 and is given in ISO 19144-1. It consists of three main parts:

1. Metadata describing the whole register.
2. A description of the meaning of each item in the register (e.g. which columns are in the tables).
3. The registered items (with information for each column in a table). The ISO register schema permits multiple tables of registered items.

   In this case there are three registered item tables,

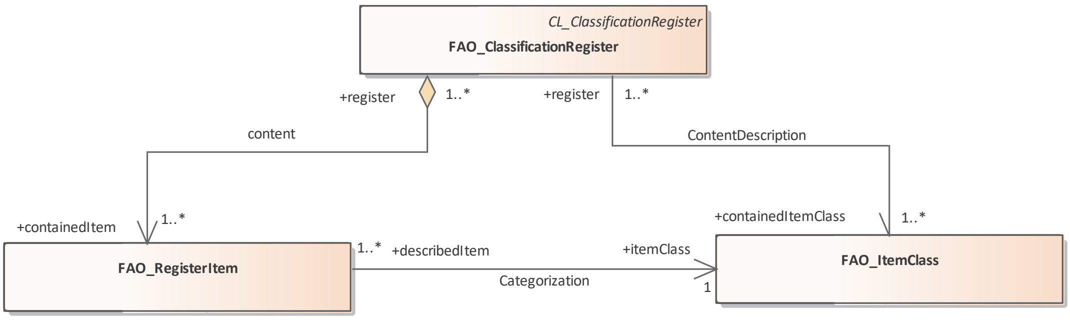
   - A list of LC datasets held by FAO or associated organizations.
   - A list of the Classifications System Legends (or schema) that are used within these datasets; and
   - A list of classification classes used within these legends.

These tables are interrelated and make reference to each other in a hierarchical manner from the dataset to legend/schema to class levels. Each of these three parts of a register may be represented as a table Figure 3-1.

In accordance with the standard ISO 19144-1 a legend or classification system is defined in a multi-part register that may include the description of classifiers and rules Figure 3-2.

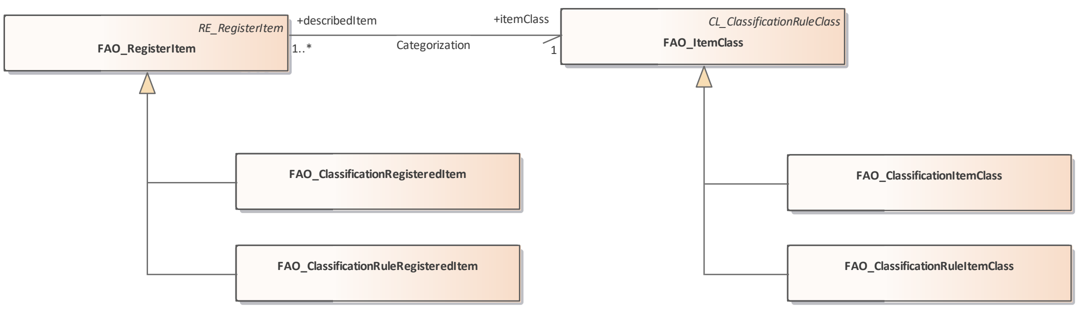
- classifier – definition and any associated code and/or name.
- rules describing the relationship between registered legend items or the relationship of classified object types to geometry.

Figure 3-1 Main parts of a register



This version of FAO LCL registry does not support classification rule entries. Rules are equivalent to operations on features and describe how the features are permitted to change state or perform an operation. For example, a dam may be adjusted to alter the water level downstream. This operation is described by a rule. Rules can be also added to a register in the future and may be important for LU classifiers, but they are not used at this time.

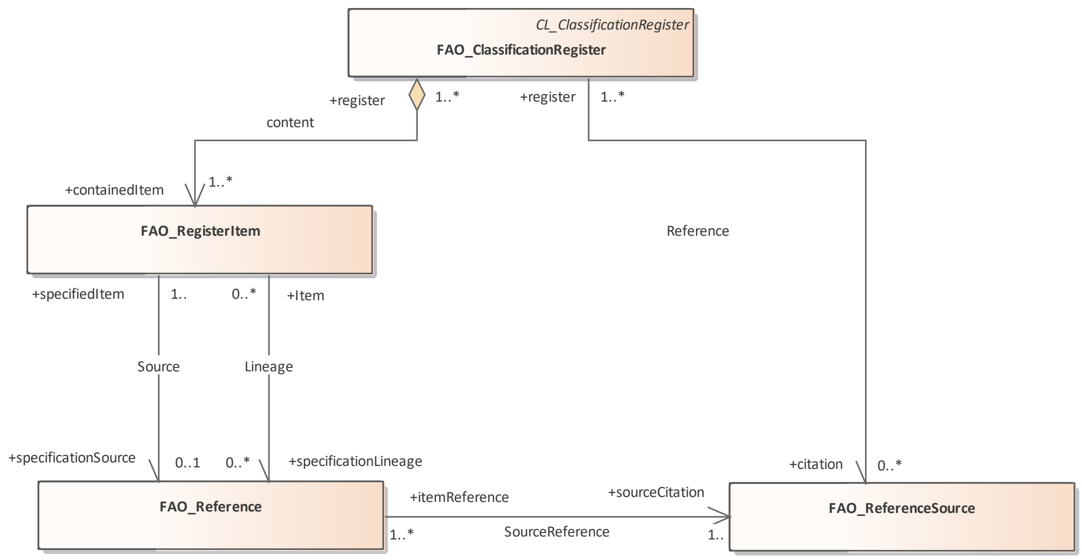
Figure 3-2 Classifiers and rules and their associated descriptions



In addition, there may be a list of sources (with full citations) and references from each registered item to its source, and possibly a historical lineage back to previous sources. This is shown in Figure 3-3. To support sourcing, there are two tables, a list of sources, and a second list giving detailed citations for each source. Lineage is the historical list of source citations. In this version of FAO LCL registry the lineage capability is not used.

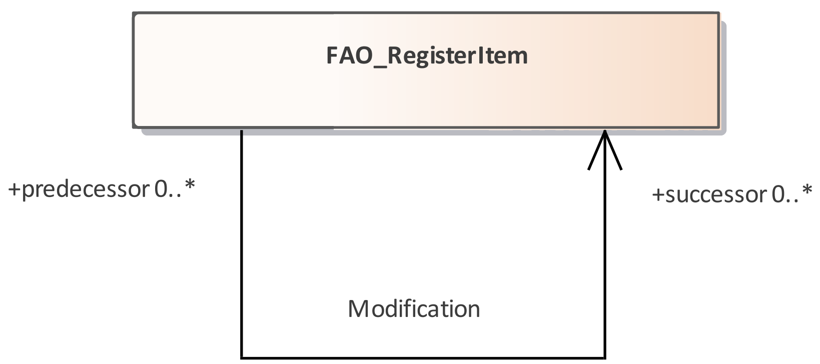
A registered item remains permanently in the register. If the registered item is modified, it is marked as SUPERSEDED (or rarely as DELETED), and a pointer or equivalent mechanism is included to identify the new item. In the model in the ISO standard this structure consists of a simple validity field and a pointer field in the register item table.

Figure 3-3 Source reference



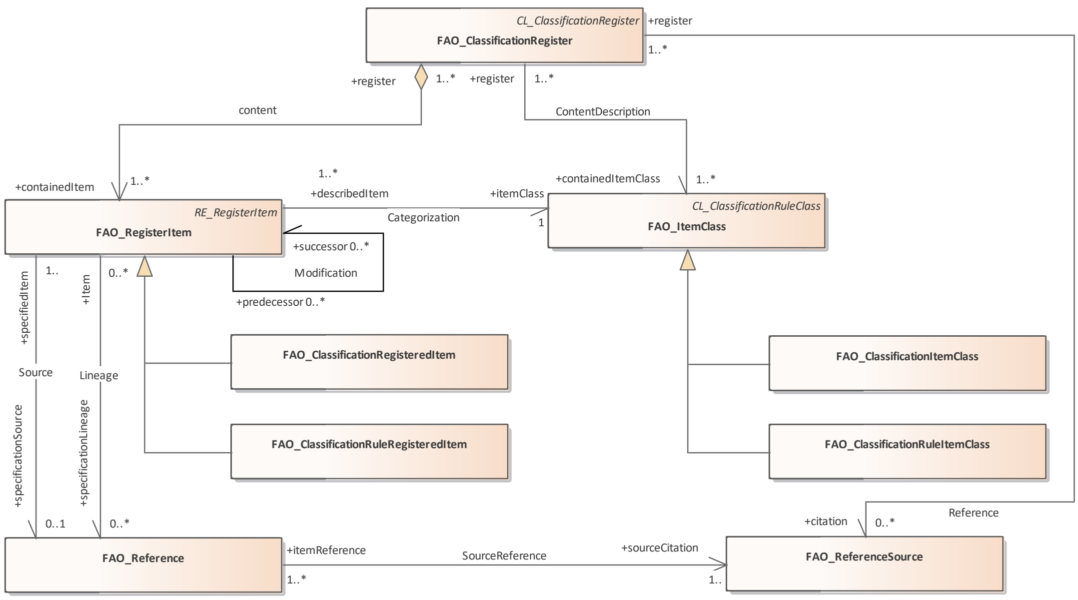
This self-referencing is shown in Figure 3-4. However, this can be implemented in a table or database implementation as a secondary ID (or name) so that a simple search can generate the pointer. This secondary permanent ID approach is used in FAO LCL register to keep the implementation to flat tables and avoid the need for potential multi to multi relationships.

