TRACES

Semantic Environmental Trajectories of Territories

Summary table of persons involved in the project:

Country	Univ. or Institution	Last Name	First Name	Current position	Role in the project	T0	T1	T2	T3	T4	T5	Involvement
FR	Université Grenoble Alpes (UGA-LIG)	GENSEL	Jérôme	Professor	FR Scient. Coord. T0 Scientific Leader	x	x	x	x	x	x	16,8p.month
		VILLANOVA-OLIVER	Marlène	Assist. Professor	Other member	x		x	x			8,4p.month
		BERNARD	Camille	Assist. Professor	T2 Scientific Leader	x	x	x	x	x		8,4p.month
		GENOUD	Philippe	Assist. Professor	Other member			x	x	x		6,3p.month
		ZIEBELIN	Danielle	Professor	Other member			x	x	x		6,3p.month
		DAVOINE	P.-Annick	Professor	Other member			x				6,3p.month
	Université de Bourgogne (UB - LIB)	CRUZ	Christophe	Professor	T3 Scientific Leader	x		x	x	x	x	8,4p.month
		CHERIFI	Hocine	Professor	Other member			x	x	x		8,4p.month
		LECLERQ	Éric	Assist. Professor	Other member			x	x	x		8,4p.month
CH	Université de Genève (UG - ISE)	DAO	Hy	Adj. Professor	CH Scientific Coord. T1 Scientific Leader	x	x	x				8,4p.month
		GIULIANI	Gregory	Senior Lecturer	Other member		x	x	x	x		8,4p.month
		RODILA	Denisa	Engineer	Other member		x					2,1p.month
		CHATENOUX	Bruno	Scientific Assist.	Other member		x					4,2p.month
	Université de Genève (UG - CUI)	DI MARZO	Giovanna	Full Professor	T5 Scientific Leader	x					x	8,4p.month
		HAMEL	Nils	Scientific Adj.	Other member			x			x	2,1p.month
		MARCHAND-MAILLET	Stéphane	Assoc. Professor	T4 Scientific Leader	x				x		8,4p.month
		FALQUET	Gilles	Assoc. Professor	Other member			x	x		x	4,2p.month
FR	UGA - LIG	non permanent staff		PhD Student	Other member		x	x				36p.month
		non permanent staff		PostDoc	Other member		x	x				12p.month
	UB - LIB	non permanent staff		Engineer	Other member				x			12p.month
CH	UG - ISE	non permanent staff		Scientific Assist.	Other member		x					4,8p.month
	UG - CUI	non permanent staff		PhD Student	Other member				x	x	x	36p.month
		non permanent staff		Postdoc	Other member					x		7,2p.month

Contents

I Proposal's context, positioning and objective(s)

I.a Objectives and research hypothesis

Human activities are the main drivers of the observed environmental megatrends (EEA 2019[1]) such as climate change, biodiversity loss, use of resources, pollutions, …, which impact territories at all scales. Better study and understand the *environmental trajectory of a territory*, by selecting and using relevant, available, and comparable over time *and* space indicators, can help to master these megatrends, which appears as one of the main challenges regarding the sustainability of humankind.

In this context, the TRACES project will adopt an Artificial Intelligence (AI) approach in order to define, model, implement, enrich, interrogate, analyse, forecast and visualize the *environmental trajectory of a territory*. To achieve such a complete processing chain, the TRACES project will rely on three domains of AI: Knowledge Representation, Machine Learning and Multi-Agent Systems.

Our first research hypothesis is that the semantic approach supported in Knowledge Representation through ontological models and languages, and Knowledge Graphs (KGs) formalism and tools, can help, not only to build and implement semantic environmental trajectories of territories (SETTs), but also to enrich them with complementary knowledge relying on the Linked (Open) Data available on the Web. Second, Machine Learning offers valuable techniques, algorithms and tools to group and categorize similar SETTs in clusters, but also to extract frequent patterns in these SETTs, and then to complete then for forecasting future SETTs. The third research hypothesis we make is that a Multi-Agent System approach helps to better understand the factors that draw territories trajectories, by modelling and simulating the way territories evolve and behave under systemic constraints.

We will consider here that an environmental trajectory can be defined on the basis of a set of environmental indicators, built themselves using official data collected from surveys, various physical sensors, satellite images, etc. The environmental trajectory object produced and analysed by the TRACES project will thus give an account of the dynamics of the studied territory, and in particular, convey its evolution in terms of biodiversity, nature conservation and resilience to climate change.

The models, algorithms, KGs and agents designed by the TRACES project will contribute to a better exploration of environmental trajectories of territories as objects of study. Thus, the proposed processing chain is intended to assist professionals and experts in spatial planning in their longitudinal and comparative analysis, decision-makers in the elaboration of future environmental policies at different territorial levels, but also citizens in their understanding of the public policies implemented and the evolution of the territories they live in.

The first scientific and technical challenge for the TRACES project concerns the modelling of environmental trajectory of territory, which to the best of our knowledge, constitutes the first attempt for representing this notion by adopting a Semantic Web-based approach. The difficulty lies in both the definition of such an environmental trajectory in terms of indicators of possibly different kinds and measures, and in the coupling with its intrinsic spatial and temporal dimensions. Therefore, as a whole, an environmental trajectory of a territory appears as a complex multidimensional and multigranular object whose modelling, visualization and analysis are difficult and yet unexplored tasks. An ontological model for such an object should also include standard vocabularies in order to facilitate data integration and ensure the reusability of the KGs representing the SETTs that will be created from both this conceptual description and the values of the indicators

Exploring how Machine Learning clustering techniques can be applied to SETTs faces three main challenges: 1) providing similarity measures adapted to the representations of multimodal data evolving along the temporal axis; 2) enabling multimodal clustering of trajectories for the definition of trajectory profiles; 3) enabling temporal prediction over multimodal data representations.

The main challenges raised by modelling environmental trajectories using an executable multi-agent system lie in: 1) leveraging SETTs to derive, possibly in an automated manner, multi-agent based models reproducing past environmental trajectories; 2) combining and visualizing in the same hybrid models, spatially-explicit agents modelling human behaviour and lattice-based territorial changes; 3) exploiting a series of scenarios highlighting relevant policies to provide useful prescriptive analysis.

Lastly, several definitions of an environmental trajectory will be proposed to handle different perspectives on this study object. Indicators and data will be provided and elaborated by partner ISE, using the Swiss Data Cube platform. Taking advantage of the fact that TRACES is a joint project between Swiss and French researchers, we will choose three types of territory to conduct longitudinal and comparative analysis: 1) Swiss territories (at the state, canton and commune levels); 2) French territories (at the department, commune and metropolitan levels); 3) on both sides of the French-Swiss border (at commune and metropolitan levels). However, our approach intends to be generic and applicable to worldwide territories.

I.b Position of the project as it relates to the state of the art

In Computer Science, the notion of *semantic trajectory* appeared in the early 2010s and focused mainly on trajectories of mobilities, movements or displacements concerning mobile objects (cars, planes, boats) or pedestrians [Alvares 2007, Parent 2013]. Thus, the semantics expressed here using ontologies and Semantic Web languages mainly describe a moving object or a person, and the activities characterizing the sequence of movements and stops that form the trajectory. Regarding environmental trajectories, Bryan et al. [Bryan 2016] propose a Land-Use Trade-Offs (LUTO) model to support a comprehensive, detailed, integrated, and quantitative scenario analysis of land-use and sustainability for Australia's agricultural land from 2013–2050. LUTO is implemented as a constrained, partial equilibrium, linear mathematical programming model which integrates multiple spatiotemporal models and data layers from a variety of sources, without any explicit semantics. Trajectories [Anquetin 2018] is an interdisciplinary project dedicated to the co-evolution processes between human societies and the environment. Focusing on the Alpine territories, the aim is to co-construct and anticipate the mutations of territories affected by climate change. Although the notion of environmental trajectory is at the very heart of this project, no semantic modelling is proposed and the trajectory has to be extracted manually from time-series of relevant indicators. ESPON research has shown the multidimensional and moving dimensions of territorial processes [Dao et al. 2012]. Lehman et al. [2010] have proposed a new spatial framework based on bio-physical characteristics for an improved monitoring of the environment, in complement to territorial statistical units. However, to the best of our knowledge, observation of environmental dynamics through trajectories has not yet been tackled using a semantic-based approach as reported in [Giuliani 2020a]. Regarding the modelling of environmental trajectories through ontologies, inspiring approaches are: 1) the 4D-fluent approach by [Welty 2006] for the representation of four-dimension objects that evolve over time; 2) the Continuum model and the TSN ontological model for the representation of spatiotemporal filiation links [Harbelot 2013] and of territorial changes [Bernard 2018]; 3) the design pattern for modelling life trajectories in the semantic Web [Noël 2017]. Standard ontological models have been defined for describing indicators observations in the Web of Data (Linked Data), and so the W3C RDF Data Cube (QB) [Cyganiak 2014], O&M[1] or the SSN Ontology[2] will be considered as good starting points regarding this aspect. From this ontological model, Knowledge Graphs (KGs) linking ontologies and indicators data will be generated. For about ten years, the paradigm of KGs has aroused great interest which has largely exceeded today Knowledge Representation, the field of AI from which it originates. As defined in [Paulheim 2017]: "A KG: 1) mainly describe real world entities and their interrelations, organized in a graph; 2) defines possible classes and relations of entities in a schema; 3) allows for potentially interrelating arbitrary entities with each other, and 4) covers various topical domains.". Based on RDF triples as elementary units, KGs naturally provide a framework for data integration, unification, analytics and sharing [Hogan 2020].

