MASTER

Multiple ASpect TrajEctoRy management and analysis

Project Acronym	***MASTER***
Project Full Name	***M**ultiple **AS**pects **T**raj**E**cto**R**y management and analysis*
Project Number	***777695***
Deliverable Title	***Revision of the Applications Scenarios***
Deliverable No.	***D5.2***
Contract Delivery Date	**29/02/2020 (M24)**
Actual Delivery Date	**29/02/2020(M24)**
Responsible Authors	*Nikos Pelekis (Unit Manager of UPRC)*
	This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie-Sklodowska Curie grant No 777695

DOCUMENT INFORMATION

GRANT AGREEMENT N.	777695
PROJECT ACRONYM	MASTER
PROJECT FULL NAME	Multiple Aspects Trajectory Management and Analysis
STARTING DATE (DUR.)	01/03/2018 (48 months)
ENDING DATE	28/2/2022
PROJECT WEBSITE	http://www.master-project-h2020.eu
COORDINATOR	Chiara Renso (CNR)
WORKPACKAGE N. | TITLE | START – END MONTH	WP5 | Application Scenarios | M1 - M48
WORKPACKAGE LEADER	UPRC
DELIVERABLE N. | TITLE	D5.2 | Revision of the applications scenarios
RESPONSIBLE AUTHOR	Nikos Pelekis (Unit Manager of UPRC)
DATE OF DELIVERY (CONTRACTUAL)	29 February 2020 (M24)
DATE OF DELIVERY (SUBMITTED)	29 February 2020 (M24)
VERSION | STATUS	1.0
NATURE	REPORT
DISSEMINATION LEVEL	PUBLIC
AUTHORS (PARTNER)	UPRC, CNR, UNIVE, HUA, DAL

CONTRIBUTORS	Nikos Pelekis (UPRC), Vinicius Cesar Monteiro de Lira (CNR), Andrea Marin (UNIVE), Salvatore Orlando (UNIVE), Alessandra Raffaetà (UNIVE), Chiara Renso (CNR), Konstantinos Tserpes (HUA), Michela Rial (CNR), Daniele Carpita (CNR)

ACRONYM LIST

MASTER	Multiple Aspects Trajectory Management and Analysis
ICT	Information and Communication Technologies
ISTI	Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
CNR	Consiglio Nazionale delle Ricerche
UNIVE	Ca' Foscari University of Venice
UVSQ	University of Versailles Saint-Quentin
UPRC	University of Piraeus Research Center
UFC	Federal University of Ceará
HUA	Harokopio University of Athens
PUC	Pontificial University of Rio de Janeiro
DAL	Dalhousie University
THIRA	Municipality of Thira
ER	Experienced Researcher
ESR	Early Stage Researcher
ACTV	Azienda del Consorzio Trasporti Veneziano (Company responsible for public transportation in province of Venice)
AIS	Automatic Identification System (AIS)

TABLE OF CONTENTS

1. INTRODUCTION

The objective of this deliverable, that is linked to WP5, is "to *revise the requirements, the application questions and the information about the data available for the three scenarios*" as stated in the MASTER Grant Agreement Annex 1 Part A pages 18-19. This activity has been carried out during secondments linked to WP5 and parallel activities of project partners. We have executed 2.7 person months during which secondees have refined requirements and information about the datasets available at the partners relative to the three application scenarios: tourism, sea monitoring and transportation. The results are reported in this document.

The work done in preparing this deliverable is mainly connected to the secondments linked to WP5 "Application Scenarios" for the requirement analysis and datasets.

WP5 has the following four objectives:

1. **to understand the application needs and perform the requirements analysis driving the design of the methods developed in WP3 and WP4**. This objective has been mainly reached, during the second year, by secondments to PUC, DAL that confirmed the requirements exposed during the first year and focused on methods closer related to application prototypes. This objective has been fully achieved relatively to the second year;

2. **to host the interaction between academic and non-academic partners to create awareness and best practices on the non-academic world needs, thus increasing the potential in career developing especially in ESRs**. This objective was reached, during the first year, with secondments to Thira. This objective has been fully achieved relatively to the second year with remote interactions with Thira for the application questions;

3. **to test the developed techniques in real-world scenarios possibly on data owned by the non-academic partner and on the basis of their actual daily problems**. This objective has been reached, during the second year, from the secondments to DAL. This objective has been fully achieved relatively to the second year;

4. **to develop software prototypes to facilitate the interaction and transfer of knowledge among academic and non-academic partners.** This objective is to be reached by developing the research prototypes to be reported in D5.3 and D5.4 due at M36 and M48, respectively. This objective has not been achieved since it was planned for M36.

