Track & Know | Issue 4 | Date: October 2019

WELCOME TO THE FOURTH TRACK & KNOW NEWSLETTER!

In this newsletter, you can find:

- Updates about recent developments in the project
- Updates on recent dissemination activities

ABOUT THE NEWSLETTER

This newsletter provides an update on the results and activities of the EU H2020 research project Track & Know. The aim is to ensure that those interested in managing big data are kept up to date with our latest developments, specifically mobility data and the tools/methods we develop to handle, analyse and visualize these datasets. Track & Know aims to answer industry generated questions in 3 test pilots in the transport/mobility, insurance and health care sectors. The business cases explored in these pilots centre on answering questions around minimizing patients travel, carpooling and electric mobility potential and driver behaviour profiling.

Project coordinator
Dr. Ibad Kureshi
Senior Research Scientist – Inlecom Systems
Square de Meeus 38/40
1000 Brussels
Belgium
Ibad.kureshi@inlecomsystems.com

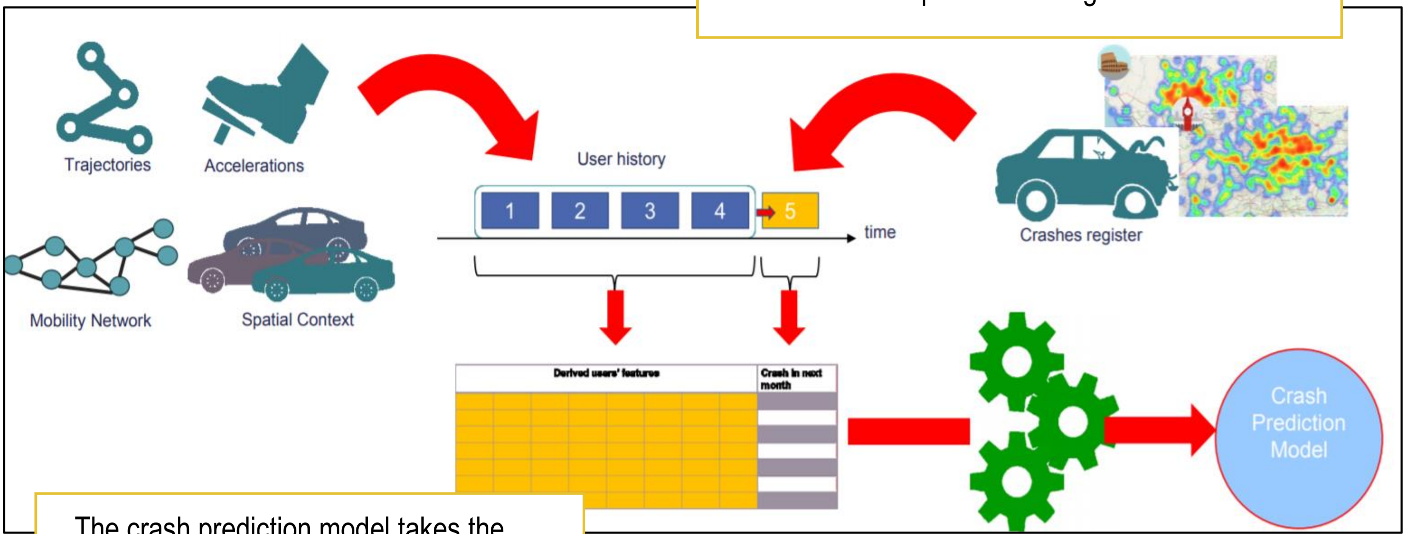
Crash prediction using vehicle mobility data and individual mobility networks (CNR and Sistematica)

Dr Mirco Nanni from CNR (Italy) together with Mr. Leonardo Longhi from Sistematica (Italy) presented their work on crash prediction at **i-CiTies 2019, the 5th Italian conference on ICT for Smart Cities and Communities**. Their research addresses the problem from an individual mobility data perspective, analysing the mobility of users based on GPS traces of private vehicles, and trying to identify features of the user's mobility and driving behaviour that correlate with accidents.

Their approach is different from most existing approaches that examine the problem either from a geographical and temporal perspective (i.e. identifying the areas and hour of the day where the risk of accident is higher) or from a real time one (i.e. identify the instantaneous conditions that are likely to lead to accident).