In the context of environmental trajectories, KGs enrichment through the Linked Data is crucial for data exchange and sharing. With the help of the W3C standards and Linked Data principles [Berners-Lee 2009], data publishers link their data to other people's data to provide context and then enrich their KGs, extending navigation, discovery and query capabilities [Nativi 2020]. Besides, inferring capability is available through ontologies with semantic graph database for fact extraction and knowledge discovery. In the TRACES project, KGs enrichment and expansion will favour automatic link discovery approaches [Nentwig 2017]. Interlinking data using Linked Data about environmental trajectories requires a set of processes that belong to knowledge engineering. These processes are demanding and require knowledge engineering skills that, in the context of spatiotemporal data, are not limited to spatial and temporal mapping, but also deal with expert view extraction and semantic matching as shown by [Prudhomme 2020]. Moreover, the enriched KGs produced

[1] https://www.ogc.org/standards/om
[2] https://www.w3.org/TR/vocab-ssn/

by the TRACES project, will comply with the principles of FAIR data (Findability, Accessibility, Interoperability, Reusability) promoted by [Wilkinson 2016]. Regarding the visualization of KGs, recent work ([Antoniazi 2018], [De Donato 2020]) shows that exploiting their potential in a useful and relevant visual analysis process still remains a challenge.

Considering the multidimensional nature of environmental trajectories, multimodal fusion, being late or early, can be used for creating representation in view of classification [Federov 2017]. Such representations may be the basis for learning in a multimodal setting [Baltrušaitis 2019]. As shown by [Huang 2019], the temporal dimension may be inserted for prediction of events guided by diverse features. In turn, the relationship between multimodality and clustering for profiling, is directly found in the context of multi-view clustering [Fua 2020]. This initial work will be used as a reference to explore how environmental trajectories can be compared, classified and extended using these Machine Learning techniques.

Spatially-explicit agent-based models, when used in conjunction with cellular automata and socio-economic models, capture human actors' behaviour and their impact on land-use change across different spatial, temporal and dynamics dimensions [Millington 2016]. Such systems are described from diverse points of view: micro-behaviour of both individuals and specific land areas or environments, their diversity and heterogeneity, and their specific impact on each other, but also meso- or macro-behaviour at a regional level for instance, providing a representation at different scales across space and time [Giuliani 2020b]. Combining agent-based models of human actors and land change models is well suited to investigate the use of policies [Le 2008], at multi-scale levels [MameLuke 2001]. Recently, knowledge graphs are used to manually inform the agent-based model [Démare 2014] or to identify its relevant features [Heckel 2017]. We will build on this body of work to address environmental trajectories through a multi-agent approach.

I.c Methodology and risk management

I.c.1 Organization of the project

To achieve the scientific and technical objectives described in section I.a, the TRACES project relies on one classical Coordination, Valorisation and Transfer Task (Task 0) plus five scientific *Tasks* (Tasks 1 to 5). Each task is led by one partner and structured in *Activities*. Figure 1 shows the six tasks and their interactions that can be interpreted in terms of input-output. Task T1 elaborates the environmental trajectories, the indicators they rely on, the associated data, and selects the case studies and the validation scenarios. Task T2 builds on the results of task T1 (arrow 1) to develop the ontological model, as well as the RDF Data cube constructed as a semantic repository of indicators and data to be used. It generates the KGs of semantic environmental trajectories and proposes an adapted multi-modal visualization. KGs developed by task T2 are then enriched by task T3 (arrow 2), which, conversely, may lead to the modification of the ontological model of the environmental trajectories (arrow 3). The analysis of the trajectories is carried out by tasks T4 and T5. Task T4 performs comparative and classificatory analysis on the environmental trajectories using Machine Learning approaches. From the KGs delivered by task T2 (arrow 4) and by T3 (arrow 5), it elaborates and displays recurrent patterns extracted from environmental trajectories and the clusters they form. It also exploits and adapts sequence prediction techniques to predict future environmental trajectories. The results of the analysis carried out in this task will be used to complete enriched graphs produced by task T3 (arrow 6). Task T5 develops an agent-based approach to describe and better understand the environmental trajectories of territories in terms of behaviour, influencing factors and interaction, and finally to propose a predictive analysis according to various evolution scenarios. Task T5 uses KGs produced by task T2 (arrow 7). Knowledge on environmental trajectories used by the agents will be provided by task T3 (arrow 8). Conversely, the knowledge produced or used by the agents will complement the graphs enriched using Linked Data produced by task T3 (arrow 9). Similarly, the knowledge resulting from the comparisons and predictions of task T4 can also be exploited by the agents (arrow 10). The agents will build on the case studies and scenarios developed by task T1 (arrow 11). Lastly, the results produced by Tasks 2 to 5 will be used by task T1 (arrow 12) for a "lessons learned" activity, which can be seen as a process of continuous improvement through a permanent feedback all along the project.

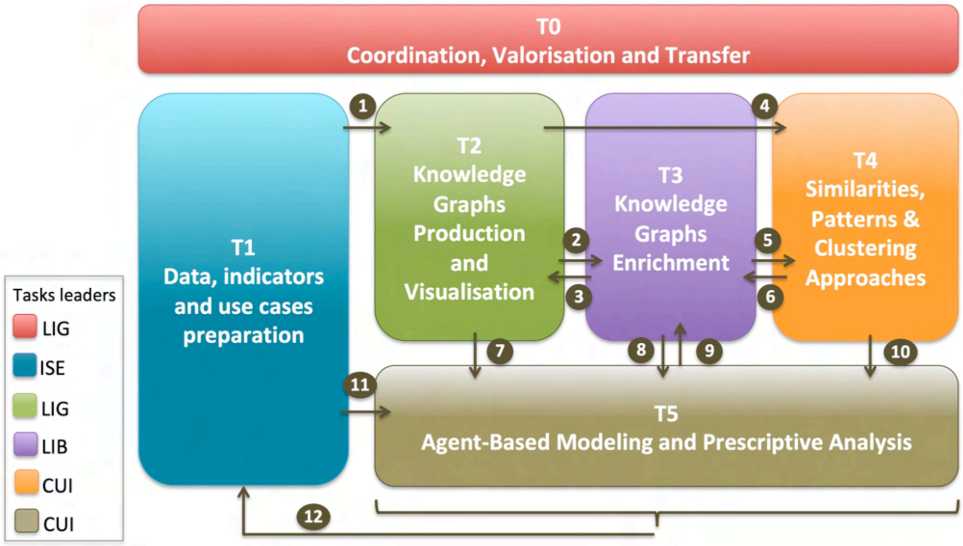

Figure 1 – Project Tasks Overview.