The past deliverable D5.1 and the present deliverable D5.2 fulfill the first, second and third objectives with reference to the progress to the goals of **WP5**. The deliverable is a revision of the already submitted and approved deliverable D5.1 (M12-**MS1**). The current deliverable will revise the application scenarios w.r.t. the goals of MS3. Indeed, deliverable D5.2 identifies those requirements and datasets that better fit for the research prototypes that will be the objective of D5.3 (M36-**MS6**) and D5.4 (M48-**MS7**). These two deliverables will provide the preliminary and the final reports on application scenarios and software prototypes, respectively.

The research activities carried out during the first 24 months of the project for these tasks have produced 1 publication reported in the Publications section. This publication joins together activities concerning two different WPs: Task 3.2 and Task 5.2. In fact, we built a set of holistic trajectories by exploiting the model developed in Task 3.2 and we used the dataset from the Sea Monitoring scenario and we coped with some application questions of such a scenario.

At M24, the number of secondments linked to WP5 is 3, as listed below in Table 1. According to the Researcher Declarations submitted to SyGMa system, the total number of person months is 2.7.

One researcher from CNR has been seconded to PUC, essentially working on Task 5.3 on transportation. Two researchers from UNIVE have been seconded at DAL (working on Task 5.2 on sea monitoring) and at PUC (working on Task 5.3 on transportation scenario), respectively.

Objectives of WP5 that have been fully achieved by this deliverable, relatively to the second year of the project, are the first, the second and the third ones listed above.

The outcome of this deliverable is to feed Deliverable D5.3 "Preliminary Software prototypes" due at M36.

Table 1: Secondments executed from M13 to M24 linked to WP5 Application scenarios from the SyGMa system

RD N.	Second. N	Secondee Name	Fellow ID	Sending Institution	Hosting Institution	From	To	PM	Task
18	6	Vinicius Cesar Monteiro de Lira	7	CNR	PUC	14/03/2019	30/03/2019	0.57	T5.3
28	29	Andrea Marin	10	UNIVE	PUC	27-06-2019	27-07-2019	1.03	T5.3
29	28	Salvatore Orlando	21	UNIVE	DAL	19-09-2019	21-10-2019	1.1	T5.2

2. TOURISM SCENARIO

This chapter of the deliverable is related to the activity of Task 5.1. about Tourism application scenario, whose responsible is CNR. The material for this deliverable has been prepared during research activity and meetings performed at the institutions, as **no secondment** has been linked to this task during the second year of the project. Research activity has focused on (1) studying the datasets that have been collected and described in deliverable D5.1 to better understand the actual usefulness and practical usage of the datasets for the collected requirements, focusing on the prototype preparation planned as D5.3 at M36; (2) studying the analysis questions reported in D5.1 page 11, focusing on the prototype; (3) revising and selecting the most suitable datasets and the application questions that will be realized in the prototype.

APPLICATION QUESTIONS

As reported in the Deliverable D5.1, the main question that Thira Municipality wants to answer is: *How can we monitor the tourism flow and support the decision-making process*?
We studied the proposed questions after having explored the available datasets and forecasted a plan for the prototype.
We therefore report here the questions already introduced in D5.1 where we comment more about our findings and the possible datasets that could contribute in answering such questions.

a) Which is the origin Country of travelers?

In order to answer this question, we need to analyse one or more datasets where the origin country of tourists is reported or can be reasonably inferred. We identified Twitter data as the most promising dataset to answer this question. As already reported in Deliverable D5.1, we have been collecting data from the Twitter API since May 8th and this collection is still ongoing. The collection has tweets containing the keyword "Santorini" and any geo-tagged tweets posted in Santorini. To address the current question, we look specifically to the information of residence which is present in the users' profiles (i.e. for the public users' accounts) and the languages which the tweets are written.

b) How long do the tourists stay in Santorini?