This work is in direct relation to the auto-insurance pilot business case, where the key objective is to **develop some telematics tools to insurance stakeholders in order to profile a personal insurance risk giving margins to reduce insurance fees to virtuous drivers.** Prediction of crash in the near future such as within the next month is based on the data of the previous 4 months. This is done via extracting several features from the history of the individual (such as trajectory features, event features, individual mobility network and spatial context) by employing machine learning technique. The framework is depicted in the figure below.

Fig. 1



The crash prediction model takes the following inputs :

1) **Individual trajectory features**, such as trips count, lengths, durations, time-of-day, average speeds etc. which can be extracted easily from GPS data
2) The dataset also provides **events that took place while driving**, such as acceleration, brakes, cornering and zigzaging
3) **Car brand and model information**

The resulting dataset has been used as input for various traditional machine learning models, including Random Forests (RF), Support Vector Machines (SVM) and Neural Networks (NN). Random Forests yielded the best and more stable results, shown in the below figure. The table also divides performances over different subsets of features (traj = travel only, evnt = events only, trev = both, all = include also brand and model). The results were computed over a sample of data, covering vehicles in Rome and London, and the corresponding model parameters were selected by grid search optimization.

As an ongoing work, the crash prediction model will further incorporate an important input which are the various features extracted from Individual Mobility Networks (IMN). An IMN is a concise graph representation of the mobility history of individuals. From raw GPS traces the trajectories of a single mobility user are reconstructed and processed to infer the relevant locations that the user visited (the nodes of IMNs) and aggregate the trips between two locations (the edges of IMNs).

clf	features	f1-score	precision	recall	test_accuracy	train_accuracy
RF	all	0.659024	0.709972	0.655093	0.712644	0.815271
RF	evnt	0.641626	0.734188	0.663095	0.672414	0.757389
RF	traj	0.636443	0.723219	0.655688	0.669540	0.748768
RF	trev	0.603573	0.651994	0.609672	0.658046	0.795567

Fig. 2

Several features are extracted that will be used as **key variables** in the model. These are as follows:

1) Number of nodes and edges
2) Node and Edge Frequency
3) Clustering co-efficeint of IMN
4) Network modularity index
5) Measure how much IMN change with repect to time that reflect changes in habits or mobility needs

For more updates, please contact **Dr. Mirco Nanni** (mirco.nanni@isti.cnr.it)

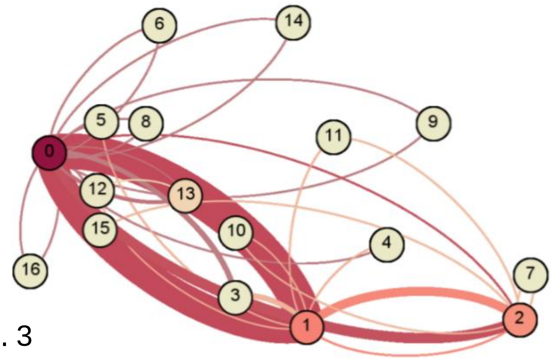


Fig. 3

Congestion detection and alternative routes (UHasselt)

Dr. Feng Liu from UHasselt and other researchers from partner organizations have presented a method to analyse congestion in the city using trajectory data. This work is in relation to fulfill an important objective of the ongoing fleet management pilot, which aims to detect congestion and proposed alternative routes to reduce travel time and other associated costs. The proposed method combines a visual analytical tool with a statistical analysis system so they tightly integrate and complement each other to allow end-users to work with extremely large volumes of mobility data and to gain more insights from the derived results. Their framework is visualized in the following figure.