Figure 3-4 Modification of registered Item



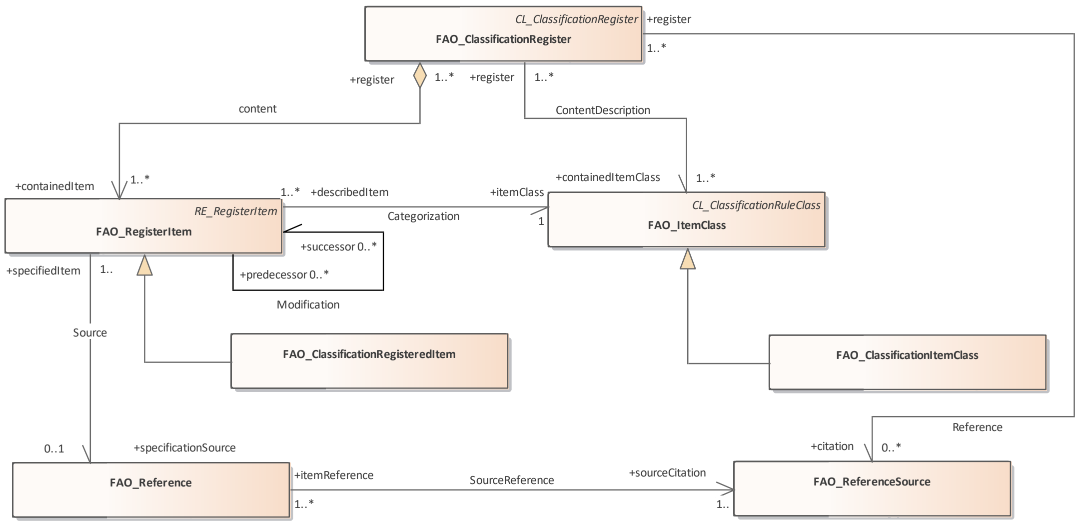
The full register schema, as derived from the standard ISO 19144-1 is shown in Figure 3-5 .

Figure 3-5 Full register schema



For FAO LCL register this schema can be simplified. Rules are not being registered (although they could be in the future). Also, lineage is not being recorded. In theory, one registered item could be superseded by more than one new item, however this is very rare and therefore the implementation will need only a single pointer (or secondary ID). An exception can be handled through notes in the description. Therefore, a flat table structure can be used to implement this register structure. This is illustrated in Figure 3-6.

Figure 3-6 Simplified register schema



There are two approaches to implementing multiple languages in a federated register. In a distributed federated register, the whole register schema can be duplicated in the alternate language, where the secondary register references the same permanent unique registered item ID from the master register. This secondary register would also have its own permanent unique ID for every registered item. This approach allows the registers in the federation to get out of synchronization, yet still functions correctly. Registers in separate registries are allowed to get out of synchronization. The only requirement is for one registry to be the master, and for any parallel registers in other languages to report any changes back to the master registry. This approach might be used where the responsibility for local languages, and even parts of the register are distributed and where the register is physically distributed over multiple computers at different locations. This distributed federated register approach is beyond the requirement for FAO LCL register that can be implemented in a single location.

A simpler approach to handling multiple languages is to have a second or multiple RegisteredItem tables in the same registry, each addressing different languages. The whole register structure would be managed together (but it is still allowed for additional language tables to get out of synchronization without fault). In this case only the RegisteredItem table needs to be duplicated. This simpler approach is used in FAO LCL register. A field will be used in the RegisteredItem table to identify the language. Therefore, only one alternate language table is needed addressing all alternate languages (for each of the three RegisteredItems (LC datasets, LC legends and LCL classes) being registered.

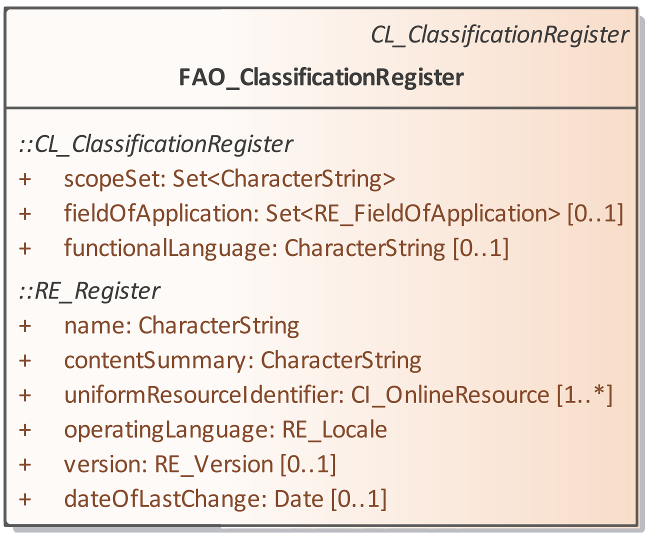
3.2. Unified Modeling Language (UML) classes with attributes

Each of the UML classes[1] has a set of attributes. Many of these attributes are inherited from the ISO standards, and many of these are optional. This document endeavors to make use of the ISO defined UML classes and attribute names. The following figures illustrate each of the UML classes and their attributes. Relationships between UML objects can be treated as pointer attributes in a table-based implementation.

3.2.1. Classification register

The root of the register system is the classificationRegister object, as shown in Figure 3-7, which inherits from the ISO standardized UML class CL_ClassificationRegister from ISO 19144-1 and RE_Register from ISO 19135-1. This object is essentially metadata about the whole register system.

Figure 3-7: Classification register attribute - Register metadata



- **name** - a character string that is used to uniquely identify a register
- **contentSummary** - a character string containing a general statement of the purpose

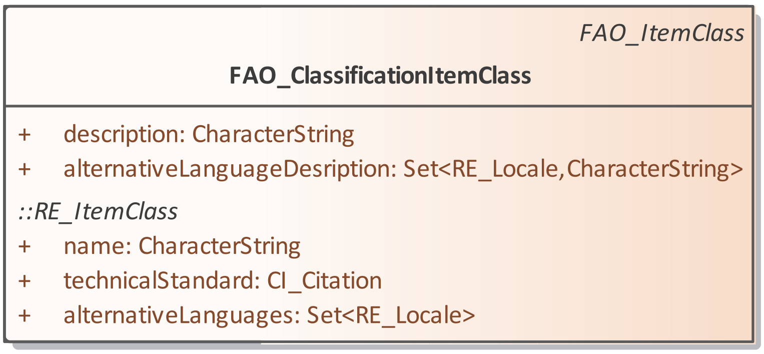
[1] There is a potential confusion between the use of the word "class" in the UML modelling language and "class" in a classification system. These are different uses of the same word in different contexts. However, in the description of a UML model of a classification system both meanings appear. To avoid potential confusion in this document the term UML will be used as an adjective to clearly identify all of the uses of the word where UML class is meant. A UML class is a template for an object, and where logically correct the term UML object will be used.

- **uniformResourceIdentifier** – URI of the online resource for a register. This field is required by the standard, but if the register is not online it can be left blank.
- **operatingLanguage -** language, country information and character encoding for the proper interpretation of the content of character strings in the register.
- **version** – state in the life of the register. This is required if the dateOfLastChange is not supplied.
- **dateOfLastChange -** the date of the most recent change to the status of an item in the register was made.
- **scopeSet -** a set of scope elements represented as character strings used to describe subject domains of the registered items. The scopeSet may be used as the basis for creating metadata for submission to search engines.
- **fieldOfApplication** - a set of character strings used to describe the kinds of use of the registered items which may be used as the basis for creating metadata for submission to search engines.
- **functionalLanguage** – a specification of the notational system used for formal definitions of rules used to relate classified objects within a classification system. This is optional and only used when rules are registered. This field is not required in FAO LCL register because rules are not registered.

3.2.2. Classification item class – contents of register items

The second UML object in the structure is the UML class classificationItemClass. This object is essentially a table of contents for the register system. This UML class inherits from the standard ISO 19144-1 which inherits from the UML class RE_ItemClass from the standard ISO 19135-1. The two attributes description and alternativeLanguageDescription have been added in FAO LCL register.

Figure 3-8 Classification item class – contents of registers



For each registeredItem attribute (column in a table implementation) there will be a row in this object explaining the attribute name and its description and the set of alternative languages used. In theory the ISO standard allows that there be a formal standard or less formal specification defining the attributes for each UML class and their meaning. However, for FAO LCL register the technical standard referenced is the ISO 19144-1 standard which does not provide sufficient detail; hence the two attributes description and alternativeLanguageDescription have been added. In addition, in this FAO LCL register an attribute has been added which contains the description. This makes the implementation self-contained. The additional attribute alternativeLanguageDescription allows an alternate language description by making use of the ISO LOCALE structure (from ISO 19115-1 metadata standard, which allows an arbitrary string of text in multiple languages, each preceded by a language code.

- **name** – the character string "Legend Class".
- **technicalStandard** - the character string "ISO 19144-1:2009, Geographic information – Classification Systems – Part 1, Classification system structure"
- **alternativeLanguages** - a set of language codes each specifying a language used in the register.
- **description** – a description of the meaning of the attribute

- **alternativeLanguageDesription** - a description of the meaning of the attribute in alternate languages, where each language is preceded by a language tag in accordance with the ISO LOCALE mechanism.

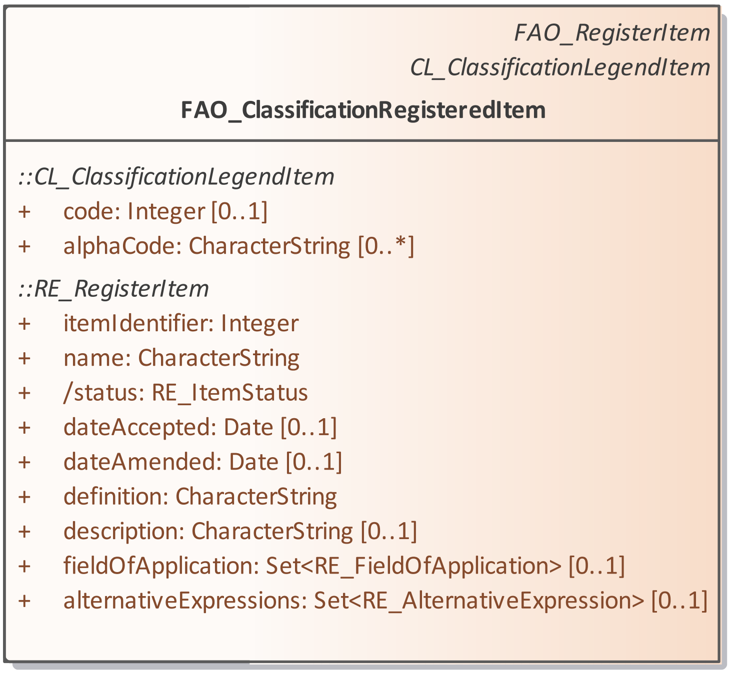
3.2.3. Classification legend registered items

The third UML object in the structure is the UML class classificationLegendRegisteredItem. This UML class is a template for each of the Registered Item tables in the system. In FAO LCL register there are three Registered Item Tables, (**LC datasets, LC legends/schemas, LCL classes**). This UML class inherits from the UML class CL_ClassificationLegendItem from the standard ISO 19144-1 which inherits from the UML class RE_RegisteredItemClass the standard ISO 19135-1. The attribute names derive from these parent standards.