We also plan to use the Twitter dataset to compute an approximation of how long the tourists stay in Santorini. To answer this question, we investigate the dates that the users have posted on Twitter while visiting the island specifically looking at the geo-tagged posts. The difference in days between the consecutive posts of a user in Santorini can give an estimation on how long the person stayed in the island. This is a lower bound approach commonly used in literature (see for example [1]) to estimate a minimum period of time that a user stayed in a location.

[1] Igo Brilhante, Jose Antonio Macedo, Franco Maria Nardini, Raffaele Perego, and Chiara Renso. Where shall we go today? Planning touristic tours with TripBuilder. In Proceedings of CIKM'13, pages 757–762, 2013.

c) Are the first-time visitors and returning visitors behaving differently?

To answer this question, we need a dataset with user-level information. Our tweets collection can be used for this purpose. The idea is to first infer the first-time visitors and returning visitors. Afterwards, we want to investigate the behaviour of these two groups of users by analyzing their trajectories in Santorini. Also, the OpenStreet map dataset can help our analysis by providing more detailed information about the locations that the users have visited.

d) **Can we predict the tourist arrival flows**?

To investigate the tourist arrival flow we plan to use the datasets below. Together, these datasets provide complementary information to analyse and predict the tourist arrival flows in Santorini.

a) Flights arrivals and departures.
b) Water consumption.
c) Ships arrivals and Departures.
d) Cruise Ships Berth Allocation.

e) **Which is the qualitative level of satisfaction / dissatisfaction of the user?**

One popular way to analyze the users' satisfaction (i.e. sentiment) respect to a subject is through the use of social media. For this reason, we plan to use our collection of tweets to discover how people feel about Santorini, particularly analyzing the textual content of the posts.

DATA

Regarding the datasets, since we did not execute secondments during the second year, we did not have variations in the datasets descriptions already reported in D5.1. In the current deliverable, we report only the datasets which we plan to use for answering the application questions, discussed in the previous section that will feed the prototype. The datasets (already introduced with details in D5.1) are:

- **Flights arrivals and departures**
- **Ships arrivals and departures**
- **Cruise Ships Berth Allocation**
- **Water Consumption**
- **OpenStreetMap**
- **Twitter streaming**

We plan NOT to use the following datasets as they probably cannot contribute to answer the application questions or they do not provide enough information to be useful:

- **AIS Data (not useful for the selected questions)**
- **Wifi (not enough information)**
- **Flickr data (too old and too small dataset)**

Other datasets that we mentioned in D5.1 page 14 as possible additional data (e.g. hospital) that could be collected have not been provided by Thira so far. However, we believe that they are not necessary to answer the application questions, therefore, at the moment, we plan not to use them. In case they will be available in the future we will try to include them in our analysis questions.

3. SEA MONITORING SCENARIO

This section of the deliverable is related to the activity of Task 5.2, namely the Sea Monitoring application Scenario, whose responsibility is UPRC. The material of this part of the deliverable is mainly prepared during secondment to DAL, but also integrated during parallel research activities of project partners. The activities have focused on (1) studying the datasets that have been collected and described in deliverable D5.1 to better understand the actual usefulness and practical usage of the datasets for the collected requirements, focusing on the prototype preparation planned as D5.3 at M36; (2) studying the analysis questions reported in D5.1, focusing on the prototype; (3) revising and selecting the most suitable datasets and the application questions that will be realized in the prototype.

The Secondment linked to the Tasks 5.2 is reported in Table 1 at the Introduction section of this document and is the following: UNIVE (Salvatore Orlando) at DAL for a total of 1,1 PM.

APPLICATION QUESTIONS

Following the approach described above, during the second year of the project, we focused our activities on three categories of application requirements, already highlighted in deliverable D5.1.

More specifically, on the category of **"Data-driven extraction and classification of maritime patterns of life"** and as the feedback from DAL partner was that the "predictive analytics" if of vital importance in the maritime domain, we focused our work on short-term ship prediction that provides track estimation on a near-term scale from the partially observed track.

