OntoLULC-SOTA: An ontology based approach to make systematic reviews for LULC data

Martin Cubaud[1], Ana-Maria Olteanu-Raimond[1], Cidalia C. Fonte[2, 3], Diogo Duarte[2, 4], Jacinto Estima[5],
Linda See[6], Nicolas Gonthier[1], Laurence Jolivet[1], Clément Mallet[1], Arnaud Le Bris[1], Vyron Antoniou[7]

[1] Univ Gustave Eiffel, Géodata Paris, IGN, LASTIG, F-77454 Marne-la-Vallée, France - (firstname.lastname)@ign.fr
[2] Institute for Systems Engineering and Computers at Coimbra (INESC Coimbra), 3030-290 Coimbra, Portugal
[3] University of Coimbra, Department of Mathematics, Apartado 3008, EC Santa Cruz, 3001-501 Coimbra, Portugal - cfonte@mat.uc.pt
[4] University of Coimbra, Department of Electrical and Computer Engineering, Polo 2, 3030-290 Coimbra - diogovad@inescc.pt
[5] University of Coimbra, CISUC, Department of Informatics Engineering,
Rua Sílvio Lima, 3030-290 Coimbra, Portugal - jacinto.estima@gmail.com
[6] International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria - see@iasa.ac.at
[7] Hellenic Army Geographical Directorate, 15561 Cholargos, Greece - byr.anton@gmail.com

Keywords: Land Use Land Cover, Ontology, Systematic review, LULC Data cycle of life.

Abstract

Land Use (LU) and Land Cover (LC) data allow us to understand the physical and human activities associated with a given land. Thus, LULC is a dynamic and highly researched field. LULC review papers are numerous and provide high-level insights about the proposed approaches, the data used, the study cases, the strengths and limitations, and the identification of new research gaps. Nevertheless, these reviews are not systematic and reproducible. The goal of this work is to propose an ontology to help the research community conduct systematic and shareable literature reviews and comparable analytical analyses of scientific papers. To achieve this, we formalize their metadata, content, strengths, and weaknesses. In particular, we consider the scientific paper as the central element of our ontology and we define formal semantics for all relevant items (data process, LULC life cycle and scientific paper). We hope to open the path to more efficient synthesis, discovery, and reuse of research outcomes from the literature. To facilitate the instantiation process and make it accessible to a broader range of researchers, we designed a tabular-based template. We used our template to simulate the process of conducting a literature review on three use cases: building function, global land cover mapping, and multi-class change detection.

1. Introduction

Land Use (LU) and Land Cover (LC) data allow us to understand the coverage (i.e., the physical and biological material present on the Earth’s surface, such as vegetation, water, bare soil, or built-up areas) and the use (i.e., the human activities or functions associated with a given land type, such as agriculture, residential, industrial, or recreational use) of the territory. LU and LC are two dual components, often confused with one another or used interchangeably. They provide useful information to support a wide range of applications, from urban planning, land artificialization control, agricultural management to biodiversity conservation.

Advances in Earth Observation technologies and computational methods have stimulated much research on LULC mapping, classification, and change detection (CD). Systematic reviews have been extensively adopted in the literature for decades and remain very important for research academics: they aim to provide an analytical presentation of all relevant studies in the field of LULC, often coupled with high-level insights about the proposed approaches, the key methodological elements, the study cases, the strengths and limitations of main methods, as well as the identification of new research gaps. The way of making reviews is also described. For example, (Kitchenham and Charters, 2007) recommend three key steps when conducting scientific reviews: planning (i.e., identify the research questions, data sources, paper search strategy), conducting (i.e., select the relevant papers and synthesize them), and reporting (i.e., write and publish the review).

The reviews may sometimes be too broad, limiting the number of papers addressed or focusing on one specific area or subject. For instance, (Wang et al., 2023) studies one hundred LULC products based on satellite to conclude that LULC products have been gradually advancing towards finer classification and higher spatial and temporal resolution. Moreover, (Mallet and Le Bris, 2020) highlight the challenges related to the need for increased resolution in an operational setup. (Wu et al., 2024) chose to focus on crowdsourced geographic information and show that the areas studied are quite limited and that the approaches are also limited. Some reviews focus on one specific area, such as Brazil (Santos et al., 2025) or China-Central Asia-West Asia (Naboureh et al., 2021) or one type of LULC, such as landslide (Pacheco Quevedo et al., 2023) or land cover (Xu et al., 2025). In addition, the description of the main methods and research gaps are often already well documented separately and biased towards authors’ own perspectives.