Since three different items are registered, there are three separate tables, one each for **LC datasets, LC legends/schemas,** and **LCL classes.** These tables all have the same basic form as shown in Figure 3-9. In addition, each of the three specific Registered Item tables contain specific attributes pertaining to that type of item. The details of these additional attributes are given in the classificationItemClass.

- **itemIdentifier** – (PID) a positive integer (i.e., greater than zero) that serves as a unique permanent identifier for the entry and is used to uniquely denote that item within the register and is intended for information processing. Values shall be assigned sequentially in the order in which items are proposed for entry into the register. Once a value has been assigned, it shall not be reused. Rows of the table are never deleted and are permanently associated with the same PID.
- **alphaCode** – (Label) an alphanumeric identifier permanently associated with the concept of a registered item. If a registered item is changed, a new entry is placed in the table with a new PID but with the same label. A pointer from an older entry to a new entry can be derived from the combination of the label, status, and date. Only one alphaCode / Label is used, and it is required.
- **code (IdLC)** - a numeric identifier permanently associated with the concept of a registered item.
- **status**- the registration status of the registerItem, (valid, superseded, deleted, proposed, rejected).
- **date** – the reference date for the entry is the dateAmended; that is, the date at which the entry was placed in the register. Only one date field is used. The dateAccepted is the same as the date amended for the first entry of an element in the register and can always be calculated by searching for the earliest date for a given unique alphaCode (Label)

Figure 3-9 Classification legend item



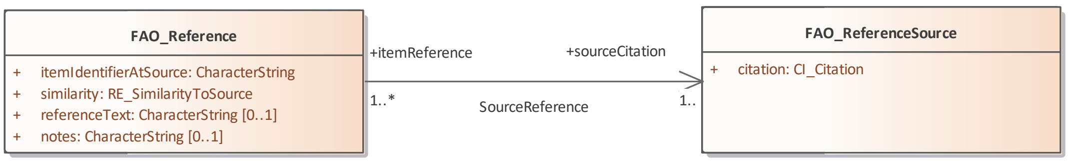
- **name** – a character string containing a compact and human-readable designator that is used to denote a register concept. The name shall be unique within a register. Multiple items of the same item class may use the same value for name but only one such item may have a status of 'valid'.
- **definition -** a character string containing the definition of the concept embodied by that item
- **description -** a character string containing a statement of the nature, properties, scope, or non-essential qualities of the concept that are realized by the item but are not specified by the definition element.
- **fieldOfApplication -** a set of character strings, each of which describes a kind of use of the item, which may be used as the basis for creating metadata for submission to search engines. This attribute is not implemented in FAO LCL register
- **alternativeExpressions – an optional character string** specifying an alternative name and optionally additional information in a locale (language or dialect or character set) different from that of the register. This can be used to replace the use of an alternative language federated register for some simple situations. In FAO LCL register this is implemented as a secondary table.

Additional attributes are defined to fully describe the registered item. These are unique to the type of registered item (LC database, LC legend/schema, LCL classes) and are described in the associated classificationItemClass (contents).

3.2.4. Reference and reference source

The fourth and fifth UML objects in the structure are the UML class Reference and referenceSource. The refenceSource object is a list of citations. The Reference object is a pointer to the citation. It also allows additional information describing the reference.

Figure 3-10 Reference to source



The reference object describes each reference. There may be multiple references to the same reference source. In this implementation the similarity to source attribute is optional and by default unspecified.

- **itemIdentifierAtSource** – A character string containing a source identifier
- **similarity** – similarity to source from the code list: identical, restyled, contextAdded, generalization, specialization, unspecified.
- **referenceText** – an optional character string describing the reference.
- **notes** - an optional character string containing notes.

The ISO standard permits the metadata field CI_Citation to be used to reference a source. This is extremely flexible and allows many (too many) possibilities, most of which are optional. Citations may include a large number of options and may be recursive. Recursion cannot be implemented in a flat table implementation and only a very small number of the allowed possibilities are required.

- **citation** – a citation to the source from the ISO metadata item CI_Citation (which provides a formal description of the citation).

The ISO CI_Citation makes use of the ISO CI_ResponsibleParty structure, all of which is optional. In this register the author's organization and publisher's organization are simply identified by an organization code eliminating the need for a general citation. The external reference to the cited source is done through a URL containing a referenced title, referenced / folder, and link.

- **refId** – Identifier of the source.
- **author** – optional name of the author
- **authorOrganization** – name of the author's organization

- **authorOrgCode** – code representing of the author's organization
- **publisherOrganization** – name of the author's organization
- **publisherOrgCode** – code representing of the author's organization
- **publishedYear** – reference date for the cited reference
- **refFileTitle** – title of the referenced source file as part of a URL
- **refFileFolder** – folder containing the referenced source file as part of a URL
- **refLink** – link to the referenced source file as part of a URL

4. Table based implementation

The structure described above can be implemented in many different ways, including a representation on an object-oriented database or represented as XML data for interchange. However, the proposed FAO implementation is as simple flat tables. There would be a table for each of the objects identified above as shown in Figure 3-7 and Figure 3-8, three tables for the three instances of the classificationRegisteredItem tables for LC datasets, LC legends/schemas and LCL classes. There will also be three auxiliary tables to these classificationRegisteredItem tables to carry the alternate language information for LC datasets, LC legends/schemas and LCL classes. References and sources will be handled in two tables. In addition, there will be a table for the code list for country codes.

An implementation in Microsoft EXCEL could make use of separate sheets (tabs) for each table, and pointers could be implemented as links between sheets. The relationships shown in the schema in Figure 3-6 would also need to be represented as columns with links between sheets.

Many of the tables can be considered as structural overhead. They are necessary to fully define the register and comply with the standard. The root table is FAO_ClassificationRegister table.

4.1. Representation in tables

Each of the tables are shown below. The order of the entries and the names of the attributes for the register items (table columns) have been changed in the register content tables in order to make the database more readable. The original attribute name, as derived from the ISO standard is also included in square brackets "[ ]" to allow traceback to the UML models.

4.1.1. Register metadata

The table FAO_ClassificationRegister is the root table of the register and provides the metadata for the register. The attribute names derive from the ISO standards. These attribute names could be changed in an implementation, but it is best to use the ISO names to minimize confusion.

4.4.2. Register contents

The table FAO_ClassificationLegendClass describes the meaning for each attribute in the table FAO_ClassificationLegendItem. In this implementation the table is called contents and also includes a description of all the other tables in the register set. The ISO defined attribute technical standard is cited for the whole register system and is ISO 19144-1:2009, therefore, it is not included in each table.

The description in an alternate language is optional. Several alternate languages may be used within the ISO LOCAL mechanism. That is each language may be preceded with a language code and optionally country code for dialects in brackets. The attributes alternateLanguage code and alternateLanguageDescription are included in FAO register; however, it is not shown in the tables below for simplicity.

Table 4-1 FAO_ClassificationRegister

ID	Attribute	Attribute Description	Value
1	name [name]	Uniquely identify a register	FAO Land Cover Legend Registry (LCLR).
2	contentSummary [contentSummary]	Statement of the purpose	Documentation of a set of Classification legends and schemas used by FAO and associated organizations.
3	uniformResourceIdentifier [uniformResourceIdentifier]	URI	Not Applicable - blank
4	operatingLanguage [operatingLanguage]	Language	English
5	version [version]	State in the life of the register ...	Under development.
6	dateOfLastChange [dateOfLastChange]	Date of the most recent change	Not Applicable - blank.
7	scopeSet [scopeSet]	Scope	Farming, biota, environment, geoscientificInformation, imageryBaseMapsEarthCover, planningCadastre, society, structure, transportation, utilitiesCommunication
8	fieldOfApplication [fieldOfApplication]	Kinds of use of the registered items	LC, LU, agriculture, environmental planning, UN sustainability goals.
9	functionalLanguage [functionalLanguage]	Functional language for rules	Not applicable.

1. FAO_ClassificationLegendClass - Contents

Some of the attributes defined in this register are required by the ISO standards, whereas others are particular to this application. These are described below. The order of the entries and the names of the attributes for the register items (table columns) have been changed in the register content tables in order to make the database more readable. The original attribute name, as derived from the ISO standard is also included in square brackets "[ ]" to allow traceback to the UML models.

Each of the Attribute names and the meaning of each attribute are described. The optional ISO defined attributes dateAccepted and dateAmended are not used, with their function folded into the attribute date. The ISO defined attributes description and fieldOfApplication are not used. An optional sourceReference attribute is included in the classificationClass tables that acts as a pointer to the Reference table. The ISO defined attributes alternateExpressions is not implemented for the datasets table.

The mandatory ISO defined attribute definition is implemented indirectly using a set of additional attributes to fully identify and describe the dataset, legend/schema or class in the tables below.

Table 4-2 FAO_ClassificationItemClass – Datasets

Name	Description
dsItemIdentifierReg [itemIdentifier]	(PID) a positive integer (i.e., greater than zero) that serves as a unique permanent identifier for the dataset and is used to uniquely denote that item within the register and is intended for information processing.
dsDate [date]	The reference date for the entry is the dateAmended; that is, the date at which the entry was placed in the register. The dateAccepted is the same as the date amended for the first entry of an element in the register.
dsName [name]	A character string containing a compact and human-readable designator that is used to denote the name of the dataset. The name shall be unique within this register. Multiple items of the same item class may use the same value for name but only one such item may have a status of 'Valid'.
dsAlphaCode [alphaCode]	(Label) an alphanumeric identifier permanently associated with the concept of a registered item. The alphaCode corresponds to the name and so the alphaCode shall be unique within this register. Multiple items of the same item class may use the same value for alphaCode but only one such item may have a status of 'Valid'.