The problem of predictive analytics over mobility data finds two broad categories of application scenarios. The first scenario involves cases where the moving entities are traced in real-time to produce analytics and compute short-term predictions, which are time-critical and need immediate response. The prediction includes either location- or trajectory-related tasks. Short-term location and trajectory prediction facilitates the efficient planning, management, and control procedures while assessing traffic conditions in the sea transportation field. The latter can be extremely important in domains where safety, credibility and cost are critical and a decision should be made by considering adversarial to the environment conditions to act immediately. The second scenario involves cases where long-term predictions are important to identify cases which exceed regular mobility patterns, detect outliers and determine a position or a sequence of positions at a given time interval in the future. In this case, although response time is not a critical factor, it is still crucial in order to identify correlations between historical mobility patterns and patterns which are expected to appear. Long-term location and trajectory prediction can assist to achieve cost efficiency or it can ensure public safety.

As the maritime domain has major impact to the global economy, a constant need is to advance the capability of systems to improve safety and effectiveness of critical operations involving a large number of moving entities in large geographical areas. Towards this goal, the exploitation of surveillance data sources, which offer vast quantities of archival and high-rate streaming data, is crucial for increasing the computations accuracy when analysing and predicting future states of moving entities. However, operational systems in the maritime domain for predicting trajectories are still limited mostly to a short-term look-ahead time frame, while facing increased uncertainty and lack of accuracy.

Motivated by these challenges, we focus our requirements (thus the scope of the envisioned prototype) on a Big data solution for online trajectory prediction by exploiting mined patterns of trajectories from historical data sources. Such an approach would offer predictions such as `predicted route of a vessel in the next hour', based on their current movement and historical motion patterns in the area. The envisioned prototype should incorporate several innovative modules, operating in streaming mode over surveillance data, to deliver accurate long-term predictions with low latency requirements.

In addition, on the category of **"Data-driven extraction and classification of maritime anomalies"** we focus our requirements towards a stream-based processing framework that can provide online summarized representations of trajectories specifically for sailing vessels. Assuming again that surveillance data monitoring their locations over a large geographical area is available in a streaming fashion, we seek for methodologies that drop any predictable positions (along trajectory segments of "normal" motion characteristics) with minimal loss in accuracy. Effectively, such a method should keep only those positions conveying salient mobility events (annotated as stop, change in speed, heading or altitude, etc.), identified when the mobility pattern of a given vessel changes significantly. Thus, such an application could employ parametrized conditions for detecting such mobility features, as well as eliminating inherent noise.

Finally, on the category of **"Improving the knowledge of the fishing activities in the Northern and Central Adriatic Sea",** during this year a publication [1] has been prepared and published at the first MASTER Workshop. This is a joint work between DAL and UNIVE resulted from the collaboration started during the secondments at DAL in 2018. The research activity focused on the creation of a set of holistic trajectories obtained by merging three data sources: trajectories from fishing vessels (obtained from terrestrial AIS data), the corresponding fish catch reports (i.e., the quantity and type of fish caught), and relevant environmental data. Then some exploratory analyses of these holistic trajectories have been investigated and an initial predictive modeling using Machine Learning has been designed to predict the Catch Per Unit Effort (CPUE), an indicator of the fishing resources exploitation useful for fisheries management. It is worth noticing that this publication reports research results from both WP5 (sea monitoring scenario) in predicting the fishing effort and WP3 since authors used a semantic enrichment method to create holistic trajectories developed as activity in T3.2 and reported in D3.1.

Moreover, a secondment has been carried out. During Secondment #28, Salvatore Orlando (UNIVE), seconded at DAL, studied how to further use the created set of holistic trajectories, with also features of the vessel databases, in order to:

1. Exploit data mining techniques to identify recurrent and seasonal fishing patterns;
2. Transform trajectory data in a temporal series of graphs, both unipartite and bipartite, where nodes are fishing regions and/or vessels. The aim is analyzing the temporal evolution of communities (or bi-communities), as well as further measures concerning graphs, and making prediction by exploiting the time series knowledge.

   The plan is to exploit season and other information to filter data and produce different graphs.

It is worth noting that our predictive results are preliminary in both the temporal data horizon that we are able to explore and in the limited set of learning techniques that are employed on this task. In the future, and towards designing a related software prototype for M36, we plan to apply learning methods from such data so, as to discover evidences and knowledge that will be useful for fisheries management.