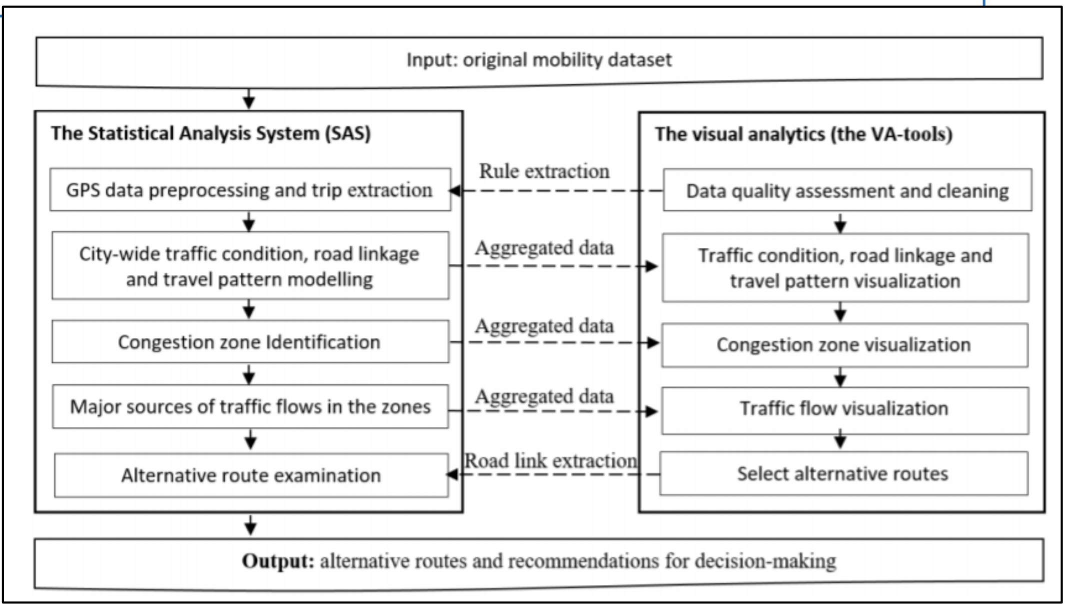


Fig. 4

The key steps involved in the method are as follows:
(1) The entire study area is divided into zones, and models characterizing traffic conditions, road linkage and travel patterns of each zone or zone pair are constructed.
(2) Based on the models, congestion zones are detected.
(3) The major sources of traffic flows in the congestion zones are analyzed.
(4) Alternative routes that avoid the congestion areas are identified, and differences between these routes and the original trips are examined.

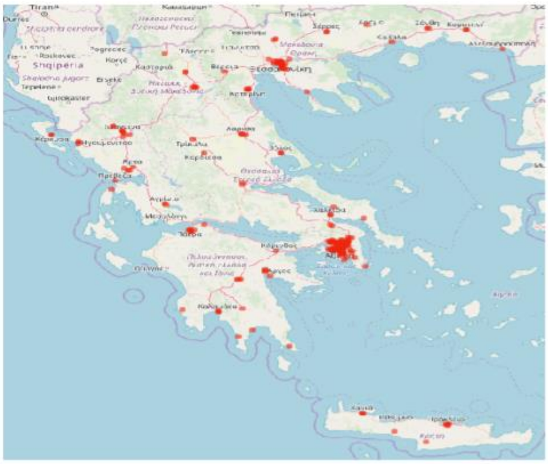
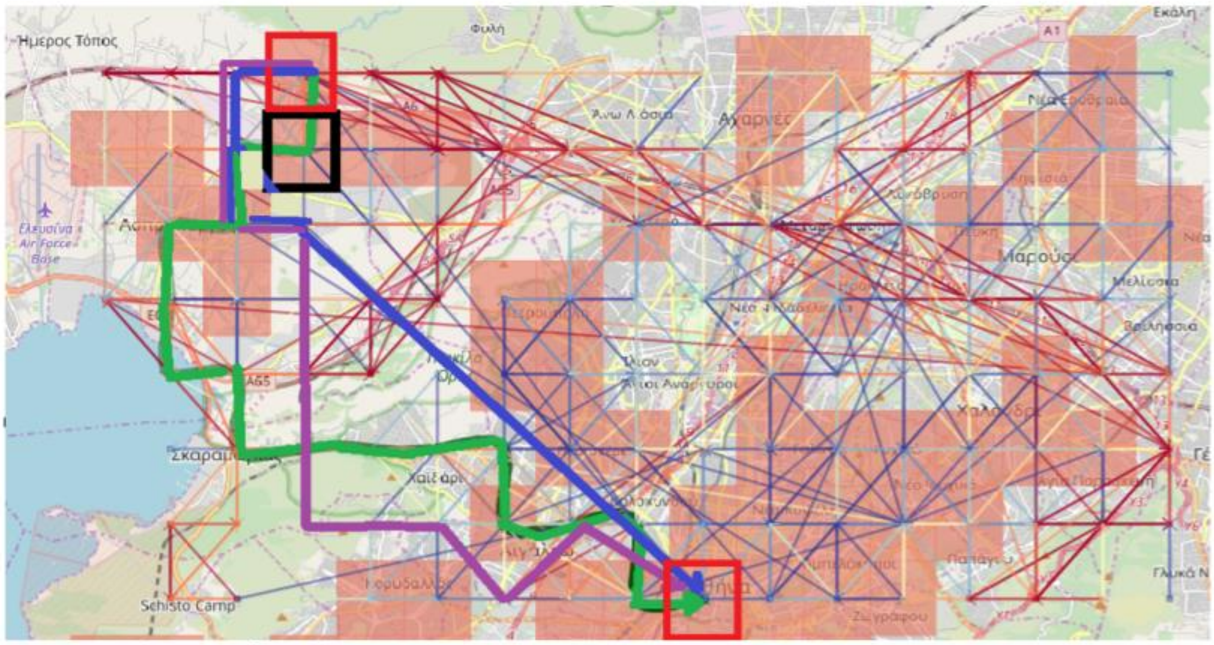