Name	Description
dsStatus [status]	The registration status of the registerItem from the code list statusCode.
dsId [code]	Unique numeric identifier of the dataset.
lgId [legendID]	Reference to the legend used.
refId [referenceID]	Reference to the ID type used.
dsCountry [country]	Name of the country or region.
dsM49CountryCode [M49CountryCode]	Country code per UN M49.
dsIso3CountryCode [Iso3CountryCode]	Country code per ISO 3166 three-character codes.
dsYear [year]	Reference year for the data.
dsSubCountryCode(GADM)	Flag indicating whether the data applies to part of a country.
dsSubCountryRegion [subCountryRegion]	Name of the part of a country.
dsUNRegion [UNRegion]	Descriptor of the UN Region.
dsSubUNRegion [subUNRegion]	Descriptor of the subregion.
dsGeonetwork [geonetwork]	Indicator of whether the data is available on Geonetwork.
dsGeonetworkIdentifier [geonetworkIdentifier]	Identifier of the dataset on Geonetwork.
dsURL [URL]	URL to access the data from FAO map catalogue.
dsFAOSupported [FAOSupported]	Indicator of whether the data is FAO supported.
dsDataAccessibility [dataAccessibility]	Indicator of whether the data is accessible.
dsDataFolder [dataFolder]	Identifier or the folder that the data is available in.
dsDataLabel [dataLabel]	Label used to describe the data set.

Name	Description
dsSourceReference [sourceReference]	Optional pointer to a reference.

Table 4-3 FAO_ClassificationItemClass – legends schemas

Name	Description
lgItemIdentifierReg [itemIdentifier]	(PID) a positive integer (i.e., greater than zero) that serves as a unique permanent identifier for the legend/schema and is used to uniquely denote that item within the register and is intended for information processing.
lgDate [date]	The reference date for the entry is the dateAmended; that is, the date at which the entry was placed in the register. The dateAccepted is the same as the date amended for the first entry of an element in the register.
lgName [name]	A character string containing a compact and human-readable designator that is used to denote the name of the legend/schema. The name shall be unique within a register. Multiple items of the same item class may use the same value for name but only one such item may have a status of 'Valid'.
lgAlphaCode [alphaCode]	(Label) an alphanumeric identifier permanently associated with the concept of a registered item. The alphaCode corresponds to the name and so the alphaCode shall be unique within this register. Multiple items of the same item class may use the same value for alphaCode but only one such item may have a status of 'Valid'.
lgStatus [status]	The registration status of the RegisterItem from the code list StatusCode.
lgId [code]	Unique numeric identifier of the legend.
DsId [datasetID]	LC Identifier.
refId [referenceID]	Legend ID for references.
lgCountry [country]	Name of the country or region.

Name	Description
lgM49CountryCode [M49CountryCode]	Country code per UN M49.
lgISO3CountryCode [ISO3CountryCode]	Country code per ISO 3166 three-character codes.
lgYear [year]	Reference year for the legend.
lgSubCountryCode(GADM)	Flag indicating whether the data applies to part of a country.
lgSubCountryRegion [subCountryRegion]	Name of the part of a country.
lgUNRegion [UNRegion]	Descriptor of the UN Region.
lgSubUNRegion [subUNRegion]	Descriptor of the subregion.
lgGeonetwork [geonetwork]	Indicator of whether the legend is available on Geonetwork.
lgFAOSupported [FAOSupported]	Indicator of whether the legend is FAO supported.
lgDataAccessibility [dataAccessibility]	Indicator of whether the legend is accessible.
lgType [legendType]	Description of the Legend Type.
lgDataFolder [dataFolder]	URL link to the legend.
lgDataLabel [dataLabel]	Label of LC legend data.
lgSourceReference [sourceReference]	Optional pointer to a reference.

Table 4-4 FAO_ClassificationItemClass – Classes_Primary

Name	Description
clPItemIdentifierReg [itemIdentifier]	(PID) a positive integer (i.e., greater than zero) that serves as a unique permanent identifier for the class description entry and is used to uniquely denote that item within the register and is intended for information processing.

Name	Description
clPDate [date]	The reference date for the entry is the dateAmended; that is, the date at which the entry was placed in the register. The dateAccepted is the same as the date amended for the first entry of an element in the register.
clPName [name]	A character string containing a compact and human-readable designator that is used to denote the name of description of the class. The name shall be unique within a register. Multiple items of the same item class may use the same value for name but only one such item may have a status of 'Valid'.
clPAlphaCode [alphaCode]	(Label) an alphanumeric identifier permanently associated with the concept of a registered item. The alphaCode corresponds to the name and so the alphaCode shall be unique within this register. Multiple items of the same item class may use the same value for alphaCode but only one such item may have a status of 'Valid'.
clPStatus [status]	The registration status of the RegisterItem from the code list StatusCode.
clPId [code]	The unique numeric identifier of the class.
clPDefinition [definition]	A character string containing the definition of the class.
clPDescription(optional) [description]	An optional character string containg a more detailed description of the class.
clPLgType [legendType]	Indicator of the type of legend used.
clPCode [classCode]	Code for LCL class.
lgName [name]	Name of the legend.
clPCountry [country]	Country or region to which this legend applies.
clPM49CountryCode [M49CountryCode]	Country code for the country or region from ISO 3166 three-character country codes.
clPISO3CountryCode [ISO3CountryCode]	Country code per ISO 3166 three-character codes.
lgId [legendID]	Legend Id for references.
clPFile [legendData]	URI for legend.

Name	Description
dsId [datasetId]	LC identifier.
clPSourceReference [SourceReference]	Optional reference to source.

2. FAO_ClassificationItemClass – Alternate Languages

This Table of Contents (registeredItemClass) table addresses all of the possible optional alternate language tables, of which there may be many. The ISO standard UML model allows for a 'SET<...>', that is, it allows many different alternate expressions. Each of the alternate language tables is identical in structure with the same attribute (column) names. Only the names of the tables are different, and the language code used in the clAlLanguage column that identified the language (or language and dialect) used in the table. These Alternate Language tables are disjoint from the primary table in that they may be maintained separately by different persons, or they may be maintained as supplemental tables in the same database.

Table 4-5 FAO_ClassificationItemClass – Alternate Languages

Name	Description
clAlItemIdentifierReg [itemIdentifier]	(PID) a positive integer (i.e., greater than zero) that serves as a unique permanent identifier for the entry and is used to uniquely denote that item within the register and is intended for information processing. This is the item identifier ID of the entry in the alternate languages table, which is separate from the primary table and which may be different from that in the primary table since the alternate language table is managed separately.
clAlDate [date]	The reference date for the entry is the dateAmended; that is, the date at which the entry was placed in the register. The dateAccepted is the same as the date amended for the first entry of an element in the register. This is the date of the entry in the alternate languages table, which is separate from the primary table and which may be different from that in the primary table since the alternate language table is managed separately.
clAlAlphaCode [alphaCode]	A character string containing a compact and human-readable designator in an alternate language that is used to denote a register concept. The name shall be unique within a register. Multiple items of the same item class may use the same value for name but only one such item may have a status of 'Valid'.

Name	Description
clAlStatus [status]	The registration status of the registerItem from the code list statusCode.
clPItemIdentifierReg [itemIdentifier]	Reference to class in base classes table. This is the external key that links the alternate language table to primary table. This is a link to the primary table registered item that was valid at the time that the alternate language record was created. It does not become obsolete. If the primary table is revised it may get ahead of the alternate language tables(s), but a trace can be done in the primary table from an old item to the current valid item. This means that the alternate language tables do not need to be managed at the same time as the primary table.
clAlLanguage [alternateLanguage]	The three letters code from ISO 639-2 identifying the alternate language.
clAlName [name]	The name of the registered item expressed in the alternate language.
clAlDescription [description]	A character string in the alternate language containing the definition or description of the concept embodied by that item.
clAlSourceReference [sourceReference]	Optional reference to source.

Table 4-6 FAO – Reference

Name	Description
referenceId [identifier]	(RPID) a unique permanent unique identifier for the entry used denote that item within the register of the form 'Ref- #' where # is a positive integer.
refItemIdentifierAtSource [itemIdentifierAtSource]	A character string containing a source identifier.
refSimilarity [similarity]	Similarity to source from the code list: identical, restyled, contextAdded, generalization, specialization, unspecified.
refText [referenceText]	An optional character string describing the reference.
refNotes [notes]	An optional character string containing notes

refSource [sourceReference]	Reference to class in base classes table.

Table 4-7 FAO – Reference source

Name	Description
refId [identifier]	(RSPID) a unique permanent unique identifier for the entry used denote that item within the register of the form "Source- #" where # is a positive integer.
refAuthor [Author]	Optional name of the author.
refAuthorOrganization [authorOrganization]	Name of the author's organization.
refAuthorOrgCode	Code representing of the author's organization.
refPublisherOrganization [publisherOrganization]	Name of the publisher's organization.
refPublisherOrgCode	Code representing of the author's organization.
refPublishedYear [year]	Reference date for the cited reference.
refFileTitle [fileName]	Title of the referenced source file as part of a URL.
refFileFolder [fileFolder]	Folder containing the referenced source file as part of a URL.
refLink [link]	Link to the referenced source file as part of a URL.

5. Conclusion

The LCL registry structure and associated registers described here provide a working solution for FAO in compliance with the developing standards on LC and Classification Systems item registration. This structure also provides an input to the ISO TC211 process for the future development of ISO 19144-4 on Registration and Implementation Aspects, and for the future revision of ISO 19144-1.

This registry also provides, *i*. a global up-to-date database of the use of remote sensing data to monitor status of natural resources using LCML in support to eradicate poverty (SDG 1) and end hunger and all forms of malnutrition (SDG 2) through FAO Hand in Hand initiative *ii.* Provides one platform to access LC legend, LC datasets and LCL classes to the geospatial community including researchers, scientists, international / national / regional / local organizations, and academia. The description of a legend in terms of the elements described in ISO 19144-2 LCML is addressed in Appendix A.