DATA

Regarding the datasets, during the current year we did not have variations in the datasets descriptions already reported in D5.1. In the current deliverable, we report only the datasets which we plan to use for answering the application questions, that will feed the prototype.

- **AIS Surveillance data from Eastern Mediterranean**
- **AIS Dataset for the Northern and central Adriatic Sea (GSA17)**
- **Fishing gears**
- **Landing data**

Among the meteorological datasets, we used:

- Copernicus is the European Union's Earth Observation Programme offering information services based on satellite Earth Observation and in situ (non-space) data. https://www.copernicus.eu/en.
  In particular, from this repository we consider (a) Sea Surface Temperature (in kelvin), (b) Sea Daily Chlorophyll-a Concentration (in mg/m^3) and (c) Spectral significant wave height (in meters).

Preliminary efforts towards answering the application questions have shown that we are able to provide promising solutions without the usage of other meteorological data than the above-mentioned Copernicus data. Nevertheless, we will continue to investigate whether the other listed datasets in Deliverable D5.1 (namely climate data from *https://www.noaa.gov/,* sea state conditions from *http://www.ionioproject.eu/ and http://www.sea-conditions.com/en/*, Mediterranean Sea Physics Analysis And Forecast Data from http://www.myocean.eu and Contextual Information from earth.google.com/gallery/kmz/marine_protected_areas.kmz) can improve our predictive models.

4. TRANSPORTATION SCENARIO

This Chapter of the deliverable is related to the activity of collecting requirements and datasets relative of Task 5.3, namely the Transportation Scenario, whose responsible is UNIVE. The material of this part of the deliverable is mainly prepared during secondments to PUC, but also integrated during parallel activities of the project partner UNIVE.

The Secondments linked to Task 5.3 are reported in Table 1 in the Introduction section of this document and are the following: UNIVE (Andrea Marin) at PUC and CNR (Vinícius Monteiro de Lira) at PUC for a total of 1,6 PM.

INTRODUCTION

From GA Annex 1 Part A: Task 5.3. This task will study methods for: creating a public transportation observatory for buses data; improving traffic prediction; improving ride sharing methods to reduce private vehicle usage. We will exploit the data from PUC and UNIVE and datasets collected from social media to develop a prototype application.

During secondments at PUC, secondees have looked in depth and discussed with PUC researchers about the datasets presented in D5.1 namely BIKE-RIO and BUSES-RIO and how these datasets can be useful for reaching task objectives and building the prototype planned at M36.

During Secondment # 29, Andrea Marin (UNIVE), seconded at PUC, has first compared the datasets available at PUC with the datasets available at UNIVE and it turned out that these datasets are somehow complementary. For example, it has been observed that the PUC datasets express travel times and events while the UNIVE datasets are better to express the people flows in the different areas of the city. Another observation is that bus data are not good predictor for traffic intensity due to traffic lights and stops. Furthermore, it has been observed that the available datasets are quite obsolete with respect to the changes in the transport infrastructure due to Olympic Games in 2016 and this should be considered for slightly revising the use of these datasets.

One proposal is to exploit the complementarities of the PUC and UNIVE datasets, for example in testing the trajectory reconstruction algorithm that UNIVE is studying in the PUC datasets. This can be done by adding noise to the BUSES-RIO dataset and hence by assessing the accuracy of the algorithm in reconstructing the original dataset.
During secondment # 6 Vinicius Monteiro de Lira (CNR), seconded at PUC, has studied the possible uses of the datasets of PUC. In consideration of the outdated data available, the proposal could be to use these datasets to understand how well the lines of buses could cover the different functions areas (e.g. tourism, shopping, and hospital areas) of the city of Rio de Janeiro.
These two proposals could be functionalities to support a transportation observatory.

PUBLIC TRANSPORTATION IN VENICE (ITALY)

UNIVE is collaborating with ACTV, the main public transport company, in order to study the flows and to understand the behavior of its users.

During this year we revised, cleaned and filtered the available data. ACTV provides two different GTFS (**General Transit Feed Specification)** data: one about the navigation service and the other one about the bus transport. The number of stops is large and this makes it difficult to find meaningful behavior patterns of users. Hence, we reduced this number by merging together stops. We used different criteria to aggregate stops. Since our focus is on navigation service and the number of bus validations is small with respect to waterbus ones, for the bus transport, we built five big areas:

- Lido,
- Airport
- Piazzale Roma
- Mestre Train Station
- Mestre (urban area)

These areas correspond to different functional areas in the mainland and Lido is an island where there is a bus network.