Fig. 5

The application of the method is demonstrated for morning rush hour (7-9am) on weekdays in the major data collection area of Attica Basin in Greece. To identify congestion zones, the threshold parameters such as TH_{mz}, TH_{vz} and TH_{con} are designated as 20, 20km/h and 0.8 respectively, leading to 212 zones being filtered out as the congestion zones. These places have undertaken at least 20 trips in the morning per weekday, and experienced average travel speeds lower than 20km/h at least for 4 in 5 of the days. The figure depicts the geographic distribution of the congestion zones (i.e. red areas in the figures).

Using the visual analytic tools, road linkage between various zones are demarcated. Furthermore, for a given origin-destination (OD) zone of the trip within a visual analytical tool , the actual trip and alternative routes are interactively marked. These data are then again analysed using a statistical tool to obtain statistics related to alternative routes. The blue and purple lines are the two proposed routes, and the grids in the red rectangles are the origin and destination zones. The green line is the original trip between this OD pair passing the congestion zone enclosed in the dark rectangular.

Fig. 6



Contextualized analysis of movement events (Fraunhofer)

Modern movement tracking technologies (e.g. GPS devices accompanied by other sensors) can not only record positional information of each point (e.g. coordinates and time), but also attributes of the moving objects (e.g. speeds and fuel levels), environment (e.g. temperature), and movement events (e.g. acceleration, braking and cornering) that occur during the movement. Understanding the conditions in which movement events happen can be important in various applications. For instance, in logistics or car insurance business, there is a great interest in detecting bad driving behaviour as well as in identifying locations and time characterized by increased risk of crashes.

Previous research dealing with movement events has been mostly focused on extraction of various events from trajectories. Once extracted, the events were analysed separately. Fraunhofer researchers propose a new perspective for investigating movement events: contextualized analysis. The approach involves a **procedure of event contextualization**, which consists of (1) extracting segments of trajectories preceding and following the events, and (2) enriching these segments with values of additional attributes characterizing the movement and/or describing the external circumstances.

To demonstrate the applicability of the proposed approach, the 'harsh braking' events in the context of the speed attribute were investigated. For event contextualization, a suitable time-window needs to be chosen. After experimenting with different values as well as discussing with domain experts, the researchers have set the time-window at 10 min. This window was further divided into 10 time-steps (i.e. m=10). Hence, each harsh braking event is characterized by a vector of 10 speed values. Based on the interactive-cluster method, they iteratively performed different clustering with the cluster number varying from 5 to 20. Finally, the number of clusters was chosen as 11, which can be seen in figure 7. Each of the 11 points in the left panel represents a cluster, and the points with shorter distances or similar colors imply more similarity of the corresponding clusters. The parameters on the right panel control the way of the projection displays and the parameters of clustering.

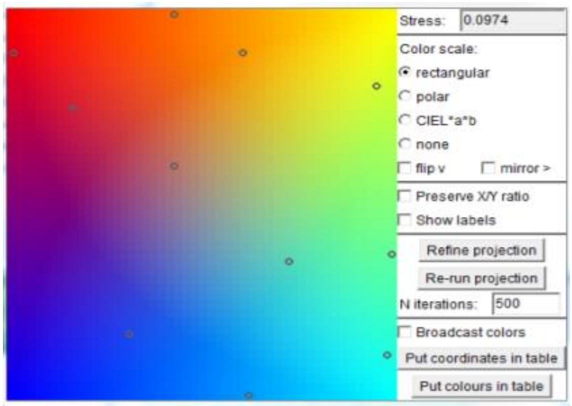


Fig. 7

The figure below demonstrates the attribute matrixes of the derived clusters; in each matrix, the rows represent events and the columns are the time-steps. From this figure, these clusters can be divided into two groups of speed patterns: low-speed (Fig 8 a) and high-speed (Fig 8 b).