6. References

Feranec, J., Solin, L., Kopecka, M., Otahel, J., Kupkova, L., Stych, P., Bicik, I., Kolar, J., Cerba, O., Soukup, T. & Brodsky, L. 2014. *Analysis and expert assessment of the semantic similarity between land cover classes*. Progress in Physical Geography: Earth and Environment, 38(3): 301–327. https://doi.org/10.1177/0309133314532001

Gregorio, A., & Jansen, J. 2000. *Land Cover Classification System (LCCS): Classification Concepts and User Manual*. A. Di Gregorio & L.J.M. Jansen, eds. Rome, Italy, Food and Agriculture Organization of the United Nations.

ISO. 2009a. *ISO 19144-1:2009 Geographic information — Classification systems — Part 1: Classification system structure.* International Organization for Standardization, Geneva Switzerland.

ISO. 2009b. *ISO 19144-2:2012 Geographic information — Classification systems — Part 2: Land Cover Meta Language (LCML).* International Organization for Standardization, Geneva Switzerland.

ISO. 2015. *ISO 19135-1:2015 Geographic information — Procedures for item registration — Part 1: Fundamentals.* International Organization for Standardization, Geneva Switzerland.

ISO. 2016. *ISO 19110:2016 Geographic information — Methodology for feature cataloguing.* International Organization for Standardization, Geneva Switzerland.

Mosca, N., Gregorio, A.D., Henry, M., Jalal, R. & Blonda, P. 2020. *Object-based similarity assessment using Land Cover Meta Language (LCML): concept, challenges and implementation*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing: 1. https://doi.org/10.1109/JSTARS.2020.3003888

USGS. 2011. *U.S. Geological Survey Gap Analysis Program, 20160513, GAP/LANDFIRE National Terrestrial Ecosystems 2011: U.S. Geological Survey* (also available at https://doi.org/10.5066/F7ZS2TM0, https://www.usgs.gov/core-science-systems/science-analytics-and-synthesis/gap/science/land-cover-data-overview?qt-science_center_objects=0#qt-science_center_objects)

7. Appendix A Representation of legend elements

The description of a legend in terms of the elements described in ISO 19144-2 LCML consists of expressing each class in terms of the component elements.

LC and LU can be represented by the combination of standardized atomic elements such as Shrubs, Herb, Trees, Building, organized into strata (or layers). Each of these elements can be further characterized by extra attributes properties and characteristics. The relationships of these elements are defined in the ISO metalanguage standard(s) for LC ISO 19144-2 (and LU ISO 19144-3). A traditional LCL can be constructed out of the basic LCML elements with their properties and characteristics.

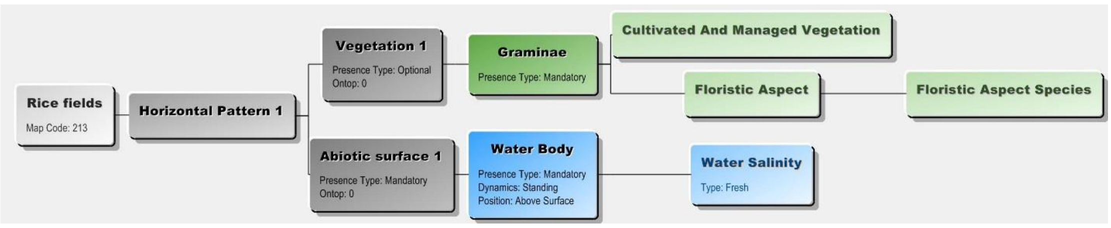
The following are three examples of LCL classes taken from the CORINE Land Cover (CLC) inventory of LC classes. This illustrates the mechanism of constructing these classes out of the basic modelling elements from the meta language.

7.1. CORINE class 213 - rice fields

Description: Land prepared for rice cultivation. Flat surfaces with irrigation channels. Surfaces periodically flooded.

Structure: This class consists of two separate layers of biotic and a-biotic elements. The first layer consists of the LCML element **herbaceous growth forms,** cultivated with specific floristic name rice. The second layer is composed by the LCML element **natural water**, fresh with a persistence period extend to the whole cultivation time.

Figure 7-1 Class 213 - rice fields

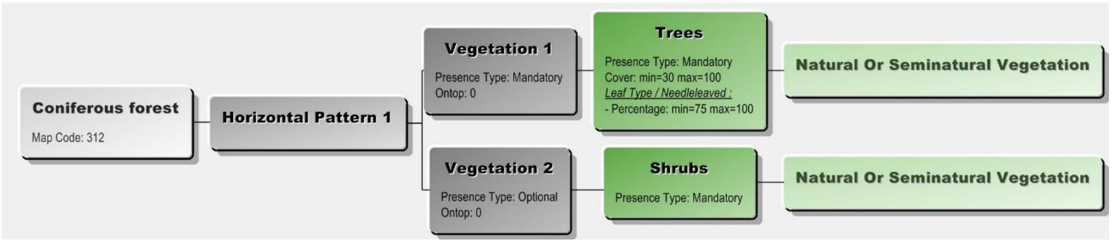


7.2. CORINE class 312 - coniferous forests

Description: Vegetation formation composed principally of trees, including shrub and bush under storeys, where coniferous species predominate.

Structure: Two layers have been considered. The first one composed by LCML element trees, needle leaved with a cover from 30 to 100% (mandatory), the second one (optional) composed by LCML element shrubs.

Figure 7-2 Class 312 - coniferous forests

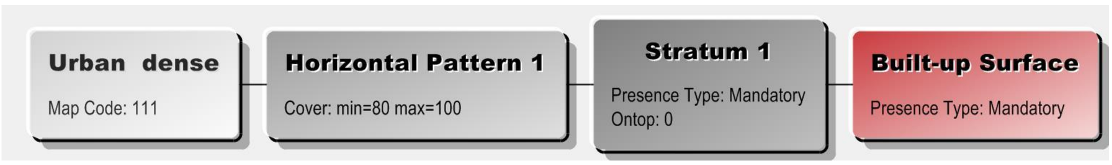


7.3. CORINE class 111- continuous urban fabric

Description: Most of the land is covered by structures and the transport network. Building, roads, and artificially surfaced areas cover more than 80 % of the total surface. Non-linear areas of vegetation and bare soil are exceptional.

Structure: Two horizontal patterns of nonlinear and linear objects has been created.

Figure 7-3 Class 111- continuous urban fabric



In LCML syntax a combination of different horizontal patterns is used when is necessary to describe natural functional units composed by different LC features. Additional descriptions of the remaining CORINE classes, as well as other legends classes are described in separate documents and referenced in the registry.

8. Appendix B
Comparison metrics

Assessing LC similarity using standard legends has always been difficult, for several reasons related to the symbolic and ambiguous nature of the symbols used for describing them. Indeed, inconsistency is a well-known problem that persists even today, decades after the introduction of GIS tools. Technological and scientific progress with new tools, higher resolution images and novel approaches, cannot provide the expected and potential leaps in accuracy and consistency in classification if the systems responsible for representing and managing the information does not evolve in the same way.

Some efforts have been made to study consistency, like in the work by Feranec et al. (2014) on analyzing and assessing the semantic similarity between LC classes. The approach, completely based on expert evaluation, without any aid by computer tools, compared nomenclatures between CLC and National Land Cover Dataset (NLCD). Deep knowledge of the two taxonomies was necessary, and the diverse expert background affected the way LC classes were compared, and eventually the results quality. In particular, the lack of a formal semantic class description proved a limiting factor, since CLC taxonomy provides only text class description. Moreover, CLC description encompasses both LC and LU concepts. The process was time-consuming for the scientists, geographers and environmentalists involved, with their heterogeneous background contributing in different interpretations of the text descriptions. After almost a week, unanimous agreement on one-to-one class matching was possible just in three cases. Eventually, experts agreed that ambiguity and subjectivity were responsible for the limited results achieved.

The approach summarized here is different because it works by exploiting the semantic richness of LCML, and harnessing, wherever possible, automatic tools to speed-up the process.

LCML standardized structure and modularity are key factors in defining an automatic similarity assessment able to understand how much two (and by extension, more) classes are similar to each other. The assessment becomes feasible as soon as the composition of LC classes in basic elements, properties, and characteristics is exploited to decompose a complex comparison in simpler parts.

A LC description in LCML is hierarchical in nature, with the selection of one or more basic LCML elements representative of the class being described. An irrigated agriculture feature, for example, starts with identifying the LCML element herbaceous vegetation and then associating cultivated and irrigated as important characteristics attached. Basic LCML elements are the bone structure around which a class is created; computing a similarity score for the whole class is first and foremost a measure on basic elements similarity. Thus, at its most basic way, e.g. by considering basic elements only, is performed by considering the following steps:

1. Correspondence
2. Extensiveness
3. Integration of correspondence and extensiveness scores

Composition rules need to consider special cases, such as optional or exclusive nodes. Moreover, properties and characteristics can be used for enhancing even more the result. However, for the sake of simplicity, this appendix is concerned about basic elements only. Additional information is available in the study conducted by Mosca *et al.* (2020).

8.1. Correspondence

Real-world LC relationships between different basic LCML elements have been used for shaping the LCML UML meta-model, including inheritance between its nodes. This same inheritance might be considered a way to compute likeliness. However, being UML a representation of a complex reality, using lookup tables enable to start from the evaluation based on path distance in the hierarchy, and then tweak individual values for better capturing expert judgment.

The approach described here for the similarity assessment exploits this by giving a different likeliness score to each different pair of basic elements, that is, as an example, it gives a higher value when comparing a tree with a shrub than when comparing a tree with a rock. The likeliness can assume a value ranging from 1 to 10.

The comparison between the input (query) LC class and a reference LC class happens by associating, for each inputs basic LCML element, the most likelihood of the reference basic LCML element, by using the correspondence value (from 1 to 10) of the related table. Hence, the input basic LCML element is the element(s) forming a query (or interrogation) class for which a similarity value is calculated in relation to one or more reference classes composed by one or more element (s) called reference basic LCML element. When two or more input basic LCML elements find two or more similar values of different reference LCML basic elements, the hierarchy established by LCML UML diagram can be used to disambiguate and automatically select the most convenient reference LCML basic element.

Table 8-1 Sample table comparing vegetated elements.

When comparing elements of the same type of the maximum score is given.