I.c.2 Description of the tasks

T0	Coordination, Valorisation and Transfer	Task leader: Jérôme GENSEL - LIG
Participants: M. Villanova-Oliver, C. Bernard (**LIG**); H. Dao (**ISE**), G. Di Marzo, S. Marchand-Maillet (**CUI**), C. Cruz (**LIB**)		

This task carries out the coordination, management and valorisation of the project. It coordinates the interactions between the other tasks and ensures that the project objectives are achieved, the expected deliverables are in line with the objectives defined and respect the deadlines set, and the results produced within the project are compatible. The project coordinator will be responsible for the representation of the consortium to the ANR authorities, the valorisation of the project, at the national and international levels, and the sustainability of the results obtained.

A0.1 - Project Web Site Management: aims to design and update a Web site in order to facilitate the valorisation of the project and the dissemination of the results. It will contain information describing the objectives of the project, the progress and results obtained, publications, prototypes, etc.

A0.2 - Animation and follow-up: aims to animate and carry out the follow-up of the project. A monitoring committee composed of Tasks leaders will be created. Bi-annual plenary meetings and annual meetings of the monitoring committee will be organized. The objective of the plenary meetings is to share and assimilate the concepts and methods proposed by the partners, and to decide on orientations according to the results obtained. The meetings of the monitoring committee will focus on financial and scientific assessments. The project coordinator will be responsible for the writing of intermediate and final reports. Bilateral exchanges and long visits between partners will be encouraged.

A0.3 - Communication, promotion and transfer: aims to communicate and valorise the project through different communication channels: publications and presentations in national and international conferences, social networks and media. A closing seminar will be organized to present the final results of the project. This activity also encompasses the description of the Data Management Plan and the definition and monitoring of the transfer strategy.

T0 Deliverables

Activity	N°	Caption	Type	Deadline
A0.1	D0.1	TRACES Project Web Site	Web Site	M1 (+ update)

A0.2	D0.2.1	Intermediary and final reports	Report	M21, M42
	D0.2.2	Data Management Plan	Report	M6
A0.3	D0.3	Closing Seminar	Seminar	M42

T1 Data, indicators and use cases preparation	Task leader: Hy Dao – ISE
Participants: G. Giuliani, D. Rodila, B. Chatenoux (**ISE**); C. Bernard, J. Gensel, (**LIG**); PhD Student (**LIG/ISE**), PostDoc (**LIG**), Scientific Assistant (**ISE**)	

This task provides the necessary conceptual and empirical territorial information that will be used throughout the project. It will ensure that the appropriate study areas, units of observation and thematic dimensions are identified, that the same definitions and data are shared by the different Tasks of the project. A high-performance spatial data infrastructure will be established for the maintenance, provision and update of statistical and geospatial data for the project partners, as well as for the dissemination to external users. This infrastructure will be based on open-source software, standards and technologies, enabling an easy by external and future users.

A1.1 - Identification of relevant themes, indicators and spatial unit: aims to identify the environmental themes (*e.g.*, land degradation, land cover change, urbanization), indicators and spatial units of relevance for the analysis of environmental trajectories of territories in Europe. A particular attention will be given to soil, land, biodiversity and climate topics, using as much as possible official national data sources (*e.g.*, swisstopo, IGN). The indicators should be as compatible as possible with national and international policy frameworks, such as Sustainable Development Goals (SDGs), multilateral environmental conventions, European environmental monitoring and reporting systems. Two main types of spatial units will be considered: irregular (*e.g.,* administrative units) and regular (*e.g.*, cells of the European 100m grid).

A1.2 - Definition of use cases: aims to define use cases in 3 territories (1 in France, 1 in Switzerland, 1 Swiss-French transboundary area). The use cases should present dynamic territorial and environmental processes subject to identified public policies, be covered by time-series of statistical and geospatial data (provided by swisstopo and IGN) with various spatial resolutions, allowing for spatiotemporal and thematic comparisons. They should ensure sufficient information quantity and variability for the analysis to be conducted in Tasks 2, 3, 4 and 5. They will be selected on the basis of the themes, indicators and spatial units defined in A1.1. A candidate case to be tested is land cover as measured by the European-wide CORINE land cover time-series (1990, 2000, 2006, 2012, 2018).

A1.3 - Data acquisition, structuration and dissemination: aims to prepare data for their efficient analysis in Tasks 2, 3, 4 and 5 and their maintenance through time. This involves data collection (*e.g.*, from statistical offices, remote sensing data portals), organization and storage of data in a high-performance spatial data infrastructure, development of data access services (based on standard technologies). A dedicated platform will be set up (using technologies such as the Swiss Data Cube (developed by ISE) for remote sensing data, and PostgreSQL/PostGIS database for statistical data…). The infrastructure will rely on widely adopted interoperability standards recommended by the Open Geospatial Consortium (OGC) and the International Organization for Standardization (ISO).

A.1.4 - Lessons learned: aims to extract the lessons learned (*e.g.*, benefits, limitations, perspectives) of the different technical components of the project (T1-2-3-4-5), as well as the different selected use cases. It will allow providing users who are interested in implementing the proposed approach with recommendations and help the way for future research collaborations and ultimately improve information delivered to decision-makers. All elements of this activity will be documented in the intermediary and final reports of T0 (Activity A0.2, Deliverable 0.2).

T1 Deliverables				
Activity	**N°**	**Caption**	**Type**	**Deadline**
A1.1	D1.1	Documented lists of themes, spatial objects and indicators	Metadata	M3
A1.2	D1.2	3 well-documented use cases	Report	M5
A1.3	D1.3	A data infrastructure of statistical and geospatial data	Data Cube, Web database	M6

T2 Knowledge Graphs Production and Visualization — Task leader: Camille Bernard – LIG

Participants: J. Gensel, D. Ziebelin, P. Genoud, M. Villanova-Oliver, PA Davoine (**LIG**); H. Cherifi, E. Leclercq, C. Cruz (**LIB**); H. Dao, G. Giuliani (**ISE**); G. Falquet, N. Hamel (**CUI**); PhD Student (**LIG/ISE**), PostDoc (**LIG**)

This task first elaborates an ontological model for the representation of semantic environmental trajectories. Second, it builds from the indicators and their observations extracted from the infrastructure (T1), semantic environmental trajectories, in the shape of KGs. Third, various modes of visualization of the semantic environmental trajectories will be designed and developed.

A2.1 - Modelling semantic environmental trajectories: aims to define an ontological model for the representation of environmental trajectories. Based on the W3C RDF Data Cube standard, a vocabulary will be defined to link indicators observations over time, even in complex cases such as time-series break due to changes in the geographical support. These multidimensional observations, linked together in time, form *semantic environmental trajectories* at a conceptual level. The spatial dimension of trajectories will be described using the TSN ontology. Coupled with the TSN-Change ontology, it will be possible to informed about changes in the administrative divisions of the territory observed in the environmental trajectory.

A.2.2 - Generating semantic environmental trajectories: aims to populate the environmental trajectory conceptual model built in A2.1 with data identified in T1, transformed in RDF and described using the vocabulary defined. A program will be developed in order to create KGs representing the semantic environmental trajectories. On the basis of input parameters (an indicator, a period of observation, and a chosen territory) describing the trajectory to be built, the program retrieves the corresponding data from the T1 infrastructure to generate the semantic environmental trajectory in the shape of a KG. All the KGs will be published in a triplestore. Several open source triplestores will be tested during this activity in order to evaluate their scalability and performance for space-time queries in collaboration with the A3.1 activity.

A2.3 - Visualizing semantic environmental trajectories: aims to propose various visualization modes of the KGs built in A2.2. Because of their format and of the multidimensional information they hold, we will explore various visualization modes for the restitution and analysis of environmental trajectories: aspatial visualization of SETTS as RDF graphs, visualization of clusters of trajectories, visualization of environmental trajectories through interactive geographical maps to account for territorial dynamics, visualization through evolving spatiotemporal cubes. The Swiss Territorial Data Lab (STDL)/swisstopo 4D framework will be considered here as well as open-source existing tools such as Gephi[3], CubeViz[4] and D3.js library.[5]

T2 Deliverables				
Activity	**N°**	**Caption**	**Type**	**Deadline**
A2.1	D2.1	Environmental Trajectories Ontological Model	Ontological model	M12
A2.2	D2.2	Knowledge Graph of raw Environmental Trajectories	KGs in a triplestore	M24
A2.3	D2.3	Prototype for visualization of semantic environmental trajectories	Software Prototype	M39

[3] https://gephi.org/
[4] http://cubeviz.aksw.org/
[5] https://d3js.org/

T3	Knowledge Graphs Enrichment	Task leader: Christophe Cruz - LIB

Participants: C. Bernard, J. Gensel, D. Ziebelin, P. Genoud, M. Villanova-Oliver (**LIG**); H. Cherifi, E. Leclercq (**LIB**); G. Giuliani (**ISE**); G. Falquet (**CUI**); PhD Student (**CUI/LIB**); Engineer (**LIB**)

This task firstly provides some technical prerequisites such as data handling and storing. Secondly, this task focuses on the enrichment process of environmental trajectories through Linked Data. Also, it makes available to expert enriched KGs views for specific data consumption requirements for T4 and T5. Environmental trajectories KGs gather several types of information and knowledge accumulated with rich multidimensional indicators from T1, from agent-based simulations, pattern extraction processes, and expert conceptualization. Expert views extraction algorithms are then required for the T1 use cases.