It was noted that the low-speed and high-speed groups correspond to intra-city and inter-city movement, respectively. This makes sense as the speeds on the inter-city roads are generally higher. Different temporal patterns are observed: the intra-city movement events happen more often at the day time while the inter-city events last till late at night. Cluster 1 and 3 represent harsh brake events that happened in the course of slow movement before the speeds started to increase.

These patterns may correspond to final parts of the movement in traffic jams. The cluster 8 is different; speeds decreased from normal to slow, and then harsh braking happened just before accelerating to normal speeds.

The domain experts interpreted these patterns as probably corresponding to stopping at red lights on street intersections or encountering slow moving vehicles ahead. These clusters can be presented in geographical regions. Based on this information, it could be recommended to avoid the identified regions at a specific time or to drive there with special attention.

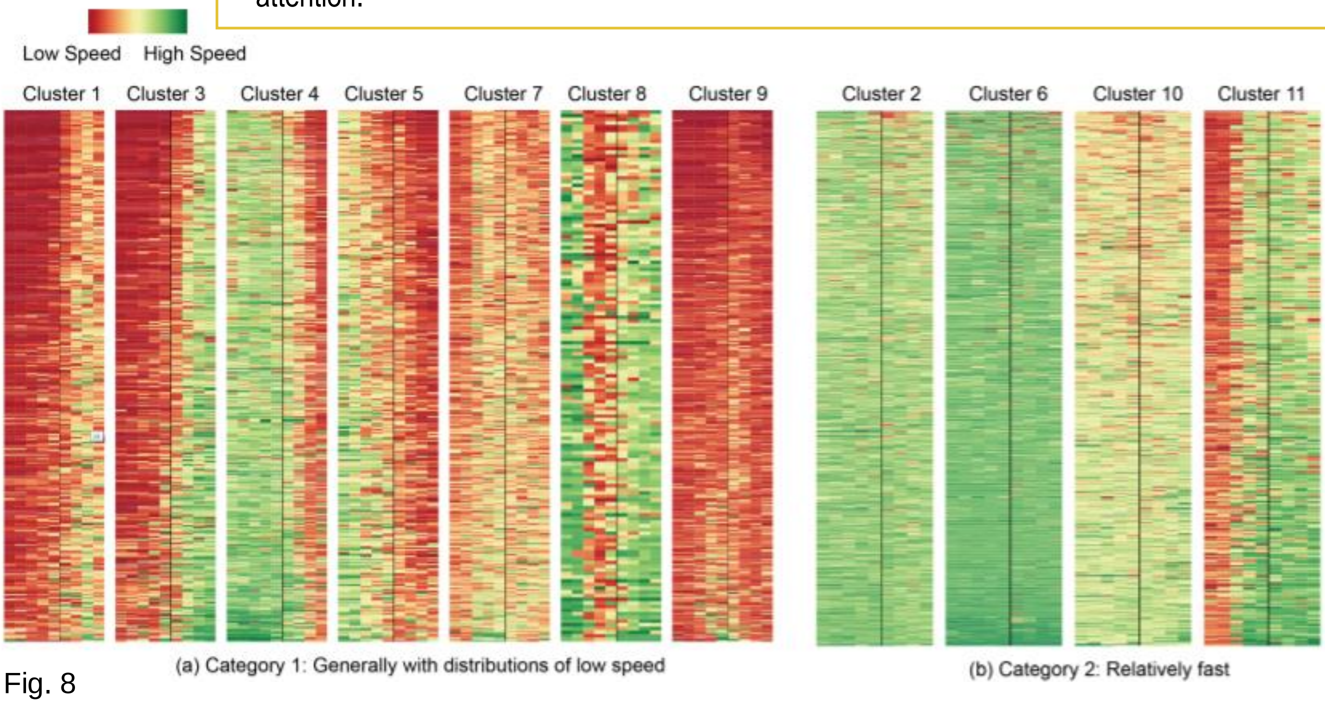


Fig. 8

Unified NoSQL Data Access Operators (NoDA) (UPRC)