Land cove class	Reference land cover class										
	LC_GrowthForms (B)	LC_WoodyGrowthForms (C)	LC_Trees (D)	LC_Shrubs (E)	LC_HerbaceousGrowthForms (F)	LC_Graminae (G)	LC_Forbs (H)	LC_LichenandMosses (I)	LC_Lichen (J)	LC_Mosses (K)	LC_Algae(L)
LC_GrowthForms (B)	10	9	9	9	9	9	9	9	9	9	9
LC_WoodyGrowthForms (C)	9	10	9	9	3	3	3	1	1	1	2
LC_Trees (D)	9	9	10	6	3	3	3	1	1	1	2
LC_Shrubs (E)	9	9	6	10	4	4	4	1	1	1	2
LC_HerbaceousGrowthForms (F)	9	3	3	4	10	9	9	4	4	4	4
LC_Graminae (G)	9	3	3	4	9	10	9	4	4	4	4
LC_Forbs (H)	9	3	3	4	9	9	10	4	4	4	4
LC_LichenandMosses (I)	9	1	1	1	4	4	4	10	9	9	4
LC_Lichen (J)	9	1	1	1	4	4	4	9	10	9	4
LC_Mosses (K)	9	1	1	1	4	4	4	9	9	10	4
LC_Algae (L)	9	2	2	2	4	4	4	4	4	4	10

8.2. Extensiveness

The computation of extensiveness compares the number of input basic LCML elements against the number of the reference basic LCML elements. The value is calculated according to the LCML extensiveness value table.

Table 8-2 Sample of basic LCML elements extensiveness value

Input corresponds to the class for which the calculation of similarity is done, reference is the reference class on which the comparison of the similarity is computed, C is the extensiveness values from 1 to 10.

Reference count	**Input count**							
	1	2	3	4	5	6	7	8
1	10	7	6	5	4	3	2	1
2	6	10	9	7	6	4	2	1
3	5	6	10	9	8	6	5	3
4	2	5	6	10	9	7	6	4
5	1	3	5	7	10	9	7	6
6	1	2	5	6	7	10	9	7
7	1	2	3	4	6	7	10	9
8	1	2	3	4	5	6	7	10

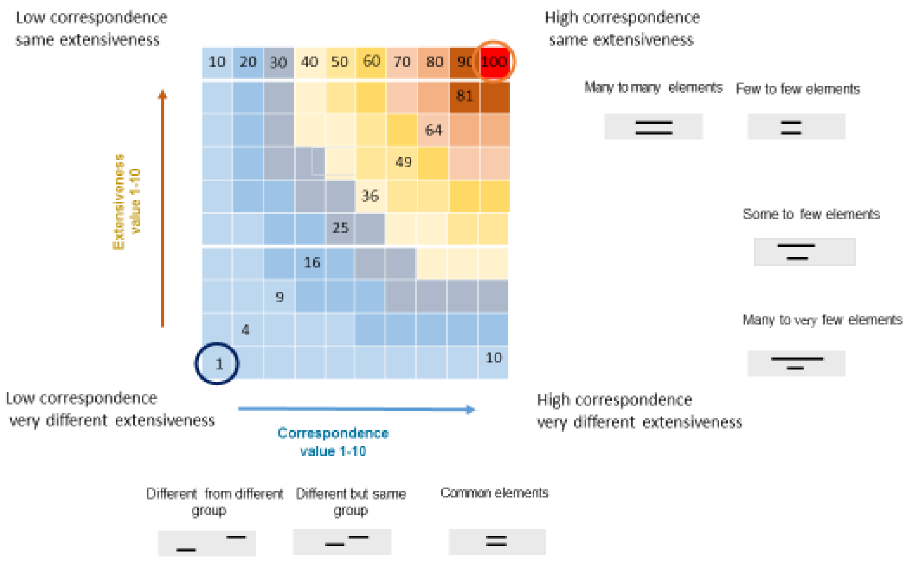
8.3. Integration of correspondence and extensiveness scores

The final computation table is prepared by multiplying the values of correspondence and extensiveness with values ranging 1- 100.

The lower similarity value represents cases with both low correspondence and extensiveness values, while the upper right represents cases with both high correspondence and extensiveness values. Extensiveness score is high when the number

of objects in the query class is comparable with the number of objects in the reference class. Correspondence is high when the type of objects to be compared between the query class and the reference class is the same and the value (or ranges) are similar.

Figure 8-1 Correspondence and extensiveness metrics



The bivariate schema representing the possible combinations of correspondence and extensiveness metrics. Abbreviations used in Table 8-3, are as follows.

- FP: Forest Plantations
- ShT: Shrubs with scattered trees
- RS: Rural settlement
- PCs: Single Crop
- SP: Salt Pan
- FMp: Mangrove Plantation
- Bwa: Brackish Water Acquaculture
- R: Rivers and Khals
- PCm: Multiple Crop
- FH: Hill Forest
- lcc1: Artificial surfaces

Table 8-3 Example of output from the similarity assessment (in %)

	lcc1	lcc2	lcc3	lcc4	lcc5	lcc6	lcc7	lcc8	lcc9	lcc10	lcc11	lcc12	lcc13	lcc14
FP	10	75	90	57	95	100	30	80	33	90	50	10	10	10
ShT	10	40	90	39	65	100	30	100	15	90	50	10	10	10
RS	4	38	41	75	43	53	77	46	76	50	37	32	40	40
PCs	10	100	30	60	100	30	15	40	60	90	10	10	10	10
SP	10	10	10	6	10	10	5	10	6	10	90	10	10	10
FMp	6	18	53	50	18	70	95	56	65	67	39	42	53	53
Bwa	8	75	23	25	75	49	39	53	60	71	41	80	100	100
R	10	10	10	6	10	10	50	10	45	10	10	80	100	100
PCm	10	100	30	39	100	65	12	70	33	90	50	10	10	10
FH	10	30	90	41	30	100	45	60	18	90	10	10	10	10

- lcc2: Herbaceous crops
- lcc3: Woody crops
- lcc4: Multiple or layered crops
- lcc5: Grassland
- lcc6: Trees covered areas
- lcc7: Mangroves
- lcc8: Shrub covered areas
- lcc9: Herbaceous vegetation acquatic or regularly flooded
- lcc10: Sparsely natural vegetated areas
- lcc11: Terrestrial barren land
- lcc12: Permanent snow and glaciers
- lcc13: Inland water bodies
- lcc14: Costal waterbodies and intertidal areas

9. Appendix C
Translation guideline

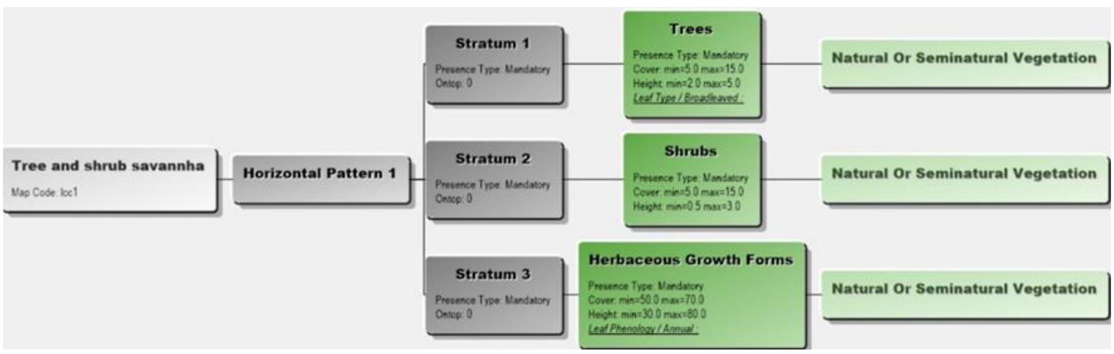
A concise guideline to support translations of conventional land cover legends (LCL) into the object oriented LCML (ISO standard 19144-2) using FAO LCCSv.3 toolbox.

9.1. Importance of Land Cover Legend (LCL) translation

A LC mapping activity can be defined as a process of extraction of different types of information (by nature rather subjective) from specific (objective) data (satellite images). In various LCL, class definitions are imprecise, ambiguous or absent and their meaning is often limited to class name or unclear description and creation of categories without allowing the end user to fully understand the different components of it, implies rigidity in the transfer and utilization of information.

In 2003 FAO submitted the LCCS concept to ISO Technical Committee 211 on Geographic Information as a contribution toward establishing an international standard for LC classification systems. Its intrinsic structure of open object oriented system allows not only an unambiguous description of real world features more consistent with the logic and structure of modern data bases but also enlarge the capability of the system to describe phenomena related to inputs and activities peoples undertake on a certain LC feature typical of agriculture and other LU (Figure 9-1). Therefore, conversion/translation of any LCL using LCML (ISO 19144-2)/LCCS can assure integration and semantic interoperability of different data sets derived from local knowledge and/or extensive field analysis such as different local, national, regional or global legends as well as field observations done without referring to any particular predefined list of classes.

Figure 9-1 Description of a LCL class using LCCS



9.2. Advantages of ISO standard – Land Cover Meta Language (LCML)

National nomenclatures are generally tailored to their specific circumstance and incorporate local knowledge and terminology. However, at the same time it is clear that:

- A legend is a dynamic feature, cannot remain untouched through time
- Stability and consistency of national systems are fundamental requirements but must not be confused with the static conservation of a certain number of classes and their fixed description
- Availability of different sensors and image processing technologies can modify the quality and the amount of information available and this must be reflected in the national legend
- A diversified and enlarged community of national and international users, new applications/utilization of LC and LU data can generate the demand of new types of information
- The advance of science in database management and interrogation pose new challenges and create new opportunities for structuring the information in a legend.

Therefore, it is important for a national system to maintain a dynamic aspect coping with the technological advances in many aspects of land resources and spatial information. The upgrading of a conventional system with the most advanced standards in LC and LU representation is thus a functional and useful approach to deal with upcoming technical advances. To be useful and interoperable this process however must adhere to some basic requirements:

- It must be fully accepted and endorsed by local experts to avoid technical gaps.
- The process must undertaken with national experts and national institutions.
- It must not create a large discontinuity with previous systems.

The upgrading of a conventional legend into an object oriented system will have the following main advantages:

- The national/regional legend will be described according to the international ISO standard 19144-2, the only accepted international standard for LC.
- From an unsystematic text description, the original legend will be converted into a coherent system where a shared dictionary of terms will allow a clear understanding of each class information content. The end user community will have an easier and reliable understanding of legend content. This will support a more efficient and democratic dialogue between map producers and data users.