A3.1 - Uplift, controlled vocabulary, and environmental trajectory storing: aims to provide the capability to semantically uplift data using relevant controlled vocabularies for trajectory enrichment. The semantic uplift step refers to the approaches allowing the transformation of data set contents into local ontologies composed of RDF triples. The LOV[6] (Linked Open Vocabularies) repository composed of more than 750 vocabularies over a huge variety of domains, will be used. Selecting the relevant controlled vocabulary to uplift data is essential to reach sharing and exchange of data and knowledge in a FAIR way. Also, providing an easy and efficient access to the environmental trajectories' KGs is also a crucial point. For this purpose, we will test several open-source triplestores in terms of storing and querying capabilities against our enriched environmental trajectories' data sets.

A3.2 - Trajectory enrichment: aims to enrich environmental KGs produced by Task 2, with vocabularies identified in A3.1. This activity refers to graph expansion by data linking using techniques such as entity resolution, duplicate detection, reference reconciliation, link prediction, ontology matching, missing values prediction/inference, that will be adapted here to the context of SETTs' KGs. It also handles knowledge validation dealing with errors and ambiguities, and, finally, knowledge discovery using automatic reasoning. This enrichment provides advanced knowledge for T4 and T5. For agent-based models, it is essential to identify the different actors involved and their interactions, including the environment. Similarly, additional knowledge can help in measuring the similarity between trajectories with multidimensional indicators.

A3.3 - Expert view extraction: aims to produce expert views extracted from the enriched KGs. Considered as upper-level application ontologies, KGs supporting multiple disparate users' domain knowledge requirements are often too large or too complex, depending on the knowledge complexity for any specific actor. Users of such ontologies are usually not ontologists or computer scientists. Thus, the global view of the KGs may not precisely match the views required by users. The concept of expert view extraction aims to provide an understandable, specialized, and lightweight portion of a KG that fits well the T3 and T4 expert needs for SETTs insight.

T3 Deliverables

Activity	**N°**	**Caption**	**Type**	**Deadline**
A3.1	D3.1	Triplestore and controlled vocabularies (FAIR) for enriched environmental trajectories	Technical reports	M18
A3.2	D3.2	Trajectory enrichment process definition	Technical reports	M30
A3.3	D3.3	Expert view extraction algorithms	Algorithms and Technical reports	M39

T4	Similarities, Patterns and Clustering Approaches	Task leader: S. Marchand-Maillet - CUI

Participants: C. Bernard, J. Gensel, D. Ziebelin, P. Genoud (**LIG**); H. Cherifi, E. Leclercq, C. Cruz (**LIB**); G. Giuliani (**ISE**); PhD Student (**CUI/LIB**); PostDoc (**CUI**)

[6] https://lov.linkeddata.es/dataset/lov

This task enables inter-trajectory matching and comparison. While the above tasks initially consider trajectories as a model (T2) and then individually (T3), there is a high interest in being able to relate, match and compare trajectories. The basis for such operations is the definition of inter-trajectory similarity measures. Based of similarity one natural operation is to group similar trajectories or, equivalently, to define profiles for categories of trajectories. The knowledge of trajectory profiles can then serve for prediction and even help for simulations (T5).

A4.1 - Similarity measures for trajectories: aims to install a metric over trajectories that will, in turn, enable similarity computation. As defined in T2 and T3, trajectories will consist of rich multidimensional indicators along a temporal axis. Then, this metric will account for essentially two main aspects. First, it will enable the fusion of metrics over the set of indicators associated with an environmental trajectory. The challenge is to account for both the rich structure of these indicators and their potential correlations. This will be based on the semantic structure defined in T2 and resolved in close collaboration this task. The second major challenge is to handle asynchronous evolutions and potentially partial temporal matching of trajectories. The definition of an internal latent structure within the trajectory (A4.2) will help enabling partial comparisons. The main contributions will be the definition of a similarity model over (potentially part of) trajectories, induced by the specific metric installed over the space of trajectories, to feed all subsequent trajectory mining operations.

A4.2 - Trajectory profiling: aims to structure a dictionary of components to encode trajectories. This encoding will naturally lead to trajectory clustering, using similarity measures produced in A4.1 and enable the grouping of similar trajectories. In parallel, such an encoding will define a trajectory profiling, which will exhibit latent factors (from the dictionary of components) that will prove useful for prediction (see A4.3). Further, the definition of a dictionary of components will help the decomposition of trajectories into base elements and therefore facilitate their partial match (A4.1). The main challenge here is to define the base dictionary components. The most basic approach is to create a dictionary of components using a low-rank decomposition applied on a set of randomly sample parts of trajectories. An evolution of this approach may use statistical learning (encoders) to learn such a dictionary. Interpretability of these components may then appear as a regularizing constraint, in relation to the semantic structure of the trajectory (T2), in order to make any encoding legible and interpretable.

A4.3 - Trajectory prediction: aims to predict future environmental trajectories. The typical Machine Learning apparatus for sequence prediction, such as LSTM using recurrent neural networks may be used for prediction, either at a global level or by interfering over local components. This activity does not aim at innovating in the field of Machine Learning tools for sequence prediction but rather to setup and evaluate the capability of such state-of-the-art sequence methods in the context of encoded multimodal temporal trajectories. It will enable trajectory forecast as a base for decision making and policy evaluation.

T4 Deliverables

Activity	N°	Caption	Type	Deadline
A4.1	D4.1	Model for measuring the similarity between trajectories	Technical Report	M18
A4.2	D4.2	A method for dictionary learning over trajectories and its use for trajectory encoding and profiling	Technical Report	M30
A4.3	D4.3	A report of the use of sequence prediction techniques for trajectory prediction	Technical Report	M39

T5 Agent-Based Modelling and Prescriptive Analysis	Task leader: Giovanna Di Marzo - CUI
Participants: J. Gensel (**LIG**); C. Cruz (**LIB**); G. Falquet, N. Hamel (**CUI**); PhD Student (**CUI/LIB**)	

This task develops spatially explicit computational agent-based models encompassing human actor behaviour, human actor-environment dynamics and its impact on land cover, based on land cover policy changes. We will develop forecasting agent-based models by integrating the past evolution (enriched KGs (from T3), trajectory profiling (from T4), and satellite images data (from T1)), with environmental and socio-economic data and investigate future scenarios for a prescriptive analysis on land cover trajectories.

A5.1 Multi-agent modelling: aims to develop a series of spatially explicit agent-based models, capturing diverse human actors/stakeholder behaviour, decisions, adaptations to policies or environmental conditions, including socio-economic factors, and providing different actors' strategies. To this end, it is essential to identify all the human actors/stakeholders involved in the observed system, the interactions among them and with the environment (land). We will also consider multi-scale aspects as actors may act individually, collectively or as an institution at diverse levels. This activity leverages information provided by enriched KGs (T3), trajectory profiling (T4), and land images (T1), and provides KGs-driven agent models.