In the last decade, change detection has shifted from image processing (Chughtai et al., 2021, Singh, 1989) to artificial intelligence and especially deep learning approaches (Peng et al., 2025). Integrated artificial intelligence technology has become a research focus in developing new change detection methods on various sensors (e.g., optical remote sensing data, synthetic aperture radar (SAR) data, street view images) (Shi et al., 2020). Most of the reviews focus on the new methods or architecture developed for this task (Peng et al., 2025), or focus on a subproblem such as heterogeneous change detection (Lv et al., 2022) whereas fewer papers provide some insights on the current challenges in this field (Shi et al., 2020, Gonthier, 2026). The challenges are heterogeneous big data processing and the reliability of AI, according to (Shi et al., 2020) and the

https://doi.org/10.5194/isprs-annals-XI-1-2026-473-2026 | © Author(s) 2026. CC BY 4.0 License.

need to move toward operational setup and to add the human into the loop (Gonthier, 2026). Such reviews become more and more difficult to carry out and time-consuming, leading to Artificial Intelligence-based solutions that do not bring further insights and exacerbate current shortcomings (Castelvecchi, 2025). Such static reviews are also generally only shared with the community through published research papers. They cannot be easily updated or queried to emphasize on some particular aspects or highlight different points of view regarding the goal of the review papers. In remote sensing, for example, now the most used approach is more oriented towards a ”meta-analysis”, i.e., performing a pure statistical-driven analysis of the plethora of papers, such as: ”77% of the papers in 2021 include the keyword ’deep learning’, while such value was 47% in 2019, showing that there is a clear increase in the use of such methods”. This avoids a more analytical and detailed analysis (in particular pointing out main research gaps), but also does not provide a complete and detailed analysis of how main methods are used. Again, few papers share their data (Ma et al., 2017) to move towards a more collaborative and incremental work on a continuously evolving domain.

Finally, there remains a significant gap in how easily data and tools can be found and compared across research communities. Thus, there is a major need for better ways to provide contextual information, curation, tools, critical analysis, and user feedback to enable meaningful reuse across communities.

Going beyond merely releasing datasets and codes openly, some communities have invested in creating ontologies or knowledge graphs. They not only formalize a shared vocabulary and establish links between resources but also enable the inference of new knowledge. For example, there are ontology in domains such as tourism (Zhang et al., 2023), landmarks (Olteanu-Raimond et al., 2023), data quality (Yılmaz et al., 2024), data sets (Färber and Lamprecht, 2021), scientific paper submission (Bucur et al., 2023), LULC nomenclature (Zhou et al., 2021), to cite just a few. The latter represents a first step toward formalizing scientific papers using the concept and technology of nanopublication (`http://www.nanopub.org/nschema#`), and proposes to formalize in a RDF-based format all stages of the publication process: submission, reviews, responses, and editorial decisions. Although this is relevant to our work, it remains general and remains at the level of the paper as a whole, focusing mainly on the article’s primary scientific claims (Bucur et al., 2023).

In addition, several automatic tools have been developed to support literature exploration and bibliometric analysis, such as VOSviewer (van Eck and Waltman, 2010), Bibliometrix (Aria and Cuccurullo, 2017) or Connected Paper. These methods exploit links between papers, for instance co-authorship, co-citation, bibliographic coupling, keyword co-occurrence, and direct citation networks. These links help identify clusters of related work, map research trends, and trace how ideas evolve over time. However, such analyses remain mostly structural and do not allow for detailed comparison of methods or the identification of specific weaknesses or gaps in the literature.

The goal of this research is to propose an ontology that enables the LULC research community to produce systematic reviews which is shareable, open and extensible. Our main research assumption is that achieving this requires comprehensive and exhaustive formalization of all elements related to LULC data and scientific papers through a well-structured ontology.

We built our approach on the work of (Kitchenham and Charters, 2007), and we propose adding three additional steps for the systematic reviews: the first is to formalize the life cycle of LULC data by adopting the concept of *process* already introduced in the literature (Follin et al., 2019). The second is to formalize the metadata of the papers and the existing criteria defined in the literature that are necessary for conducting a systematic review (Naboureh et al., 2021). Finally, the third step is to instantiate the ontology with a selection of papers aligned with the goal of the systematic review and to allow user feedback at different scales (e.g., paper, algorithms, parameters, study area).

Thus, the main contribution of this paper is to define an open source LULC data ontology for systematic review, named **OntoLULC-SOTA** which represents a first step toward building a multi-community platform to support systematic reviews. Its scope includes any research work related to LU or LC data. Beyond state of the art, the ontology may also serve as a decision-making tool.

2. Theoretic Ontology Background

In this section, we briefly introduce the ontology theoretical background necessary for our approach. With respect to the work of Hogan, an ontology is defined as ”a formal representation of knowledge that forms a shared conceptualization of a given domain” (Hogan, 2020). The ontology defines the conceptual model of a domain and semantically describes what exists (i.e., ”the things”) and how these things are related one to each other. In technical implementations, this model is standardly formalized using the Web Ontology Language (OWL) and Resource Description Framework (RDF), which are two semantic web formats for knowledge modelling using a graph-based structure for data description. In an ontology, there are four main items.

The first one is the concept (named class in OWL) which describes semantically an entity of the domain (e.g., LU nomenclature). A class is defined by a minimum of concepts which are: a machine-readable URI (e.g., the Uniform Resource Identifier), multilingual labels which are human-readable, a comment which defines the semantic explanation of what the class represents, and relationships subClassOf (i.e., a class A is the daughter of the class B). The second item represents the instances (named individual in OWL) i.e., the concrete realizations of concepts (e.g., the Inspire LU nomenclature of LU classes). The third concerns the properties, which describe the attributes (named datatype properties in OWL) or the relationships between concepts (named object properties in OWL). Finally, axioms and constraints represent the formal rules that allow asserting facts and restrictions about classes (e.g., an author of a paper is a Human), properties (e.g., an author has an Affiliation), or individuals (e.g., PaperLULC, Douglas Alex meaning that Douglas Alex is the author of PaperLULC), cardinality (e.g., a paper has n Authors), domain (e.g., the property hasAuthor has domain Paper and range Human, meaning that only humans can be authors of papers).