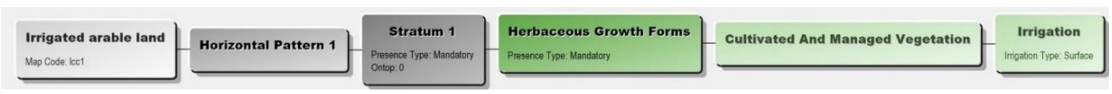
- The conversion will allow the original legend to be represented with the last advances in the modelling languages (UML, Markup language as xsd, xml etc.), this will make all set of information machine readable.
- The conversion will make possible and initiate the harmonization process between different national, regional and global LCL. The exchange and utilization of LC data will be more efficient and reliable. It will foster the process within the LC community to create a common understanding of LC nomenclatures with the aim to produce global, regional, national and sub - national datasets able to be reconciled at different scales, levels of detail and geographic location.
- The passage from an unsystematic legend development to a comprehensive consistent system will permit the evaluation of the limitations and/or inconsistencies of the actual list of classes (or their description) in the original legend.

9.3. Basic guidelines for legend translation process

The conversion/translation process is never a straightforward simple procedure, some main suggestions are listed to make this process easier and correct:

- Full comprehensive knowledge of the syntax and shared dictionary of the LCML.
- LCML is an object oriented language based on a physiognomic approach. It is mainly composed by basic elements that can be further enriched in their semantic meaning with a series of attributes. The basic elements specify the biophysical entities that populate the Earth surface (like trees, herbs, rocks, waterbodies, buildings etc.).
- Full understanding of the text LCL class description of the original legend because without a clear understanding of semantic meaning of the original LCL class, translation is not possible.
- Any LCL class translated into the LCML **must** always start from the definition of one or more basic element. If for instance in a traditional LC legend exist a class called **irrigated arable land** the translation process, first must identify the main basic element that characterize the biophysical aspect of this feature, afterwards specific attributes will further characterize the whole semantic meaning of the class. For example, in Figure 9-2, the basic element is **herbaceous vegetation** to which attributes on vegetation management and irrigation types are attached.

Figure 9-2 Schema of the translated class



9.4. Steps for a standardized translation and outputs presentation

For each translated legend a series of standardized outputs must be generated. These include:

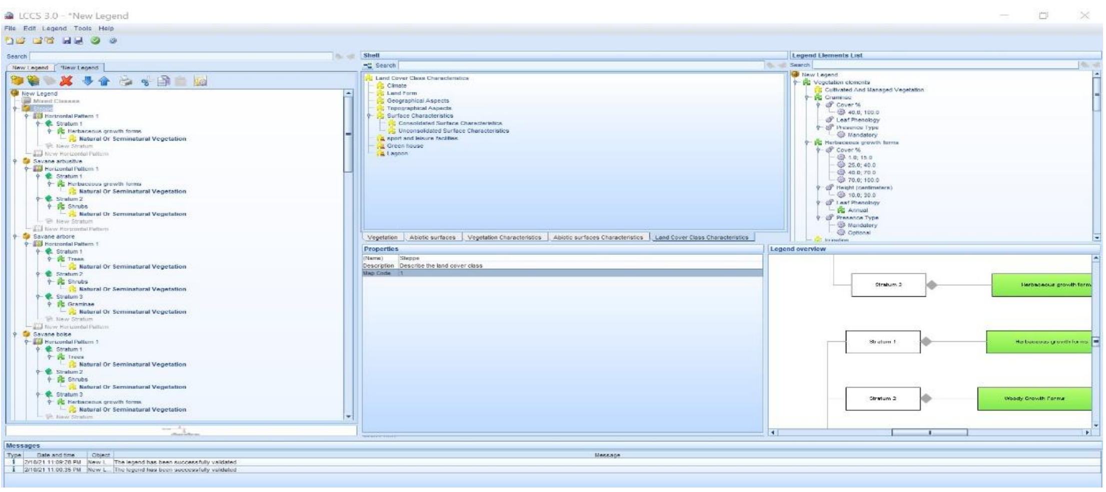
- LCCSv3 file of the translated legend,
- UML like graph for each translated class,
- UML schema (using the Enterprise Architect software) for each translated class.

9.4.1. Preparation of legend file using LCCS3 tool

The process of translation may be undertaken using FAO LCCS v.3 tool and is accomplished in two steps:

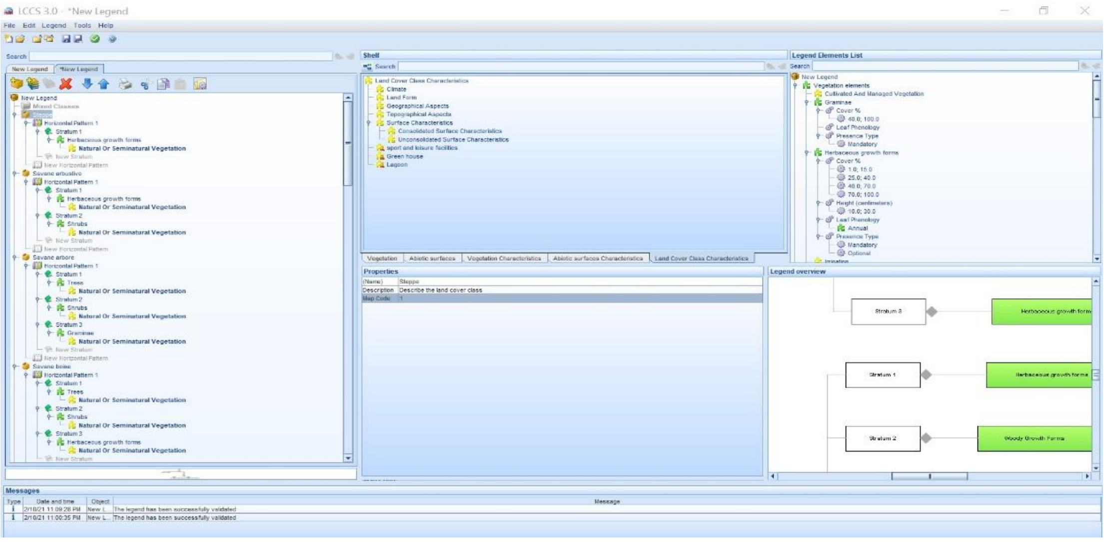
Step 1: The translation is done using the LCCS v.3 tool that allow the application of the ISO 19144-2 standard in an easy and user-friendly way (see Figure 9-3).

Figure 9-3 Translation of a legend into the LCCS v.3 tool



Step 2: After completing the translation and before saving the final file, all the processes must be validated to understand if the conversion process has been done according to the rules of the ISO standard 19144-2 LCML. To do so, the validation process must be activated using the green button in the left upper part of the software (see Figure 9-4), if no mistakes have been made a message will appear in the lower part of the software screen. At this point the ‘LCCS v’3 file can be saved.

Figure 9-4 Validation process (LCCSv3 tool)

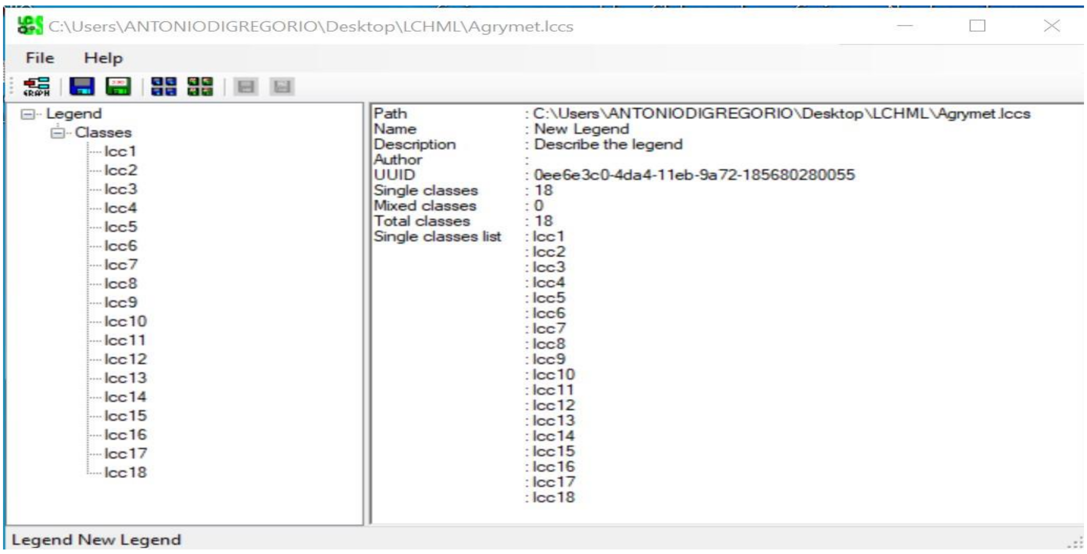


9.4.2. Generation of a Unified Modeling Language (UML)

UML is a graphic representation of the class structure of each translated legend category. The most complete graph can be produced using FAO LCCS tool software in two steps.

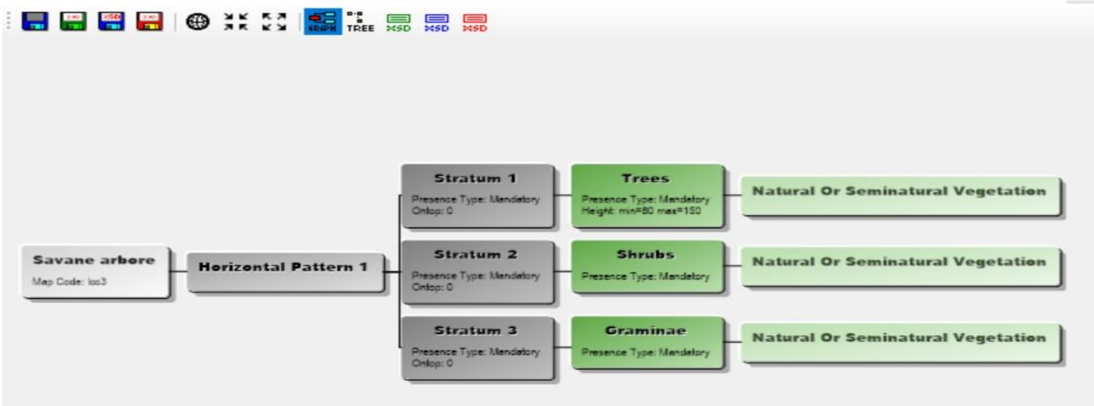
Step 1: After saving .lccs file from FAO LCCS v.3 software, the file must be loaded into FAO LCCS tool software. When this is successfully done the software will show the list of the translated classes (see Figure 9-5**Error! Reference source not found.**).