A5.2 - Impact on land cover and trajectories: aims to model the impact of human actors' behaviour on the environment, as well as various dynamic aspects and behaviour such as choice of crops, urban development and their effect on land degradation. Similarly, we will consider the effect of land change on the human actors' behaviour (*e.g.,* changing strategy). We will attempt at reproducing land change, land use, or urban sprawl patterns, identified from enriched KGs (T3), trajectory profiling (T4) and land images (T1). These hybrid models, combining human actors' behaviour and land changes, will display spatiotemporal evolution and multi-scale aspects (micro, meso and macro-scale, local vs regional). For land cover trajectories, we will use 2D multi-scale lattice agent-based models (*e.g.*, cellular automata). This activity provides a combined model of human actors/stakeholder behaviour and their impact on land trajectories. We will validate our models by identifying to which extent they: 1) exhibit similar behaviour, profile, or effect on the land (T4); 2) follow history of land trajectory as provided by Enriched KGs (T3); 3) share spatiotemporal similarities with the observed satellite data (T1).

A5.3 - Forecasting and prescriptive analysis: aims to automate the development of various policies scenarios, by integrating those policies in the models above (A5.1), in order to derive spatiotemporal forecasting of the land cover in various scenarios. Policies are important to consider because they have an impact on human actor's behaviour, and, through them, on land use and land change. We will, on the one hand, identify the mechanisms at work behind the dynamics of land-use change (*e.g.*, explaining the changes), and, on the other hand, anticipate the effect of policies on a land trajectory (*i.e.* providing a prescriptive analysis).

T5 Deliverables

Activity	**N°**	**Caption**	**Type**	**Deadline**
A5.1	D5.1	Agent-based models	Executable models	M18
A5.2	D5.2	Impact on land cover and trajectories	Executable models	M30
A5.3	D5.3	Forecasting and prescriptive analysis	Executable models, report, Phd thesis	M39

I.c.3 Gantt diagram

Figure 2 presents a Gantt diagram that covers the project duration (42 months). Please note that the activities that compose tasks may have two phases: one main phase (in a darker colour) which leads to the release of the associated deliverable (white stars in the Gantt correspond to the "Deadline" column in the tables of task deliverables above), and another phase (in a clearer colour) during which some additional work and upgrade could be required for adaptation purpose to other activities results. This second phase might then lead to a revision of the deliverable (black stars in the Gantt).

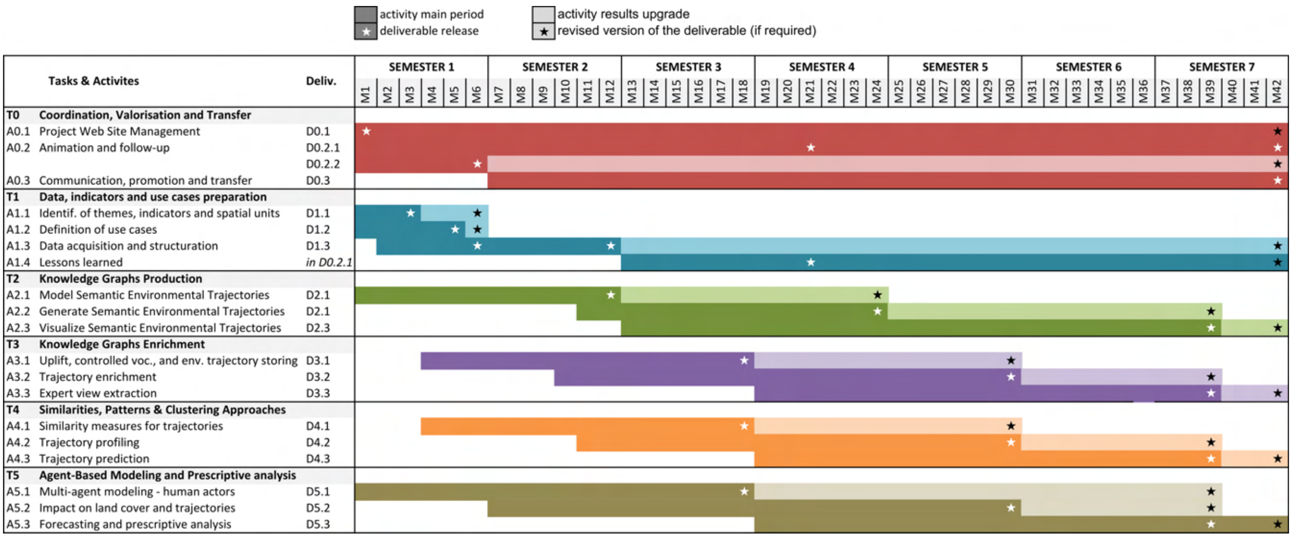

Figure 2 – Project Gantt Diagram

I.c.4 Risk management

Task	Risk	Likelihood	Impact	Mitigation
All	Staff Recruitment Difficulties	Medium	High	Anticipate the publication of position offers. Large diffusion through specialized mailing lists. Creation of multi-partners recruitment committees
All	Risks related to the sanitary situation	Medium	Low	All the work in the project can be done using virtual and remote communication between partners. No on-site or field survey/study is needed
T1	Difficulties to identify relevant indicators, time periods and study areas	Low	Low	Use a core indicator whose data is free and open online and that covers large territories (e.g., Corine Land Cover)
T2	Lack of genericity of the SETTs model to handle the description of any trajectory	Medium	Low	LIG/ISE thesis co-supervision will help. Adopt a cyclic approach to develop the model from a representative panel of basic environmental indicators, then progressively consider additional indicators
T2	Technical issues in generating KGs representing SETTs	Low	Low	A lot of open-source tools exist to assist the transfer from a DB schema or an ontology model towards KGs.
T2	Failure in handling visualization of SETTs	Low	Low	Free tools and libraries exist for graphs and KGs visualization. Also, partner LIG has a strong experience in geovisualization
T3	Difficulties in Semantic uplift of SETTs	Low	Medium	More than one thousand data sets are available (*e.g.,* Wikidata) in the Linked Open Data Cloud, as well as vocabularies and open-source triplestores.
T3	Lack of interoperability with T4 and T5	Medium	Medium	CUI/LIB thesis co-supervision will help.
T3	Difficulties in defining expert view extraction algorithms	Low	Low	Rely on a set of core existing algorithms available in the literature
T4	Define Similarity Measures for multidimensional data	Low	Low	Rely on the expertise of CUI and LIG partners in similarity measures definition. Consider the simple case of a SETTs based on the Corine Land Cover indicator only.
T4	Lack of training data to build the model and make predictions	Medium	High	The Swiss Data Cube data warehouse (designed and developed by partner ISE) contains a large amount of data already considered useful for the project
T5	Agent-based model and impact on land cover too complex to develop	Medium	High	Develop the agent-based model step by step, starting with series of well-identified territories and trajectories, exemplifying various behaviours and land changes
T5	Difficulties in building the agent-model dynamically from the enriched KGs	Medium	Low	Revert to identifying and manually programming the changes that occurred upon the territory considered

Task	Risk	Likelihood	Impact	Mitigation
T5	Forecasting and predictive analysis difficult to validate	Medium	Low	Validation of the models on historical data to refine and adapt the models, so that future forecasting is sound

I.d Ability of the project to address the research issues covered by the chosen research theme

The TRACES project is an interdisciplinary project bringing together computer scientists and geographers who propose to explore the notion of *environmental trajectory of a territory* by leveraging three fields of Artificial Intelligence. For each of them, we point out here the methods, languages or techniques used, and, in return, we indicate the expected contributions of the TRACES project in each field, *i.e.,* the avenues to be explored, the new perspectives and expected results.

In the field of Knowledge Representation, more particularly the Semantic Web, ontologies and KGs will be exploited jointly to produce a conceptual model of environmental trajectories populated by observations data. The RDF Data Cube vocabulary will be used to build a multidimensional semantic cube describing indicators, territories and periods constituting the environmental trajectories studied, based on the Swiss Data Cube spatial data infrastructure containing all the data used by the project. Then, in the field of Knowledge representation and KGs, the TRACES project aims to:

- develop a generic ontological model of trajectories of territory, one specialization of which will be an ontological model dedicated to environmental trajectories of territory;
- propose a methodology based on the exploitation of RDF Data Cube as a vocabulary facilitating the semantisation of territorial statistical data warehouses;
- show that KGs are well suited to the representation, navigation and interrogation of multidimensional and multi-granular data, such as environmental trajectories;
- test the scaling performance of the main open-source KGs management tools (triplestores);
- explore the modalities of an interactive, synchronized and multi-modal (graphs, curves and time graphs, dynamic and animated maps) visualization for environmental trajectories and spatiotemporal KGs;
- adapt and extend existing methods (entity linking, link prediction, etc.) for a controlled enrichment (quality, relevance, etc.) of environmental trajectories by searching relevant KGs in the Linked Open Data Cloud.

