



# Learning from Our Movements – The Mobility Data Analytics Era

Yannis Theodoridis<sup>(✉)</sup>

Data Science Laboratory, University of Piraeus, Piraeus, Greece

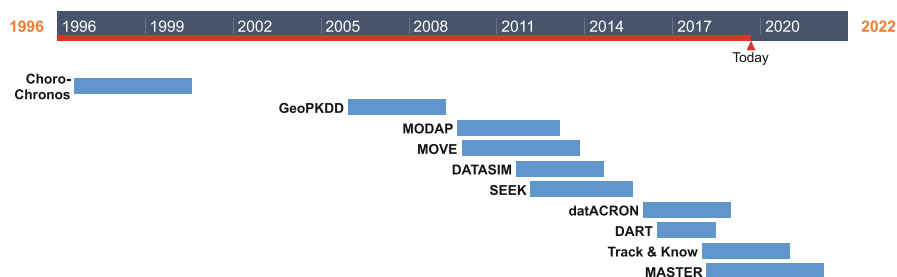
ytheod@unipi.gr

<http://www.datastories.org/>

**Abstract.** From the pioneering works on spatiotemporal databases back in ‘90s to the era of Big Mobility Data Analytics nowadays, in this paper we try to follow the thread of research objectives and initiatives in the field. Initially, we provide a flashback to the 25 past years (though from a personal, hence, biased point of view). Then, we discuss in brief the challenges related to mobility data processing, management, analytics, and visualization to be addressed in modern applications tracking populations of moving objects in real time.

## 1 Introduction

Once upon a time, it was the ChoroChronos EU research project. The challenge at that time (‘90s) was to bring spatial and temporal database aspects together in a, then emerging, integrated spatio-temporal domain. As time was passing, new challenges appeared and addressed by the researchers of the field: efficient system architectures, knowledge discovery from mobility data, privacy aspects, etc. Nowadays, in the era of Data Science and Big Data, mobility data analytics aims at learning from objects’ movements, covering a range of methods and solutions, from de-noising of location information and integrating with multiple



**Fig. 1.** Related research projects timeline (mid ‘90s - today).

heterogeneous related sources to predictive analytics, offline and online. Figure 1 illustrates a timeline of related research projects covering the period from mid ‘90s until today<sup>1</sup>.

## 2 Flashback to the Past

Let’s catch the thread from the beginning... Back in ‘90s, research in spatial and temporal databases, separately, resulted in pretty mature results to contribute in real-world DBMSs. In the spatial database field, Oracle introduced Spatial Data Option in 1996, OGC released its first specifications in 1997, and PostGIS was launched in 2001. On the other hand, TSQL2 language specification was developed in 1993. The “marriage” of the two fields was led by research projects in both US and Europe; focusing on the latter, the notable ChoroChronos EU project (1996–2000) aimed at bringing together the two communities and integrate their ideas in the so-called **spatio-temporal databases**, where time would be considered a first-class citizen [2, 4, 5].

What followed was the focus of research in point objects due to the popularity of related applications (tracking of moving objects via GPS technology), which led to the “**moving object trajectory**” concept<sup>2</sup> and, as expected, raised challenges on knowledge discovery from this new type of data as well as on personal data privacy. This brought the “dialogue” with other than database management domains, including machine learning/data mining and data privacy and security. For instance, the GeoPKDD (2005–2009), MODAP (2009–2012) and MOVE (2009–2013) EU projects aimed at **devising knowledge discovery and (privacy-preserving) analysis methods for trajectories of moving objects**, bringing together ICT researchers and domain specialists [3, 8]. The advances in social networks and linked open data in ‘00s also resulted in the so-called location-based social networking and relevant applications. Thus, a new trajectory variation was born, the **semantic trajectories** [6], studied by e.g. the SEEK EU project (2012–2015), the objective of which was to envisage a new semantic enriched knowledge discovery process where the semantic aspect (in the sense of the meaning of the movement) would be embedded in each step.

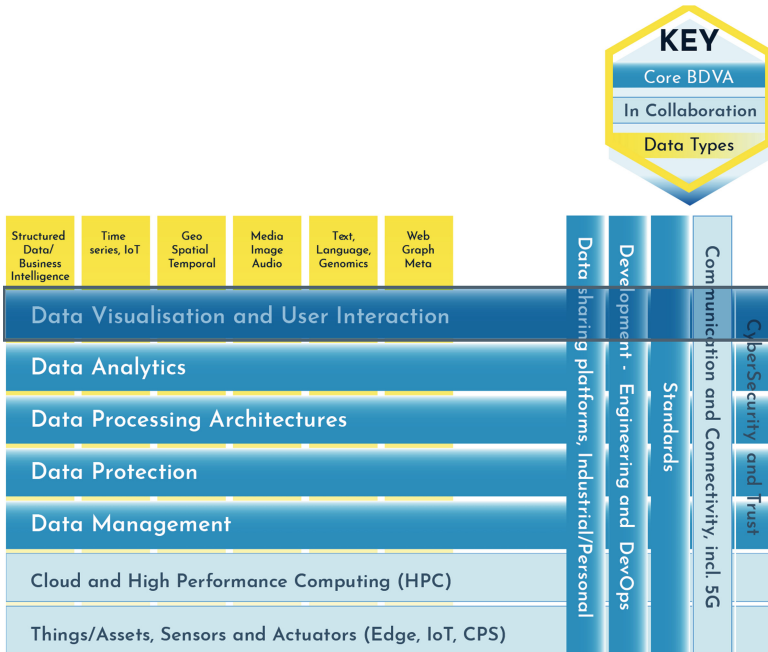
## 3 Nowadays - Mobility Data Analytics

Nowadays, one can find plenty of sensor data and open sources of related information, mature Big Data technologies, plethora of Data Science methods and tools. In this environment, a hot research topic is that of **Mobility Data Analytics** (MDA). The range of processes covered under this term includes data

<sup>1</sup> Disclaimer: the flashback in past is biased since it only refers by name to projects where the author has participated. The author’s intention is by no means to provide an exhaustive survey of research activities related to the topic of the article.