Figure 9-5 List of translated classes



Step 2: For each translated graph an UML like diagram will be generated (see Figure 9-6), all LCL classes can be saved in .jpeg or .tiff file to create the second output of the translation process.

Figure 9-6 Example of UML like schema of a translated LCL class

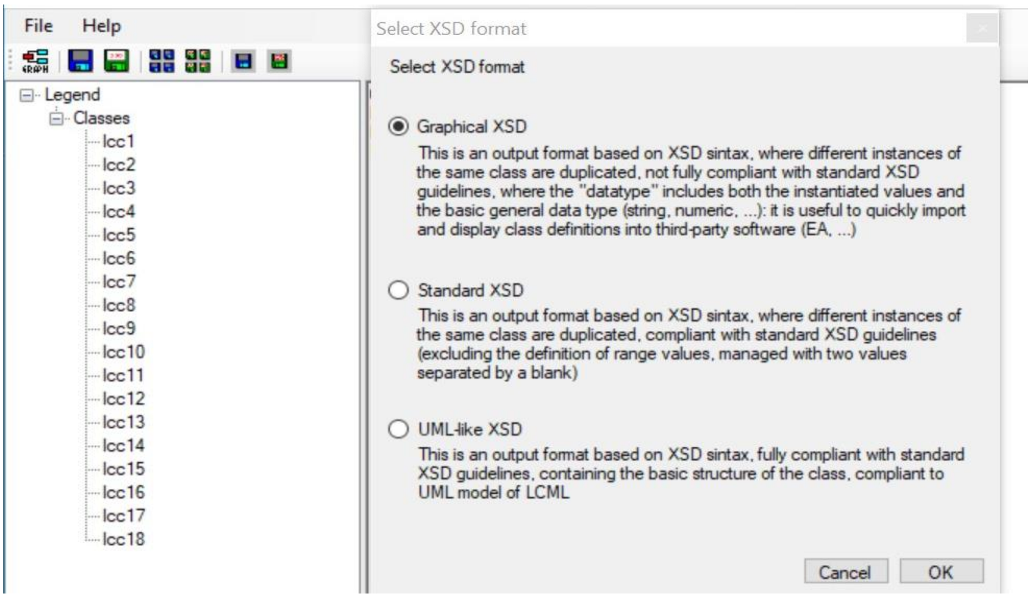


9.4.3. Preparation of legend file using Enterprise Architect software

This can be done in two steps which are as follows:

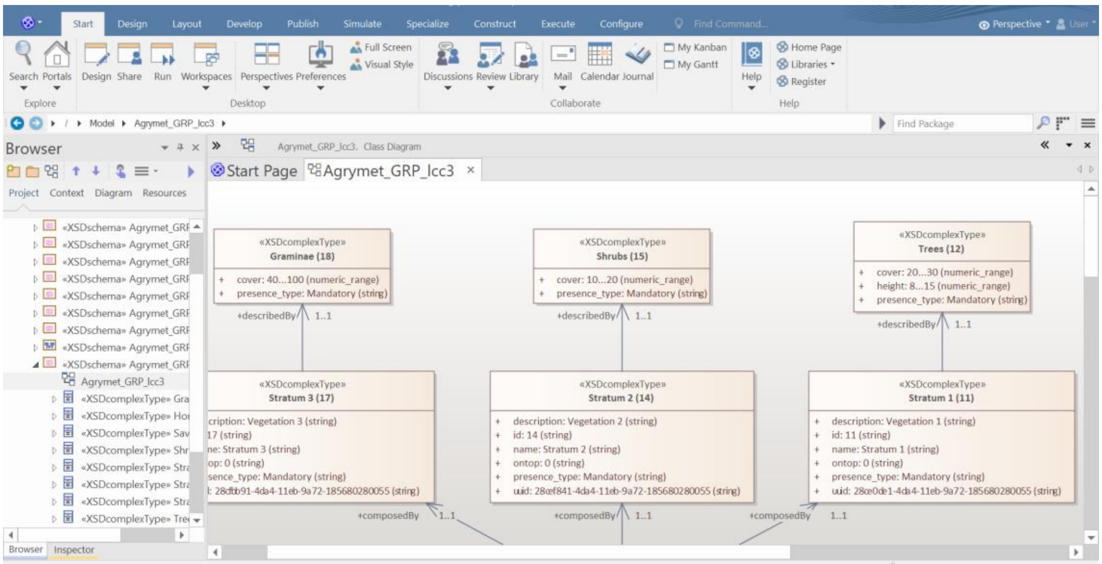
Step 1: When the .lccs v3 file has been successfully imported into the software each legend class file can be converted/saved into the .xsd format (the best format to import into the enterprise architect software is graphical xsd) (see Figure 9-7).

Figure 9-7 .xsd file



Step 2: Import and generate translated LC legend containing an UML schema for each translated class in enterprise architect (see Figure 9-8).

Figure 9-8 .xsd file into Enterprise Architect software

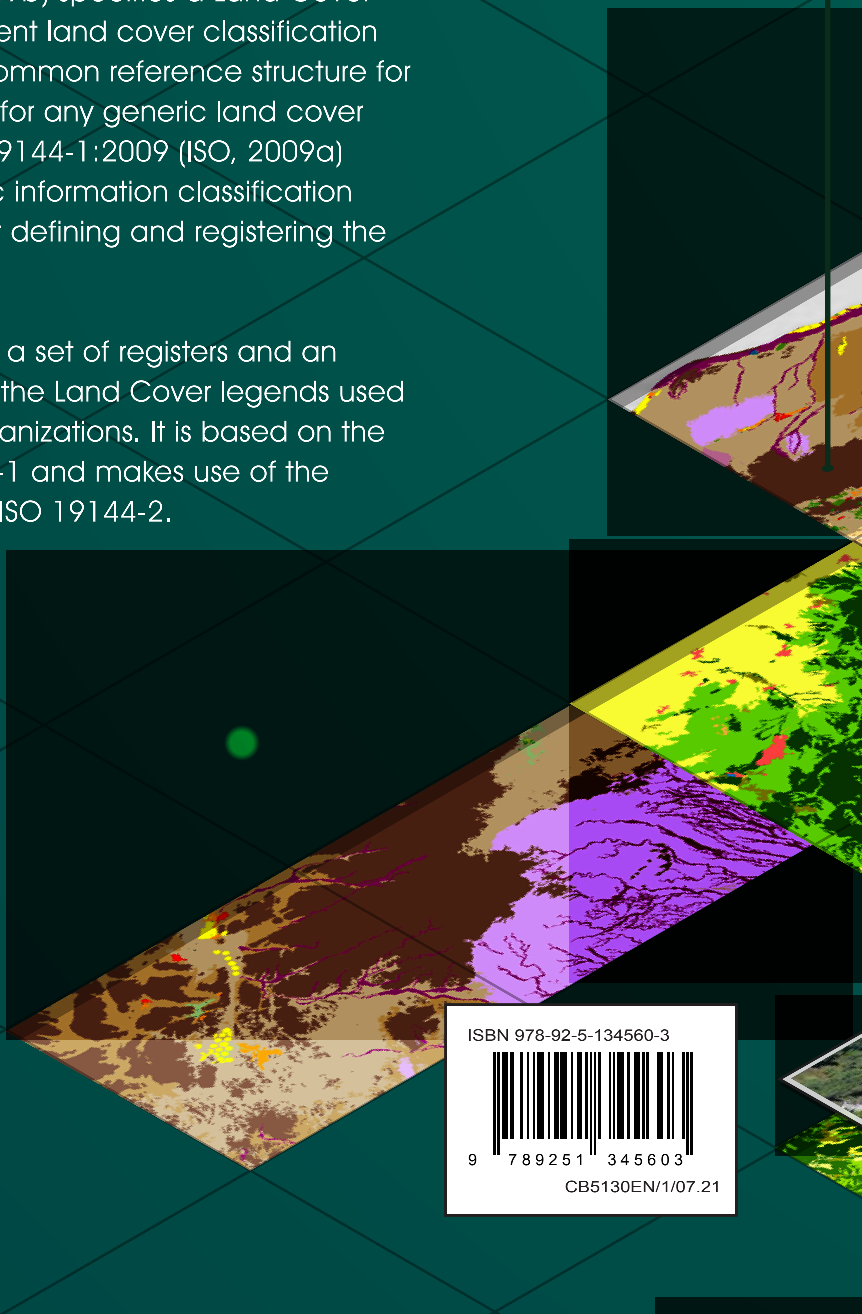


Register implementation for land cover legends

Creating global land cover data is very difficult because data is gathered for different purposes through different survey and acquisition projects and various nations use different classification schemes to represent that data. There are many different legends used to classify data. However, the integration of data is required to address regional and world wide requirements.

The UN FAO together with ISO Technical Committee 211 on Geographic Information / Geomatics developed a series of standards for the management of Land Cover and related classification systems. The standard ISO 19144-2:2012 (ISO, 2009b) specifies a Land Cover Meta Language (LCML) that allows different land cover classification systems to be described. It provides a common reference structure for the comparison and integration of data for any generic land cover classification system. The standard ISO 19144-1:2009 (ISO, 2009a) establishes the structure of a geographic information classification system, together with the mechanism for defining and registering the classifiers for such a system.

This document describes the structure of a set of registers and an associated registry for the description of the Land Cover legends used by the UN FAO and associated other organizations. It is based on the registry concepts identified in ISO 19144-1 and makes use of the descriptive metalanguage described in ISO 19144-2.



ISBN 978-92-5-134560-3
9 789251 345603
CB5130EN/1/07.21