It is important to make the difference between a domain ontology and an application ontology. The last is the deployment of the conceptual model defined in the domain ontology.

How to build an ontology is another important aspect. Research articles provide an overview of methodologies for defining an ontology (Sure et al., 2009), (ElHassouni and El Qadi, 2022). The latter concludes that no methodology has reached a consensus and suggests using ontology design patterns. This methodology is not suitable for our purpose because the LULC data lifecycle is too specific and does not constitute a true workflow. In practice, very few papers address the complete lifecycle of the data. The Agile Methodology for Ontology Development (AMOD) (Abdelghany et al., 2019) is often criticized for being time-consuming; however, its simplified version (SAMOD) (Peroni, 2016), which is an iterative methodology bringing together designers and end users and is easy to adapt (Olteanu-Raimond et al., 2023). To go forward, another relevant aspect for our work is how to publish a scientific paper proposing an ontology. Precise recommendations are made by (Ferrario and Grüninger, 2020) who propose the following key recommendations: (1) clearly define the motivation and use case of the ontology; (2) provide a detailed conceptualization of its structure; (3) describe the implementation in a formal language (e.g., OWL); (4) perform evaluation and validation of the ontology; and (5) ensure its publication and accessibility through a public repository.

In this work, we follow the SAMOD method (Peroni, 2016) to build the ontology and the (Ferrario and Grüninger, 2020) recommendations to publish the ontology paper. Thus, our method is a collaborative work where a community brainstorm to produce the ontology. The authors of the paper are divided into groups: those working on ontology conceptualization, those working on ontology instantiation (with papers), and those testing. This is an iterative process.

3. Designing LULC ontology

3.1 Motivation

Our motivation is to propose a LULC ontology that supports the LULC research community in conducting systematic and shareable literature reviews.

To achieve this, we consider the scientific paper as the central element of our ontology and we formalize its metadata, its content, and its strengths and weaknesses. While modeling metadata is relatively straightforward thanks to existing standardized elements, formalizing the content of a paper and capturing its strengths and weaknesses presents challenges. Moreover, since the role of the LULC ontology is to support systematic reviews, we also consider the indicators needed to review and analyze existing papers.

Thus, first, we consider the full LULC life cycle, which contains five main components: modeling, mapping, validation, management, and updating. Second, our approach is based on the process-oriented concept proposed by (Follin et al., 2019), where each element of the LULC life cycle and its sub-elements are seen as processes characterized by input data, output data, and resources (i.e. methods and tools to perform a process). A process can be understood as one of the scientific objectives addressed in a scientific paper. Third, to establish explicit links between the scientific literature and the components of the LULC life cycle, we define formal semantics for all relevant items, including LULC components, scientific publications, types of produced data, algorithms (used or developed), available code (if any), and tools (e.g., Geo-Wiki, QGIS). Representing these examples as instances within an ontology enables the inference of new knowledge and the automatic identification of semantic relationships between research papers, thereby facilitating more efficient synthesis, discovery, and reuse of LULC research outcomes. Finally, end-users are encouraged to provide feedback on all processes, contributing to the continuous refinement and evolution of the ontology.

3.2 Knowledge acquisition

The goal of this step is to collect the knowledge required to formalize the LULC ontology. To achieve this, we build upon the systematic review conducted by (Naboureh et al., 2021), which provides a comprehensive synthesis covering a wide range of indicators relevant to LULC research. This review is particularly suitable for our purposes, as the proposed indicators can be directly mapped to ontology elements such as classes, data properties, and object properties. This alignment ensures that our ontology reflects both the breadth of LULC research and the depth needed for semantic representation and inference. The study proposed 17 indicators:

- Publisher: Journal name.
- Year: Year of publication.
- Title: Paper’s or book title.
- Author: Author’s name.
- Citation: Number of citations.
- Case study area: Information about the region (country, province, city).
- Sensor: Name and resolution.
- Classification method: Pixel-based, object-based, machine learning algorithms, or others.
- Reference data: Source of training and validation data.
- Classification system: Number of classified categories.
- Data type: Remote sensing data, GIS data, topographic data, etc.
- Accuracy assessment: Methods for accuracy assessment.
- Research focus: Main focus of the research.
- Challenge: Identified weaknesses and strengths.
- Major cause of LULC: Main reasons of LULC (e.g., urbanization).
- Environmental issue of LULC: Environmental issues caused by LULC.
- Main LC change type: LC classes showing highest change rates.

The second source of information consisted of panel experts (i.e., the authors of this paper having a strong experience in LULC research), with two of them contributing to conceptualization and the remaining providing feedback and testing. We thus extend this list of indicators by adding five new criteria:

- Nomenclature: The naming and classification nomenclature used for LULC categories
- Code availability: The accessibility of source code or processing scripts
- Data availability: The openness and accessibility of the datasets

- Metadata availability: The presence of descriptive information about data content, quality, and provenance
- Storage: The format and platform used for storage and management.