<sup>2</sup> According to DBLP ([dblp.uni-trier.de](http://dblp.uni-trier.de)), the first papers with this term in their title appeared in 2000.

acquisition and processing (typically from multiple and varying data sources), data management (storage and indexing, as usually...), data mining, data privacy, data visualization and user interaction [7].



**Fig. 2.** The BDVA reference model (source: <http://www.bdva.eu>).

A typical MDA application tracks in real-time a population of humans / vehicles / vessels / aircrafts and handles the resulted trajectories, enriched with heterogeneous context, in order, for instance, to be able to assess traffic situation or drivers' behavior, forecast anticipated movements in short-term or schedule traffic in long-term, react as soon as an “anomaly” is detected, and so on [9]. In this framework, challenges touch almost each architectural layer of an MDA system, compatible with the EU BDVA reference model (illustrated in Fig. 2):

- **data sources** of interest include, on the one hand, streaming information, such as the GPS signals transmitted by the objects themselves, the objects' locations tracked by an external device (e.g. radar), and live weather information, and, on the other hand, archive collections of census, meteorological, etc. data; all of this information to be correctly and smoothly integrated;
- **data processing** requires cleansing (de-noising, smoothing) and semantic enrichment of incoming data as already mentioned, segmentation into trajectories that make sense (e.g. from/to predefined places of interest), and storage in appropriate (relational or NoSQL) stores;

- **data management** includes efficient querying and retrieval, which assume implementation of query algorithms and maintenance of indexing mechanisms suitable for these purposes;
- **data analytics** requires both offline and online algorithms: cluster and frequent pattern analysis for detecting typical movement patterns as well as interesting outliers can run offline on the historical data whereas movement prediction and anomaly detection over the incoming stream should run online; in order for the online methods to be effective, they should take into consideration the results of the offline methods (a now tracked object may be considered to have anomalous behavior if it is quite dissimilar to past typical patterns or quite similar to past outliers) and this makes things even more challenging;
- **data visualization** and user interaction is an essential component in order for the mobility analyst to get familiar with the data he/she is requested to analyze and interact with the above methods and tools; visual analytics (VA) is challenging, especially in mobility data [1].

This is, more or less, the context of a number of recent EU projects, including datAcron (<http://datacron-project.eu>; 2016–2018), where the use cases are on maritime and aviation, Track & Know (<https://trackandknowproject.eu>; 2018–2020) on drivers in urban environment, and MASTER (<http://www.master-project-h2020.eu>; 2018–2022) on land transportation, sea monitoring, and tourism.

## 4 What’s Next

Fortunately, research is an everlasting story. So, what’s next? In the near future, we expect to see advances, for instance, in self-organizing and self-cleansing information integration tools (do we know what information is relevant to enrich a location or assess its accuracy better than a crawler searching the Semantic Web?), close to storage-less architectures following the IoT paradigm (if data resides in its original source, what is the need to replicate it in our storage? what if data is unaffordable huge and only meta-data as well as patterns are stored locally?), and Explainable AI (why this and not that technique in order to analyze location data? our right to explanation). We expect these and other even more exciting research outcomes to appear in the years to come.

## References

1. Andrienko, G.L., Andrienko, N.V., Bak, P., Keim, D.A., Wrobel, S.: Visual Analytics of Movement. Springer, Berlin (2013). [https://doi.org/10.1007/978-3-540-75177-9\\_14](https://doi.org/10.1007/978-3-540-75177-9_14)
2. Frank, A.U., et al.: Chorochronos: a research network for spatiotemporal database systems. SIGMOD Rec. **28**(3), 12–21 (1999)

3. Giannotti, F., Pedreschi, D. (eds.): *Mobility, Data Mining and Privacy - Geographic Knowledge Discovery*. Springer, Berlin (2008). <https://doi.org/10.1007/978-3-540-75177-9>
4. Güting, R.H., et al.: A foundation for representing and querying moving objects. *ACM Trans. Database Syst.* **25**(1), 1–42 (2000)
5. Koubarakis, M., et al.: *Spatio-Temporal Databases*. LNCS, vol. 2520. Springer, Heidelberg (2003). <https://doi.org/10.1007/b83622>
6. Parent, C., et al.: Semantic trajectories modeling and analysis. *ACM Comput. Surv.* **45**(4), 42:1–42:32 (2013)
7. Pelekis, N., Theodoridis, Y.: *Mobility Data Management and Exploration*. Springer, New York (2014). <https://doi.org/10.1007/978-1-4939-0392-4>
8. Renso, C., Spaccapietra, S., Zimányi, E. (eds.): *Mobility Data: Modeling Management and Understanding*. Cambridge University Press, Cambridge (2013)
9. Vouros, G.A., et al.: Big data analytics for time critical mobility forecasting: recent progress and research challenges. In: *EDBT*, pp. 612–623 (2018) [OpenProceedings.org](https://openproceedings.org)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