In the field of Machine Learning, the project addresses the modelling, mining and prediction of semantic trajectories under several perspectives, corresponding to the full Machine Learning pipeline:

- extract adequate (compact and informative) representations from enriched trajectories, taken as complex objects akin to multimodal information along a temporal dimension. Such a representation is key to all subsequent learning and mining operations.
- recognize similar trajectories in order to define profiles that will enhance the legibility of the corpus of trajectories. Installing adequate similarity will open the way to applying and developing further access and interaction mechanisms with the data.
- understand and model the information and temporal dynamics of trajectories to install predictive techniques over the corpus of trajectories and serve subsequent tasks for simulation and forecasting.

In the field of Multi-Agents systems, models that combine both human actor behaviour and their impact on the land, but also land-specific parameters and socioeconomic factors, provide an important complement to current approaches in order to analyse targeted policies scenarios. Besides forecasting and predictive analysis of various scenarios, these models give spatiotemporal insights into the roles played by policies, decision makers, individuals, institutions, and the dynamics behind land-use change. The use of KGs to derive agent-based models is still a research field in its infancy. The TRACES project aims to advance the state of the art as follows:

- derive agent-based models for both the human actors and the impact on land, by identifying and extracting the underlying driving mechanisms, and the main features of the models, directly from the enriched KGs.
- provide insights (explanations and visualization) into the dynamics of the system, through agent-based environmental trajectory models.
- automate the construction of the agent-based models in order to investigate large numbers of prescriptive/forecasting scenarios probing various policies.

II Organisation and implementation of the project

II.a Scientific coordinator and its consortium / its team

II.a.1 Presentation of the scientific coordinator

French scientific coordinator: Jérôme Gensel is a full-professor in Computer Science since 2007 at the Université Grenoble Alpes (UGA). He is a member of the LIG (Grenoble Computer Science Laboratory), where he led the Steamer research team from 2010 to 2014 (about twenty members then). With an extensive research experience in artificial intelligence, knowledge representation and constraint programming, he conducts research on spatial and temporal information systems and in geomatics (or geographic information science). His research is mainly applied to the fields of land use planning and natural hazard prevention. He has supervised 16 doctoral theses. Three of his most recent theses are related to the proposed project. He has been the coordinator of a dozen European and international research projects, notably in the framework of the European Commission's ESPON programme, dedicated to spatial planning and environment themes. He was director of the CNRS MAGIS French national research group on spatial information and geomatics from 2013 to 2016.

Swiss scientific coordinator: Hy Dao is Adjunct Professor at the University of Geneva and Head of Unit at UNEP/GRID-Geneva. He holds a post-graduate diploma from the International Institute for Aerospace Surveys and Earth Sciences (ITC, Enschede, The Netherlands) and a PhD in human geography from the University of Geneva. He was leader and expert in several EU and Swiss funded projects on data and indicators of territorial cohesion and environmental situation. He led several studies on Planetary Boundaries at the Swiss, European and global levels. He contributed to the risk analysis in the UNISDR Global Assessment Report series. He recently was responsible of the University of Geneva team in the H2020 PLACARD project – PLAtform for Climate Adaptation and Risk reduction.

II.a.2 Presentation of the consortium

Université Grenoble Alpes (UGA) - Grenoble Computer Science Laboratory (LIG) / STeamer. STeamer is a research group in Computer Science which leads research in Geographic Information Science, focusing on the spatial and temporal dimensions of data. Research studies of STeamer focus on collecting, modelling, querying, reasoning and visualizing spatial and temporal data, in particular using Semantic Web languages and technologies. STeamer has led or been involved in many funded scientific projects at the national, European and international levels.

- Marlène Villanova-Oliver is an Assistant Professor in Computer Science at UGA. She's the leader of the STeamer research group. Her research interests and skills concern spatiotemporal information modelling, visualization and reasoning.
- Camille Bernard is an Assistant Professor in Computer Science at UGA. Her PhD Thesis defended in 2019 dealt with the evolution of administrative divisions and their immersion in the semantic Web using KGs.
- Paule-Annick Davoine is a Full-Professor in Geography at UGA. She graduated in Computer Science and geographical information sciences. Her research interests and competencies are in spatiotemporal visualization, dynamic cartography and interfaces.

- Philippe Genoud is an Assistant Professor in Computer Science at UGA. His research interests and skills are in the Semantic Web, programming tools and technologies.
- Danielle Ziébelin is a Full-Professor in Computer Science at UGA. Her research interests and expertise are in AI fields of knowledge representation and reasoning.

The **Centre Universitaire d'Informatique, of the University of Geneva (CUI)** is an interfaculty center for research and teaching in Computer Science, Information Systems and Informatics Linguistics. As relevant for TRACES, CUI develops state of the art scientific knowledge in ontologies, context-aware systems, machine learning, and agent-based systems. Established in 1976, CUI has a long experience of competitive research projects funded by various research agencies (among others SNF, H2020, COST, Innosuisse, etc.).

- Prof. Giovana Di Marzo Serugendo is Full Professor in Computer Science and Information Systems and currently the Director of CUI. She is an expert in self-adaptive, context-aware systems, distributed collective AI systems with multi-agent systems modelling and engineering. She is involved in various projects (SNF, Innosuisse, private mandates) in domains, such as the Internet of Things, ecosystems of services, smart grids, robotics, and spatial data. Her work develops state of the art research involving digital twins, ontologies, and multi-agent systems.
- Prof. Gilles Falquet is an expert in knowledge engineering, using ontologies in various settings (texts, images, social networks, etc.). His research activities include ontology management systems, point of views in ontologies, knowledge-based indexing and information retrieval, and integrating knowledge visualization in 3D interfaces.
- Prof. Stéphane Marchand-Maillet is Associate Professor and leads the CUI's Viper research group. His research is directed towards large-scale, high-dimensional distributed machine learning and information mining and indexing, with applications to data modelling, recommendation and prediction. He and his group are part of several national and European and international projects in the domain. He is involved in the organization and steering of several international scientific events.
- Mr Nils Hamel is scientific assistant at CUI, working on the Swiss Territorial Data Lab (STDL) project in collaboration with swisstopo. He developed the STDL/swisstopo 4D framework and contributes to its extension. He has advanced expertise on data science activities analysing various signal data (Lidar, satellite, hyperspectral, land register).

The **Institute for Environmental Sciences (ISE)** is an inter-faculty entity of the University of Geneva that undertakes research and teaching activities in the numerous inter-connected domains of the environment, such as climate, water, biodiversity, health, energy, urban ecology, environmental governance and territorial development. ISE is specialized in performing spatially-explicit modelling, by handling and analysing spatial and statistical data on environmental and natural resource issues through Geographic Information Systems (GIS) and remotely-sensed imagery. The team also specialized in down-scaling and upscaling of environmental data and indicators, and in developing geocomputation workflows with distributed computing infrastructures linked to spatial data infrastructures, using international (meta)data standards (*e.g.,* ISO, OGC) and initiatives (*e.g.,* GEOSS, INSPIRE). Capacity building on Spatial Data Infrastructures (SDI), Open Data Cube (ODC) and GEOSS- related issues is also a key asset of the team.

- Dr. Gregory Giuliani, a Senior Lecturer at the University of Geneva, is a geologist and environmental scientist who specializes in Geographical Information Systems analysis and Spatial Data Infrastructures (SDI). He also works at GRID-Geneva of the United Nations Environment Programme since 2001, where he was previously the focal point for SDI. He is currently the Head of the Digital Earth Unit and Swiss Data Cube project leader. His research focuses on Land Change Science and how Earth observations can be used to monitor and assess environmental changes and support sustainable development.
- Dr. Denisa Rodila is a Scientific Collaborator in the Swiss Data Cube project and works on algorithms regarding integration of satellite and non-satellite sources, and on management of large-scale data sets. She is involved in the parallelization and performance improvement of the Swiss Data Cube project, while making use of High-Performance Computing technologies.