The stops of the navigation service are 144 but a lot of them correspond to the same landing stage, so we aggregated them into a unique stop like in the following example: S. Zaccaria (Pieta') "A", S. Zaccaria (M.V.E.), "B", S. Zaccaria (Jolanda) "C", S. Zaccaria (Jolanda) "D", S. Zaccaria (Danieli) "E", S. Zaccaria (Danieli) "F" are mapped into the stop named "S. Zaccaria".

The number of navigation stops is reduced to 69.

After this aggregation process, from 2500 stops, we obtained a set containing only 74 stops.

The stamp datasets appear like big CSV files where each line represents a single validation of a specific user. Each validation includes the following pieces of information:

- Date and hour of the stamp
- Serial number of the user
- Code_Profile
- Code of the stop in which the ticket has been stamped

- Name of the stop.

As already highlighted in D5.1, for privacy reason, the Serial number of the user is anonymized, in a way we cannot disclosure the identity of the user.

From these data we built trajectories, by grouping the stamps associated with the same user. Hence a trajectory is a sequence of couples of this form: (*stop, time*), where *stop* is the identifier of the stop in which the user gets on the waterbus/bus and *time* is the moment when this action takes place. Notice that we replaced the original stops with the ones we created according to the aggregation process we have just described. For instance, in Figure 1, the red points are the different stops named from left to right S. Zaccaria (Pietà) "A", S. Zaccaria (M.V.E.), "B", S. Zaccaria (Jolanda) "C", S. Zaccaria (Jolanda) "D", S. Zaccaria (Danieli) "E", S. Zaccaria (Danieli) "F". They are very close: the distance between the first and the last stop is around 250 meters. After the aggregation process, all these stops are replaced with the name "S. Zaccaria" and in the visualization we plan to show the stop circled in yellow.

Figure 1. Example of the aggregation process. In yellow we show the location resulting from the aggregation process.

The sequence of couples is ordered with respect to the time component. Some cleaning operations are done to avoid useless stamps. For instance, when for a user there is several stamps having close time stamps, few minutes of difference, we keep only the last stamp.

Moreover, we use the code Profile of the ticket in order to select different typology of users. In particular, we consider:

1. One-day ticket
2. Two days ticket
3. Three days ticket
4. Seven days ticket
5. Monthly ticket
6. Yearly ticket

Based on the knowledge of ACTV we assume that tickets belonging to categories 1 to 4 (called *time limited tickets*) are typically bought by tourists whereas categories 5 and 6 are purchased by people living, working or studying in Venice. In this way, we built different datasets that can be used to answer our application questions. In particular, in order to extract patterns concerning tourists, we created four datasets, one for each category of time limited tickets. On the other hand, we built separate datasets for the monthly and yearly tickets in order to investigate habits about people living, studying and working in Venice. These datasets are our starting point for the analyses that we would like to cope with.

Concerning the application questions listed in D5.1 we proceeded as follows.

a) **How do the stamps of the users vary in space and time?**
Before applying any mining technique to extract patterns, it is fundamental to visualize the locations (i.e., the stops) of the stamps along the different time periods. We plan to use QGIS, an open source GIS, and its plug-in Time Manager, to produce animations showing the heatmaps of the stamps at certain time intervals. Time Manager for QGIS enables comfortable browsing through spatio-temporal data and its dock widget provides a time slider and a configuration dialog to manage spatio-temporal layers. These animations will be part of the preliminary prototype we are going to implement for the next year and some visual results will be shown in Deliverable D5.3.

b) **Is it possible to classify the users into categories, like workers, students, tourists?**
In order to answer this question, we are going to exploit the Profile of the ticket stored in the stamps. As we have already written the time limited tickets are usually bought by tourists. On the other hand, to distinguish between workers and students we plan to consider the hour of the stamps, the location and the length of the stay, obtained as difference between consecutive stamps of the same user. Unfortunately, we know where the user gets on the bus/waterbus but not where s/he gets off. The missing of the arrival stop makes this kind of analysis complex and challenging.