Researchers from UPRC developed an approach which they called a NoDA, which is an abstraction layer in the form of an API for querying NoSQL stores with native support for geospatial operations. The salient features of this approach are as follows:

- Query different NoSQL stores using a single code base
- Seamless transition of application code from one NoSQL store to another
- NoDA is JVM-compliant and can return results as Spark Data frame
- NoDA enables operator ‘push-down’ (filter, project, aggregate) to the storage engine to increase efficiency
- Supports spatio-temporal operations

More details can be found in this [paper.]

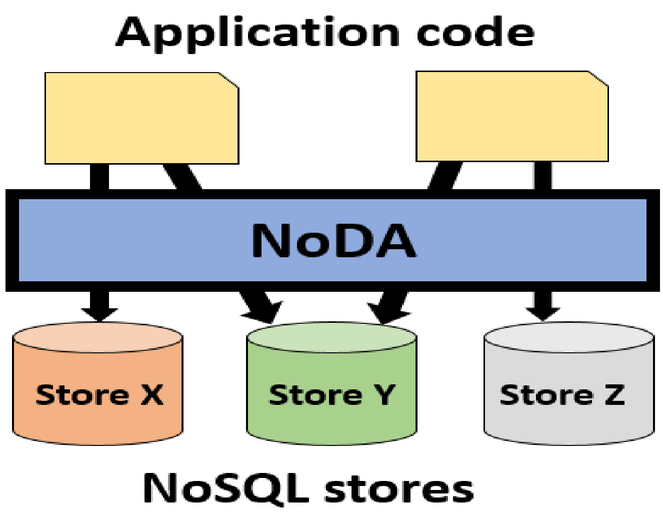


Fig. 9

Key Dissemination and Liaison Activities

Prof. Yannis Theodoridis (from UPRC) delivered a lecture at the Austrian Institute of Technology

As part of an AIT lecture series, Prof. Yannis Theodoridis has delivered a lecture on 'Learning from movements - A roadmap towards mobility data analytics' on the 6th of June 2019. Prof. Yannis discussed the research progress of the Track & Know project and what will be done in the future. In addition to this, ongoing research activities at both institutes were discussed at length to assess possible future collaboration.

Health pilot case: Network lunch session in Cambridge to engage with stakeholders

On June 17, Royal Papworth OSA service and Cambridge Medical Academy team (as project partners of Track & Know) organised a network lunch session to engage with stakeholders. The event was held at Cambridge Biomedical Campus (UK).

The seminar attracted new stakeholders for feedback and further project engagement, such as an Operations Manager of OSA (Obstructive Sleep Apnea) sleep services, a CPAP (Continuous Positive Airway Pressure) practitioner (OSA therapy machine), a Philips (CPAP manufacturer) Clinical Services Manager, a Public Health/IHR, R&D Manager Cambridge University Hospitals, Engineering Department University of Cambridge, an NHS project manager and a Psychology researcher University of Cambridge.

Valuable feedback was received for the service optimisation concept, with additional suggestions for transferability of the results to several other domains in health care. Support has been offered for the further development of the OSA risk map and the analysis of actual demand based on existing data. The dashboard development triggered a productive discussion and further user requirements were identified. The feedback on the concept of the driving profile study was very positive.

Track & Know story shortlisted for 'Success Story Awards 2019'

Next to participating and showcasing the Track & Know project at the BDV PPP summit 2019 in Riga (Latvia), Track & Know also reached the shortlist of the 'success stories awards', organized by BDVE Project and BDVA. Track & Know liaison project 'Transforming Transport' ended up scoring the first prize.

The Track & Know consortium submitted two stories for this contest of which one story, titled **'Sleep well – drive safely – a tale of mobility tracking and Big Data'**, successfully made it to the shortlist. Dr. Angelos Liapis, CEO of Track & Know partner Konnekt-able Technologies Ltd., did a great job presenting the Track & Know success story. His presentation attracted excitement and gained visibility for the entire project. Please find the story here.