- Bruno Chatenoux is a civil-engineer and geologist specialized in GIS in environmental sciences. He has been involved in various projects and also worked as a consultant for the Small Arms Survey (SHBA project), and as a GIS advisor for the French Red Cross. He is currently the Data Analyst of the Swiss Data Cube (http://www.swissdatacube.ch)

Université de Bourgogne - Burgundy Computer Science Laboratory (LIB) is a research laboratory developing and structuring research on three major topics: Geometric Modelling, Combinatorics and networks, Data Science. This partner institution, made up of members of the Data Science research team, brings its experience in knowledge modelling, knowledge enrichment, knowledge evolution, spatiotemporal and trajectories modelling, dynamics ontologies, large-scale data storage solution.

- Christophe Cruz is a Full-Professor in Computer Science. His research interests are knowledge modelling, knowledge enrichment and spatiotemporal trajectory modelling. He developed expertise and research projects on agent knowledge modelling and spatiotemporal trajectory insight.
- Hocine Cherifi is a Full-Professor in Computer Science. His primary work and research interests are in the areas of Hypergraph and Network Science.
- Eric Leclerq is an Associate Professor in Computer Science. His research work concerns the definition of models to link data, knowledge and analysis tools, and multi-paradigm models and polyglot storage system (or polystore) to deal with the variability of large-scale data.

II.a.3 Complementarity of the consortium

The partners form a multidisciplinary consortium of researchers confirmed in Computer Science in the three fields of AI addressed (Knowledge Representation, Machine Learning, Multi-Agent System), and in geography and geomatics, in the processing of geographical information, particularly in the production and management of statistical and environmental data. The consortium brings together a great deal of expertise in various technologies (semantic web, machine learning, multi-agent systems, environmental databases, use and analysis of satellite images, etc.) which will be used to build the proposed processing chain. Some of these researchers have already worked together in national or European projects or for PhD supervision on related subjects, but all of them are gathered for the first time around this ambitious project. The partners have to their credit numerous robust and widely used IT projects. Although mainly made up of computer scientists, the consortium has a good knowledge of environmental issues and practices or policies at different territorial levels (municipalities, metropolitan areas, cantons, departments, State, Europe). Finally, the geographical proximity of the partners appears to be an asset, and motivates in particular the choice of environmental trajectories allowing comparative analysis to be carried out on the border regions between France and Switzerland.

II.a.4 Implication of the scientific coordinator and partner's scientific leader in on-going projects

Name of the researcher	Person. month	Call, funding agency, grant allocated	Project's title	Name of the scientific coordinator	Start - End
Jérôme Gensel	4p.month	ANR PRC Défi 7	Choucas	Ana-Maria Olteanu-Raimond	01/01/2016-15/04/2022
Hy Dao	4.2p.month	ESPON	ESPON 2020 Database Portal	Hy Dao	12/04/2021-12/07/2022
Giovanna Di Marzo Serugendo	1.2p.month	Innosuisse	Addmin	Stéphane Journot	01/07/2021-31/06/2022
Christophe Cruz	2p.month	I-SITE UBFC	HERMES	Frédéric Demoly	01/09/2018-31/08/2021

II.b Implemented and requested resources to reach the objectives

II.b.1 Partner 1: LIG

A 36p.month funding is requested for a Ph.D. thesis, co-directed by J. Gensel (LIG) and G. Giuliani (ISE). The aim is to propose a generic ontological trajectory model for representing the any thematic evolution of any territory over a given period. A specific instance of this model will be dedicated to environmental trajectories developed by T1 and T2. The priority will be to use standard RDF vocabularies and, more particularly, RDF Data Cube to link indicators, time and territories. Once this ontological model has been designed and implemented as KGs, adapted similarity measures will be evaluated, as key elements of clustering-type classification processes or frequent pattern searches, such as those addressed in task T4 for the particular case of environmental trajectories.

A 12p.month funding is requested to propose a postdoc position to address challenges identified in Activity A2.3 Visualizing semantic environmental trajectories.

18p.months are requested for 3 internships at Master level to tackle specific topics such as tools/technologies benchmarking or for a proof-of-concept development.

II.b.2 Partner 2: LIB

A 12p.month funding is requested to recruit an engineer to tackle the technical requirements of the task T3. The engineer will focus on data infrastructure for storing KGs to make available data online, to provide uplift capability, trajectory enrichment, expert view extraction (M7 to M24).

18p.month are requested for three internships at Master level to support the Ph.D. for the development of the tasks respectively T3.1, T3.2, and T3.3.

II.b.3 Partner 3: ISE

Staff expenses A 4.8p.month funding is requested for a Scientific assistant to develop set of T1 activities on Data, indicators and use cases preparation.

II.b.4 Partner 4: CUI

A 36p.month funding is requested for a Ph.D. thesis, co-directed by G. Di Marzo (CUI) and C. Cruz (LIB). The PhD aims at producing knowledge-graph driven agent-based models of environmental trajectories, including a prescriptive analysis of various scenarios (T5). The PhD activity spans T3, T4, and T5 as the PhD student will also participate to knowledge enrichment and expert view extraction (T3), and on similarity between trajectories and trajectories encoding profiling (T4).

A 7.2p.month funding is requested for a postdoc position to develop set of T4 activities on Similarities, Patterns & Clustering Approaches.

III Impact and benefits of the project

The modelling and analysis of environmental trajectories expressed and enriched in the form of KGs is the overall contribution expected from the TRACES project. In the field of geographic information and the Semantic Web, this contribution may be of interest to working groups within the standardization bodies that are the OGC (Open Geospatial Consortium) and the W3C. In the field of Linked Open Data, the RDF vocabulary associated with the ontological model produced will be referenced on the LOV site, ready to be used as a reference vocabulary for trajectories of territories.

The proposed processing chain (from modelling to prediction, via visualization) and the associated methodology will have to be disseminated towards the Human and Social Sciences, in particular geography, ecology, town and country planning, economics, etc., which are increasingly interested in trajectories of territories as objects of study.

Communities/cities face many challenges, among other the ecological transition, climate change or economic issues. In view of reaching their targets in these areas, municipalities will benefit from the TRACES project results in order to measure and understand the impact of decisions, be they related to new multi-modal possibilities for their citizens, building new public transport facilities, or urban gardens, etc. The economic impact of the project will not be immediate as TRACES will not deliver a complete and mature software package that assists the user in the modelling and in the various analysis proposed. However, we expect that the software programs and prototypes developed in the TRACES project will open the way to a more robust and integrated tool that facilitates the implementation of each stage of the processing chain. This software could then serve as a decision-making tool for professionals working in private or public organizations that produce or use statistical data on territories

In terms of dissemination to the general public, first of all, the choice to place the SETTS' KGs in the Linked Open Data Cloud is in line with the Open Data and INSPIRE directives to which institutions in Europe are subject. The enrichment of KGs is also an essential step in the dissemination to citizens of expert data associated with metadata, allowing these citizens to understand the meaning of such data, thus facilitating their participation in public debate and participatory democracy initiatives.

CUI developed since 2019 a reach out activity for primary and secondary schools of the Canton Geneva, called the Infoscope. It proposes workshops for school teachers and their classes on well identified computer science topics. In the framework of the TRACES project, the Infoscope could develop a new activity at the crossroad of AI and environmental studies. The University of Geneva in general participate every two years to a large public event on two days and nights, aiming at providing research results to the general public through ludic and interactive activities. Visualization results of the TRACES project will fit perfectly this type of activities. The intermediate and final results of the TRACES project will be presented at scientific, professional or general public events (scientific popularization) such as the Geodatadays[7], an annual national reference and independent event for digital geography in France, under the aegis of Afigéo (French association for geographic information), and mainly aimed at local authorities. This will contribute to make the results of the TRACES project better known to French professionals' networks. Others events involving geographic information and environment professionals can be envisioned as (for instance Geosummit, or the geo-information day, geonight, or hackathon[8]). Publications in independent and specialized magazines or media (for instance, The Conversation, DécryptaGéo[9])) will also be targeted.

The work and results of the TRACES project will be published in high-ranking international conferences (ACM SIGSPATIAL, ISWC, ICML, AAMAS, SIMUL...) and journals (Journal of Semantic Web, IJDE, IJGIS, IJDAEOG, Pattern Recognition, ACM TAAS, Land...) in the three AI domains addressed and in the field of geographic information, with a particular emphasis on open access journals or books. These publications will be listed and deposited in the HAL open archive. Workshops will be organized to bring the project's themes and results onto the international scene and initiate new collaborations.