c) **Are there typical patterns in the movements of the users?**
To extract patterns, first we would like to apply different clustering techniques, such as K-Means, DBscan, Hierarchical cluster, considering only the spatial component of the trajectory. For each algorithm we would like to compare the results obtained by using several distance metrics, such as Jaccard, Edit, LCSS (longest common subsequence).
After that, we would like to take into consideration the temporal component, too. In order to detect meaningful clusters, we should partition the day in temporal slots. Then, we can further partition the most significant spatial clusters based on the temporal information.

d) **Are there different behaviors during the weekday and at the weekend?**
We can separate the data between weekday and weekend and apply the previous algorithms to compare and verify whether there are different patterns.

e) **Which are the common itineraries for tourists?**
This is the most difficult query because the quality of itineraries we can detect is related to the sparsity of our datasets. In fact, if for each tourist we have a few stamps, we would mine very short paths. As already remarked in point (b) the missing of the arrival stop makes it very complex to reconstruct where the person is going. We could investigate the use of probabilities to go from stop A at time t_1 to stop B at time t_2 crossing some Points of interest. To this aim it could be useful to create an estimate distance matrix of the walk time between POIs and bus/waterbus stops of ACTV.

f) **Emphasize the differences between workers and tourists with respect to the most used stops, the time period and the duration of the stay in the town.**
Once we mined the different patterns we can compare them with respect to these criteria.

g) Detect the behavior of tourists during the different days of their stay.
We can partition tourists based on the duration of their tickets: one day, two days, three days and seven days. Then, we plan to apply to each of these groups the technique we are developing and we will compare the mined behaviours.

DATA

TRANSPORTATION IN RIO DE JANEIRO

Regarding the datasets, during the current year, we did not have variations in the datasets descriptions already reported in Deliverable D5.1.

TRANSPORTATION IN VENICE

Regarding the datasets, during the current year, we did not have variations in the datasets descriptions already reported in Deliverable D5.1.

5. CONCLUSIONS

This deliverable reports "the *revision of the requirements, the application questions and the information about the data available for the three scenarios.*". As reported in the MASTER Grant Agreement Annex 1 Part A and B we have identified three application scenarios linked to three relative tasks: T5.1 tourism, T5.2 sea monitoring and T5.3 transportation. In this deliverable, we have reported the revision of the application requirements, questions and datasets collected so far for the three scenarios.

The content of the deliverable has been produced during secondments linked to WP5 and parallel activities of project partners. The total effort in PM for secondments linked to WP5 is 2,7. No PM is linked to T5.1 Tourism, 1,1 PM are linked to Task 5.2 Sea Monitoring, while 1,6 PM are linked to Task 5.3 Transportation Scenario. During the secondments of researchers from UNIVE at DAL they compared the AIS data of the Northern and Central Adriatic Sea obtained from ExactEarth with the datasets available at UNIVE. They discovered that the datasets at UNIVE consist of a much larger number of trajectories which are more accurate since they are obtained from Terrestrial AIS stations whereas ExactEarth provides Satellite AIS data. Moreover, in this part of the project, concerning the fishing activities in the Adriatic Sea, the expertise from the environmental scientists at UNIVE has been fundamental to identify meaningful analyses. In particular, they suggested to predict the Catch Per Unit Effort (CPUE), an indicator of the fishing resources exploitation useful for fishery management purposes, and they assessed the experimental results.

The current deliverable is therefore an input to D5.3 at M36 that in turn will be an input to D5.4 at M48. The deliverable D5.3 will report about the research prototypes to be developed starting from the requirements and the datasets collected that in turn will be an input to D5.4 due at M48. This last deliverable will describe the developed prototypes and the experiences conducted in the diverse application scenarios.

6. PUBLICATIONS

[1] Pedram Adibi, Fabio Pranovi, Alessandra Raffaetà, Elisabetta Russo, Claudio Silvestri, Marta Simeoni, Amílcar Soares, Stan Matwin. Predicting Fishing Effort and Catch Using Semantic Trajectories and Machine Learning. Proc. of MASTER 2019, Würzburg, Germany. This is publication N. 20 in SyGMA chapter in a book (workshop proceedings). Link to repository https://zenodo.org/record/3678159#.XlO41i2ZNUM