Indeed, not only the number of categories but also the names of these categories are essential to compare classification research works. Knowing if code, data and metadata are available, as well as how they are stored, is crucial for asserting if the research papers are reproducible and align with the FAIR (Findable, Accessible, Interoperable and Reusable) principles (Wilkinson et al., 2016).

3.3 Conceptualization of the LULC Ontology

Class and properties. In our ontology, a class is defined by a human-readable *label* and *comment*, and by an *URI* that uniquely identifies it. Each class may also have multiple associated *data properties* and *object properties*, which specify its attributes and relationships with other classes, respectively. To determine whether to model an element as a class, a data property, or an object property, we adopt the following rule: use a class for things that can have instances, data properties to represent attributes or characteristics, and object properties to represent relationships.

In contrast to the systematic review methodology proposed by (Kitchenham and Charters, 2007), which consists of three main stages: defining research questions, identifying data sources, and establishing a strategy for searching, selecting and reporting papers, we introduce two additional preliminary steps.

Process. The first step consists of formalizing the notion of process as defined by (Follin et al., 2019). In our context, a process represents the main objective of a research paper (e.g., LU change detection).

A concrete example is used to illustrate the concept of a process in Figure 1. Suppose that one of the goals of the paper is to propose an approach for LU change validation. The process is Change Detection. The input data consist of a set of detected changes and two orthophotos (from 2016 and 2019). The outputs of this process include the validated changes, the corresponding LU class, and the associated contributor agreement. The workflow follows a procedure structured into four main steps: sampling design, response design, photo-interpretation (performed at a computer), and quality assessment, which evaluates both contributor performance and accuracy. The instruments required for this process include the Lacowiki tool, which supports collaborative validation, and two algorithms: one for computing contributor agreement and another for computing accuracy. Two types of operators are involved: researchers, who design the algorithms and define the workflow, and contributors, who participate in the validation of detected changes.

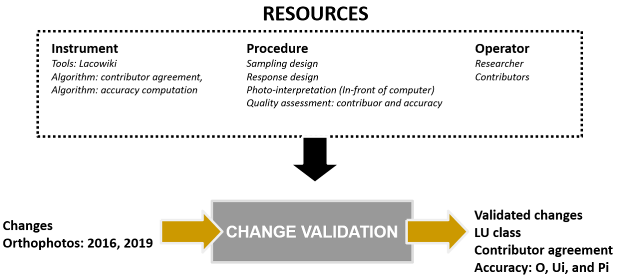

Figure 1. Example of a process: change validation.

LULC life cycle. The second step involves formalizing all elements that define the governance of the LULC data lifecycle and its associated processes. This includes specifying the relevant concepts, relationships, properties, and instances that structure the governance framework. As illustrated in Figure 2, LULC data life cycle comprises five main stages represented in light gray : *Modelling* (i.e., the process that turns a given source into a more informative level: semantization, attribute estimation, enriched encoding (for Artificial Intelligence - AI - models)), *Mapping* (i.e., the process that allows the production of geospatial data such as LU or LC), *Validation* (i.e., the process that allows to validate the geospatial data), *Manage* (i.e., the process for managing LULC data), and *Update* (i.e., the process used to update geospatial data). Each stage is seen as a type of process. Moreover, each stage can contain different other stages (depicted in dark gray in 2. For example, the stage named Update includes, Change Detection, Change validation, Integrated changes, and Remapping.

As noted before, our hypothesis is that the LULC data life cycle is not a linear workflow, each stage being considered independently, since research papers rarely address the full life cycle of LULC data. Thus, we introduce two concepts: Use (in sky blue) and User Feedback (in light blue), to describe how each process is applied and to incorporate feedback on its execution. Such feedback may come from the paper itself or from other researchers who apply the same processes in different contexts.

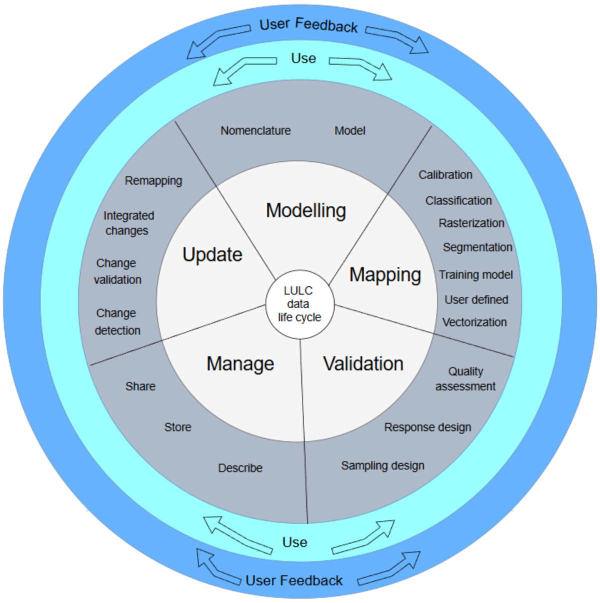

Figure 2. Formalization of the LULC data life cycle through processes and sub-processes (in light and dark gray). Each process can be used independently (sky blue), and users can provide feedback (light blue) for every process and sub-process.