A transfer to the world of industry and software could be envisaged, for example by participating at the end of the project in the Out of Labs Challenge launched each year by the SATT Linksium[10], which offers to support researchers in the valorisation of their results, from maturation to start-up creation.

The TRACES consortium is made up of research teams with complementary expertise and skills, some of which already have fruitful multidisciplinary collaborations with each other, either in funded scientific projects or in the supervision of theses in the field of computer processing of geographical or environmental data. This mutual knowledge makes the consortium confident in its ability to work together. The data that will be used by the project are data made available by the national mapping agencies swisstopo for Switzerland and IGN for France. This is why the focus will be on the evolution of the Swiss and French territories, with a particular focus on the border territories, which we believe is justified by the PRCI call and constitutes a contribution for each of the countries independently - Switzerland and France - but also for the neighbourhood relations and the common actions and policies that these two border countries carry out or could carry out in the future. It should be noted, however, that the results of the TRACES project can be easily generalized to other territories, particularly in Europe, where similar indicator data are available. The scope of the TRACES project's work therefore goes far beyond the Franco-Swiss case studies chosen. Similarly, the generality of the model and the proposed algorithms will be of interest to the scientific communities in the fields covered on an international level.

---

[7] https://www.geodatadays.fr/
[8] https://www.geosummit.ch/- https://georomandie.com/ - https://www.geonight.net/ - e.g. https://opengeneva.org/
[9] https://theconversation.com- https://decryptageo.fr/
[10] http://outoflabs.linksium.fr/

IV References related to the project

[Alvares 2007] L. O. Alvares *et al.* Towards semantic trajectory knowledge discovery. Data Mining and Knowledge Discovery, 2007.

[Anquetin 2018] S. Anquetin *et al.* TRAJECTORIES: Social-ecological trajectories of French alpine valleys under climate variability", EGU2018, European GeoSciences Union, Vienna, 2018.

[Antoniazi 2018] F. Antoniazzi *et al.* RDF Graph Visualization Tools: a Survey. 2018. 23rd Conference of Open Innovations Association (FRUCT), pp. 25-36, 2018, doi: 10.23919/FRUCT.2018.8588069.

[Baltrušaitis 2019] T. Baltrušaitis *et al.* Multimodal Machine Learning: A Survey and Taxonomy. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423-443, 2019, doi: 10.1109/TPAMI.2018.2798607.

[Bernard 2018] C. Bernard *et al.* Modeling changes in territorial partitions over time: Ontologies TSN and TSN-Change. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing, SAC'18, pp. 866–875. ACM, 2018, doi: 10.1145/3167132 .3167227.

[Berners-Lee 2009] https://www.w3.org/DesignIssues/LinkedData.html

[Bryan 2016] B. A. Bryan *et al.* Land-use and sustainability under intersecting global change and domestic policy scenarios: Trajectories for Australia to 2050, Global Environmental Change, Volume 38, pp. 130-152, 2016, https://doi.org/10.1016/j.gloenvcha.2016.03.002.

[Cyganiak 2014] R. Cyganiak and D. Reynolds. The RDF Data Cube Vocabulary, 2014, https://www.w3.org/TR/vocab-data-cube/.

[Dao 2012] H. Dao *et al.* Indicators of territorial cohesion. Final Report. ESPON. 2012. https://www.espon.eu/programme/projects/espon-2013/scientific-platform/interco-indicators-territorial-cohesion

[De Donato 2020] R. De Donato *et al.* QueDI: From Knowledge Graph Querying to Data Visualization. In: Blomqvist E. et al. (eds) Semantic Systems. In the Era of Knowledge Graphs. SEMANTICS 2020. LNCS vol 12378, 2020.

[Démare 2014] T. Démare *et al.* Agent-based model and networks : the modelling of the maritime and metropolitan interfaces of the Seine axis. Conference: Devport international conference, 2014.

[Federov 2017] I. Fedorov *et al.* Multimodal sparse Bayesian dictionary learning applied to multimodal data classification. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, pp. 2237-2241, 2017, doi: 10.1109/ICASSP.2017.7952554.

[Fua 2020] L. Fua *et al.* An overview of recent multi-view clustering Neurocomputing. Vol 402, 2020, pp. 148-161, 2020, https://doi.org/10.1016/j.neucom.2020.02.104

[Giuliani 2020a] G. Giuliani *et al.* Knowledge generation using satellite Earth Observations to support Sustainable Development Goals (SDG): a use case on Land Degradation, International Journal of Applied Earth Observation and Geoinformation, 2020, doi 10.1016/j.jag.2020.102068

[Giuliani 2020b] G. Giuliani *et al.* Essential Variables for environmental monitoring: What are the possible contributions of Earth Observation Data Cubes, Data 5(4):100, 2020, doI: 10.3390/data5040100

[Harbelot 2013] B. Harbelot *et al*. Continuum: A spatiotemporal data model to represent and qualify filiation relationships. In Proceedings of the 4th ACM SIGSPATIAL International Workshop on GeoStreaming, pp. 76–85. ACM, 2013.

[Heckel 2017] R. Heckel *et al*. Features of Agent-based Models. In Proceedings Third Workshop on Graphs as Models, GaM@ETAPS Uppsala, Sweden, EPTCS 263, pp. 31-37, 2017, https://arxiv.org/abs/1712.09496v1

[Hogan 2020] A. Hogan *et al.* Knowledge Graphs. CoRR 2020, abs/2003.02320, [2003.02320].

[Huang 2019] C. Huang *et al*. MiST: A Multiview and Multimodal Spatial-Temporal Learning Framework for Citywide Abnormal Event Forecasting. In The World Wide Web Conference (WWW '19), 2019. https://doi.org/10.1145/3308558.3313730

[Le 2008] Q Le *et al.* Land-Use Dynamic Simulator (LUDAS): A multi-agent system model for simulating spatio-temporal dynamics of coupled human–landscape system. I. Structure and theoretical specification, Ecol. Informatics 3(2), pp. 135-153, 2008.

[Lehmann 2010] A. Lehmann *et al.* Swiss Environmental Domains : A new spatial framework for reporting on the environment. Federal Office for the Environment (FOEN). 2010. https://archive-ouverte.unige.ch/unige:32451

[MameLuke 2001] The MameLuke Project. Spatially Explicit Multi-Agent Modeling of Land-Use Change in the Sierra, Madre, Philippines. In Agent-Based Models of Land-Use and Land-Cover Change International Workshop. Irvine, CA, USA. Edited by: D. C. Parker *et al.*, LUCC Report Series No. 6, 2001, https://www.researchgate.net/publication/228871179

[Millington 2016] J. D.A Millington *et al.* (Guest Editors). Special Issue "Agent-Based Modelling and Landscape Change", Land 2016.

[Nativi 2020] S. Nativi *et al.* Towards a knowledge base to support global change policy goals, International Journal of Digital Earth, 13:2, 188-216. 2020. DOI: 10.1080/17538947.2018.1559367

[Nentwig 2017] M. Nentwig *et al.* A survey of current Link Discovery frameworks, Semantic Web, 8, pp. 419–436. 2017. doi:10.3233/SW-150210.

[Noël 2017] D. Noël *et al.* Design patterns for modelling life trajectories in the semantic web. In International Symposium on Web and Wireless Geographical Information Systems, pp.51-65. Springer, 2017.

[Parent 2013] C. Parent *et al*. Semantic trajectories modeling and analysis. ACM Computing Surveys, vol. 45, no. 4, pp. 1-42, 2013.
[Prudhomme 2020] C. Prudhomme *et al.* Interpretation and automatic integration of geospatial data into the Semantic Web. 2020. Computing 102, 365–391. https://doi.org/10.1007/s00607-019-00701-y
[Paulheim 2017] H. Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web 8(3), pp. 489-508, 2017.
[Welty 2006] C. Welty *et al.* A reusable ontology for fluents in OWL. In FOIS, volume 150, pp. 226–236, 2006.
[Wilkinson 2016] M. D. Wilkinson *et al.* The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 3: 160018, 2016, doi:10.1038/sdata.2016.18. OCLC 961158301. PMC 4792175. PMID 26978244.