The annual BDV PPP Summit is the primary event for driving European innovation in **Big Data** and **Artificial Intelligence**. Key European industry, academia and policy-making players gathered in Riga at the end of June to foster cross-sector collaboration and shape strategies for European leadership in data-driven Artificial Intelligence. Each year, the Summit welcomes hundreds of organisations involved in Big Data Public Private Partnership as well as all those who want to be part of the thriving European Big Data Ecosystem.

Track & Know results presented at BigMedilytics event in Valencia

On September 4 and 5, the intermediate results from the health pilot of Track & Know project were presented at the BigMedilytics event titled 'Big Data: Fueling the transformation of Europe's Healthcare Sector' which was held at the Polytechnic City of Innovation in Valencia (Spain). Jenny Rainbird from Inlecom Systems (coordinator of the Track & Know Project) presented Track & Know results.

The event involved organizations covering the key players in the healthcare sector such as healthcare providers, health technology companies, payers, research institutes and academia from across Europe. The event also brought together high-level speakers.

Track & Know health case pilot presented at Taizhou Medical Expo 2019

Toni Staykova from Track & Know partner, Cambridge Medical Academy, visited the Thaizhou Medical Expo 2019 – China Medical City on September 18 and 19, 2019. At the expo, she explained the big data use for service optimization in regards to the health case pilot.
The expo welcomed over 1000 visitors, mostly coming from industry and production of supplies and pharmaceuticals. Medical professionals and nursing staff also visited the expo.

PUBLICATIONS 2019

- Nikitopoulos P., Sfyris G. A., Vlachou, A., Doulkeridis C., Telelis O. (2019) Parallel and Distributed Processing of Reverse Top-k Queries, In *Proceedings of the 35th IEEE International Conference on Data Engineering* (ICDE 2019). Download **here**
- Andrienko N., Andrienko G., Garcia J. M. C., Scarlatti D. (2019) Analysis of Flight Variability: a Systematic Approach, *IEEE Transactions on Visualization and Computer Graphics*, vol. 25(1), pp.54-64. Download **here**
- Liua S., Andrienko G., Wu Y., Cao N., Jiang L., Shi C., Wang Y. S., Hong S. (2019) Steering Data Quality with Visual Analytics: the Complexity Challenge, *Visual Informatics*, vol 2(4), pp.191-197. Download **here**
- Koutroumanis N., Santipantakis G., Glenis A., Doulkeridis C.,Vouros G. (2019) Integration of Mobility Data with Weather Information*, In proceedings of the International Conference on Extending Database Technology (EDBT) and International Conference on Data Theory (ICDT) Workshops 2019*, Lisbon, Portugal. Download **here**
- Guidotti, R, Monreale A, Cariaggi L (2019). Investigating Neighborhood Generation Methods for Explanations of Obscure Image-Classifiers, In: Yang Q., Zhou ZH., Gong Z., Zhang ML., Huang SJ. (eds) *Advances in Knowledge Discovery and Data Mining. PAKDD 2019. Lecture Notes in Computer Science, vol 11439*. Springer, Cham. Download **here**
- Chen S., Andrienko G., Andrienko N., Doulkeridis C., Koumparos A., (2019) Contextualized Analysis of Movement Events, In: T. V. Landesberger and C.Turkay (Eds) *EuroVis Workshop on Visual Analytics (EuroVA)*, pp: 49- 53, The Eurographics Association. ISBN: 978-3-03868-087-1, DOI: 10.2312/eurova.20191124. Download **here** (This publication received the 'EUROVA 2019 best paper award')
- Koutroumanis N., Nikitopoulos P., Vlachou A., Doulkeridis C. (2019): NoDA: Unified NoSQL Data Access Operators for Mobility Data. In *Proceedings of the 16th International Symposium on Spatial and Temporal Databases (SSTD'19).* (to appear). (to appear)
- Tsilionis E., Koutroumanis N., Nikitopoulos P., Doulkeridis C., Artikis A. (2019): Online Event Recognition from Moving Vehicle. In *Proceedings of the 35th International Conference on Logic Programming (ICLP'19)*. (conditionally accepted)
- Nanni, M., Longhi, L. (2019): Vehicle mobility data analysis and Individual Mobility Networks for crash prediction. Download **here**

Find & Follow us
Website: https:// trackandknowproject.eu
Twitter: @TrackandKnow
LinkedIn: Track & Know

The project has received funding from the European Union's
Horizon 2020 Research and Innovation Programme under Grant Agreement No 780754.