Scientific Paper. The third step focuses on formalizing the concept of a scientific paper and all the entities it involves. We begin by modeling the standard metadata elements that characterize a scientific publication, as well as the proposed algorithms, used data, geographic extent, etc.

3.4 Implementation of the LULC Ontology in OWL

In this section, we present the implementation of our LULC ontology, which was developed using the well-established OWL formalism. The ontology contains 151 classes, 59 object properties, and 44 data properties. Due to space constraints, it is not possible to provide a detailed description of all *classes*, *object properties*, and *data properties* defined in the ontology. All of these elements are fully documented and available in the published version of the ontology. Following the same structure as in Section 3.3, we therefore describe here only the most relevant classes, object properties, and data properties.

Figure 3 illustrates the first level classes of the LULC ontology (i.e., they are subclasses of *owl:Thing*.

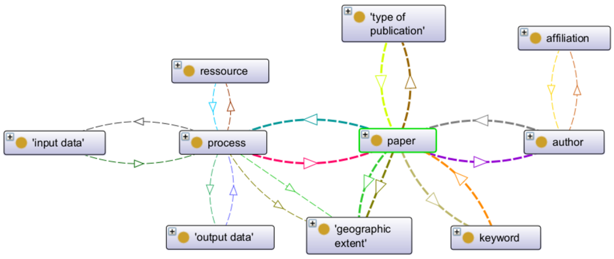

Figure 3. First levels classes of LULC Ontology and their object properties.

Process. The main ontology classes related to *Process* and their descriptions are as follows:

- *Process*: A step within the LULC data life cycle, defined by its input data, output data, and required resources. In our context, a process represents any research activity that contributes to achieving a specific research goal.
- *InputData*: Represents the data required to execute a process.
- *OutputData*: Represents the data produced by a process.
- Class *Resource*: Represents any element needed to carry out a process.
- *Instrument*: A resource that supports the execution of a process, such as tools or algorithms. It is a subclass of *Resource*.
- *Operator*: A resource describing the agent (e.g., researcher, system) responsible for performing the process. It is a subclass of *Resource*.
- *Procedure*: A resource describing the workflow or sequence of actions followed within a given process. It is a subclass of *Resource*.

LULC data life cycle. Each of the stages illustrated in Figure 2 is formalized as a subclass of the class *Process*, and thus represents both a process and a container for its own set of more specific sub-processes.

Scientific Paper. To formalize the concepts linked to a research paper, we chose the following: the concepts *Paper*, *Author*, *Affiliation*, *Keyword*, and *Type of publication* are represented as classes (*rdfs:subClassOf Thing*), while information such as *abstract*, *name* and *address of affiliation*, *country* and *research unit appearance*, *DOI*, *author names*, *ORCID*, *publisher name*, *number of citations*, *title*, and *year of publication* are formalized as data properties. This distinction ensures a clear semantic separation between entities (modeled as classes) and their descriptive attributes (modeled as data properties), facilitating the instantiation of the ontology with real scientific papers. Note that the paper also includes a set of data properties formalized through a specific criterion. This criterion helps provide a richer description of each paper and makes it easier to answer questions such as: “Which papers provide code and data for change detection?” The criterion includes the following subproperties: application, challenge, codeAvailability, dataAvailability, metadataAvailability, strength, and weakness (see Figure 4.

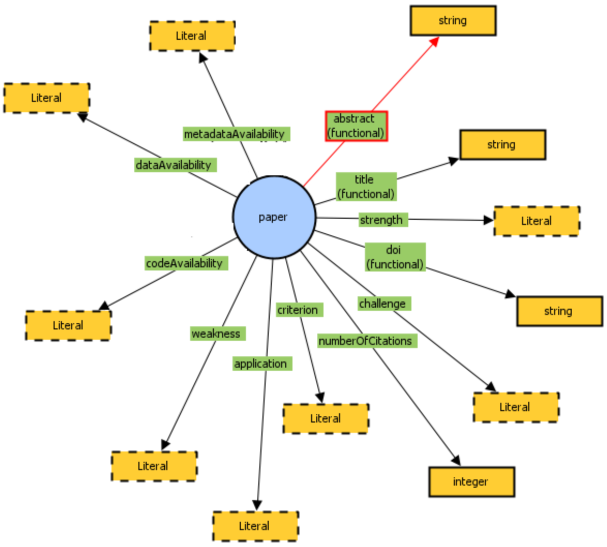

Figure 4. Data properties of the class *Paper*.

To establish the link between the process and the research paper, we also defined the following first-level classes in the ontology, which are the subclasses of `owl:Thing`:

- *Data*: This class represents the type of data which can be used in the life cycle of LULC Data. It can be spatial or aspatial data. The *SpatialData* class represents the types of spatial data that may be used throughout the *LULC* data life cycle. It is one among several data categories, including *Backbone*, *ElevationData*, *HumanStatistics*, *LandUse*, *LandCover*, *Zoning*, and others. Additionally, *AspatialData* (i.e., data without a spatial component, such as text) may also be used. Furthermore, *UserDefined data* refers to data types not yet covered by the ontology, which users may define as needed.
- *Date*: It defines the date of a product. Different temporal granularities are defined such as day, month, period, year, user-defined. This class allows to manage the temporal component in the LULC data life cycle.
- *GeographicExtent*: Defines the geographic extent coverage of a data set. Generally, it formalizes the case studies of a paper. It can be local, regional, national and global.

Let us revisit the example illustrated in Figure 1 to examine how the formalization has been carried out. The classes *Tool* and *Algorithm* are defined as subclasses of *Instrument* using the relation *rdfs:subClassOf*, and the individuals *Lacowiki*, *contributor agreement*, and *accuracy computation* are instantiated using *rdf:type* as members of their corresponding classes. Similarly, the classes *Sampling design*, *Response design*, *Photo-interpretation*, and *Quality assessment* are defined as subclasses of *Procedure* (*rdfs:subClassOf*). For example, *Blind sampling design* can be represented as an individual of the class *Sampling design* (*rdf:type*). Finally, the classes *Researcher* and *Contributor* are defined as subclasses of *Operator* (*rdfs:subClassOf*), and the authors of the paper are instantiated as individuals of the class *Researcher* (*rdf:type*). Finally, for the sake of understanding, Table 1 summarizes the formalization we proposed for the state of the art criteria defined by (Naboureh et al., 2021).

3.5 Instantiation process

The LULC ontology was implemented in OWL 2 using Protégé and WebProtégé, which provides a collaborative environment for ontology development. However, manual instantiation of individuals directly within Protégé is time-consuming and not well suited for users who are unfamiliar with ontology-editing tools. To facilitate the instantiation process and make it accessible to a broader range of researchers, we designed a tabular-based template. Each row of the tabular file represents a process defined by a paper. The values in each column describe the procedure, the instruments, the input and output data, the operator, the study case, the potential LULC classes and nomenclatures, and the values of the evaluation metrics. To automatically instantiate the ontology based on these structured files, a Python script has been developed using the Owlready2 library[1] and is available with the ontology OWL file at `https://github.com/umrlastig/Land-Use-Land-Cover-ontology-instantiation`.

4. Evaluation

The evaluation of the ontology was conducted in two iterations that allowed us to improve and strengthen its structure.
First, the group of authors who did not participate in the definition of the ontology instantiated the same paper using the provided template. We reviewed the different versions and identified missing elements as well as the difficulties encountered. A new version of the ontology and the template were produced. Second, we examined whether it could be effectively used to conduct LULC literature reviews and to facilitate subsequent analyses. We used the following protocol:

1. We collaboratively defined three use cases representing three themes to simulate the process of conducting a literature review: building function, global land cover mapping, and multi-class change detection.

2. For each use case, we formulated a set of research questions that a literature review on that topic should be able to address. These included both open and closed questions (see Table 2).

3. For each use case, we selected a sample of representative papers and instantiated them in the ontology. The number of instantiated papers per use case is provided in Table 2.

4. For each question, we developed queries to assess whether the ontology can be used to answer the questions based on the instantiated papers.

Indicator	**Formalization**
(Naboureh et al., 2021)	**(Naboureh et al., 2021)**
Publisher	isPublishedIn: *object properties* of *Paper*
Year	year date: *data properties* of *Paper*
Title	title: *data properties* of *Paper*
Author	Author: *class*
Citation	number of citation: *data properties* of *Paper*
Case Study Area	GeographicExtent: *class* with subclasses: *Global, National, Regional, Local*
Sensor Name, Resolution, MMU	Sensor data: *class*, resolution, minimum mapping unit: *data properties*
Classification Method	Classification: *class*; Algorithm: *class*
Reference Data	ReferenceDataset: *class* and its subclasses TrainingData and ValidationData
Classification System	hasNumberOfClasses: *data properties* for Algorithm and SpatialData classes
Data Type	polygonGeometricRepresentation, pixelGeometricRepresentation, pointsGeometricRepresentation, tripletGeometricRepresentation: *data properties* of SpatialData classe
Accuracy Assessment	Accuracy: *class*
Research Focus	Process: *class*
Challenges, Weaknesses, and Strengths	challenge, weakness, strength: *data properties* of Paper class
Major Cause of LULC	application: *data properties* of Paper class
Environmental Issue of LULC	Environmental: *class*
Main LC Change Type	LandUse, LandCover: *class*
New criteria	**New criteria**
Nomenclature	Nomenclature: *class*
Code Availability	codeAvailability: *data properties* of Paper class
Data Availability	dataAvailability: *data properties* of Paper class
Metadata Availability	metadataAvailability *data properties* of Paper class; has two sub properties: textReadable, machineReadable.
Storage	Store: *class*

Table 1. Enhanced indicators for systematic review and their formalization for LULC research papers: an overview.

For most of the questions, it was possible to develop a query that could accurately provide an answer based on the selected set of papers. For example, the answer given by the query to the question “What are the most frequent building function classes?” is that the class “residential” is used in 7 papers out of the 8 instantiated. The second most-frequent is “non-residential” in 3 papers. However, two questions could not be answered in the current state of the ontology. Indeed, for “What are the latest global land cover products that have appeared since the year 2015?”, the concept of product is not defined in the ontology. Moreover, the date of production of the data is not indicated,

---

[1] owlready2.readthedocs.io

only that represented by the data and the paper date. For the question “What are the papers dealing with multi-class change detection?”, in the current state of the ontology, it was not possible to differentiate between *multi-class change detection* and *binary change detection for which a LULC nomenclature was provided*. Moreover, the tabular template was not perfectly fitted for the instantiation of articles about change detection, leading to ambiguities that required manual inspection.

Use case	Question
Building function: 8 instantiated papers	Which are the articles that proposed methods to define building function?
	Which are the datasets commonly used?
	Which are the algorithms allowing to define building functions?
	What are the most frequent classes?
	What are the accuracies in predicting building function?
Global LULC mapping: 7 instantiated papers	What are the latest global land cover products that have appeared since the year 2015?
	Of these studies, which have spatial resolutions of 30 or finer?
	What are the overall accuracies of these maps?
	What are the classes used or classification systems used?
	Which validation methods have been used to validate these global land cover maps?
	I’m looking for training data for vineyards. Which are the existing research training data for vineyards?
	Which are the Voluntary Geographic Information sources used for LC mapping?
	Which are the most studied LC classes in the last 10 years?
Multi-class change detection: 3 instantiated papers	What are the papers dealing with multi-class change detection?
	Which papers use a given dataset?
	Which are the papers that validate their approach on more than 1,000,000 pixels?
	Which are the papers that do not use change vector analysis?
	Which are the datasets commonly used?
	Which change detection methods provide the best results?
	What are the metrics used to evaluate multi class change detection?
	What are the semantic classes considered?
	What are the failure cases?

Table 2. Use cases and questions defined for the evaluation of the ontology.

5. Conclusion and future works

In this paper, we propose the **OntoLULC-SOTA** ontology designed for the research community working on LULC data to support the production of systematic reviews. We formalize in *OWL*, the metadata of scientific papers, the life cycle of LULC data including methods, tools, data characteristics, their availability, strengths and weaknesses based on user feedback. To evaluate the ontology, we manually instantiate it using a selection of papers by considering three use cases: building function, global LULC mapping, and multi-class change detection.

This work is only the first step toward achieving our main goal, and one could imagine analyzing papers based on this approach by tracing all the citations from a few key review articles.

Thus, much research is still necessary to refine the ontology, automate its instantiation, and engage the community in its continuous development and validation.

Therefore, future works must address the gaps in the current instantiation of the ontology. There are two directions we plan to explore. In the short term, we aim to enhance the tabular template presented in Section 3.5 by: adding the concept of product, creating subclasses of the change detection process for semantic and binary change detection, and including a data property for the production date. Indeed, a dataset or a product can have a different date than that of the paper.

However, this approach still requires a manual step, as users must fill in the tabular file themselves, which leaves room for errors. Thus, the second long-term direction is to use Large Language Models (LLMs) to automatically extract ontology instances directly from text. This approach would allow us to build a knowledge graph (i.e. a fully automatically instantiated ontology). Powerful tools such as the Neo4j LLM Knowledge Graph Builder can extract nodes and relationships from unstructured text while leveraging an existing ontology. However, our initial tests indicate that further research is needed to correctly assign text segments to the appropriate ontology elements. Once the knowledge graph is created, we plan to implement GraphRAG technologies and develop a chatbot connected to Neo4j such as *https://scispace.com/* tool. This would offer strong advantages, including enabling the community to maintain the knowledge graph, improving explainability and traceability, and allowing users to quickly query for relevant papers, models, benchmarks, and other resources, as well as to give feedbacks. Another perspective is to link our ontology with existing vocabularies such as nanopub, wikidata or DCAT for datasets publication (Färber and Lamprecht, 2021). Future works will extend the evaluation of the ontology by defining protocols to quantitatively measure its quality. Finally, this approach will succeed if there is an active community to provide ongoing feedback on the formalization and to keep it alive. Therefore, we intend to establish new collaborations with the LULC community. Our approach enables the ontology to be enhanced with new concepts, objects, and data properties.

References

Abdelghany, A., Darwish, N., Hefni, H., 2019. An Agile Methodology for Ontology Development. *International Journal of Intelligent Engineering and Systems*, 12, 170–181.

Aria, M., Cuccurullo, C., 2017. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975.

Bucur, C.-I., Kuhn, T., Ceolin, D., van Ossenbruggen, J., 2023. Nanopublication-based semantic publishing and reviewing: a field study with formalization papers. *PeerJ Computer Science*, 9, 1–30.

Castelvecchi, D., 2025. Preprint site arXiv is banning computer-science reviews: here’s why. *Nature News (Accessed November 16, 2025).*

Chughtai, A. H., Abbasi, H., Karas, I. R., 2021. A review on change detection method and accuracy assessment for land use land cover. *Remote Sensing Applications: Society and Environment*, 22, 100482.

ElHassouni, J., El Qadi, A., 2022. Ontology engineering methodologies: State of the art. *Proceedings of the 5th International Conference on Big Data and Internet of Things (BDIoT 2021)*, Lecture Notes in Networks and Systems, 489, Springer, Cham, 59–72.

Färber, M., Lamprecht, D., 2021. The data set knowledge graph: Creating a linked open data source for data sets. *Quantitative Science Studies*, 2(4), 1324-1355. https://doi.org/10.1162/qss_a_00161.

Ferrario, R., Grüninger, M., 2020. Proposed Guidelines for Publishing Ontology Papers. *Applied Ontology*, 15(1), 1–5.

Follin, J., Girres, J., Olteanu-Raimond, A., Sheeren, D., 2019. The origins of imperfection in geographic data. M. Batton-Hubert, Éric Desjardin, F. Pinet (eds), *Geographic Data Imperfection1: From Theory to Applications*, Wiley / ISTE, 25–44.

Gonthier, N., 2026. Operational Change Detection for Geographical Information: Overview and Challenges. *International Journal of Applied Earth Observation and Geoinformation*.

Hogan, A., 2020. *The Web of Data*. 1st edn, Springer, Cham.

Kitchenham, B., Charters, S., 2007. Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE-2007-01, Keele University and Durham University Joint Report.

Lv, Z., Huang, H., Li, X., Zhao, M., Benediktsson, J. A., Sun, W., Falco, N., 2022. Land cover change detection with heterogeneous remote sensing images: Review, progress, and perspective. *Proceedings of the IEEE*, 110(12), 1976–1991.

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., Liu, Y., 2017. A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, 277-293.

Mallet, C., Le Bris, A., 2020. Current challenges in operational very high resolution land-cover mapping. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 703–710.

Naboureh, A., Bian, J., Lei, Li, 2021. A review of land use/land cover change mapping in the China–Central Asia–West Asia Economic Corridor countries. *Big Earth Data*, 5(2), 237–257.

Olteanu-Raimond, A.-M., Bunel, M., Dominguès, C., Duchêne, C., Jolivet, L., Mustière, S., Van Damme, M.-D., Favre, O., Gendner, V., 2023. A lightweight ontology for landmarks to assist rescue in mountainous areas. *Advances in Cartography and GIScience of the ICA*, 4, 15. https://ica-adv.copernicus.org/articles/4/15/2023/.

Pacheco Quevedo, R., Velastegui-Montoya, A., Montalván-Burbano, N., Morante-Carballo, F., Korup, O., Daleles Rennó, C., 2023. Land use and land cover as a conditioning factor in landslide susceptibility: a literature review. *Landslides*, 20(5), 967–982.

Peng, D., Liu, X., Zhang, Y., Guan, H., Li, Y., Bruzzone, L., 2025. Deep learning change detection techniques for optical remote sensing imagery: Status, perspectives and challenges. *International Journal of Applied Earth Observation and Geoinformation*, 136, 104282.

Peroni, S., 2016. A simplified agile methodology for ontology development. G. Schreiber, T. Kawamura, C. M. Keet, V. Presutti (eds), *Proceedings of the 12th OWL: Experiences and Directions Workshop and 4th OWL Reasoner Evaluation Workshop (OWLED-ORE 2016)*, CEUR Workshop Proceedings, 1963, CEUR-WS.org, Bologna, Italy.

Santos, P. A. d., Adami, M., Picoli, M. C. A., Prudente, V. H. R., Esquerdo, J. C. D. M., Queiroz, G. R. d., Carneiro de Santana, C. T., Chaves, M. E. D., 2025. Land use and land cover products for agricultural mapping applications in Brazil: Challenges and limitations. *Remote Sensing*, 17(13), 2324.

Shi, W., Zhang, M., Zhang, R., Chen, S., Zhan, Z., 2020. Change detection based on artificial intelligence: State-of-the-art and challenges. *Remote Sensing*, 12(10), 1688.

Singh, A., 1989. Review article digital change detection techniques using remotely-sensed data. *International journal of remote sensing*, 10(6), 989–1003.

Sure, Y., Staab, S., Studer, R., 2009. Ontology engineering methodology. S. Staab, R. Studer (eds), *Handbook on Ontologies*, International Handbooks on Information Systems, 18, Springer, Berlin, Heidelberg, 135–152.

van Eck, N. J., Waltman, L., 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.

Wang, Y., Sun, Y., Cao, X., Wang, Y., Zhang, W., Cheng, X., 2023. A review of regional and Global scale Land Use/Land Cover (LULC) mapping products generated from satellite remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 206, 311–334.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E. et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1–9.

Wu, H., Li, Y., Lin, A., Fan, H., Fan, K., Xie, J., Luo, W., 2024. A review of crowdsourced geographic information for land-use and land-cover mapping: current progress and challenges. *International Journal of Geographical Information Science*, 38(11), 2183–2215.

Xu, Q., Yordanov, V., Bruzzone, L., Brovelli, M. A., 2025. High-Resolution Global Land Cover Maps and Their Assessment Strategies. *ISPRS International Journal of Geo-Information*, 14(6). https://www.mdpi.com/2220-9964/14/6/235.

Yılmaz, C., Cömert, , Yıldırım, D., 2024. Ontology-Based Spatial Data Quality Assessment Framework. *Applied Sciences*, 14(21). https://www.mdpi.com/2076-3417/14/21/10045.

Zhang, P., Wang, J., Li, R., 2023. Tourism-type ontology framework for tourism-type classification, naming, and knowledge organization. *Heliyon*, 9(4), e15192. https://www.sciencedirect.com/science/article/pii/S240584402302399X.

Zhou, X., Xie, X., Xue, Y., Xue, B., 2021. Ontology-Based Probabilistic Estimation for Assessing Semantic Similarity of Land Use/Land Cover Classification Systems. *Land*, 10(9), 920.