The technologies described above provide basic tools for efficient multivariate data analysis. Such an analysis can be performed in order to enhance the process of decision support by extracting new knowledge from the data. OLAP systems are supplemented by data mining processes which, in contrast to OLAP systems where data analysis is driven by user queries, allow for automatic data analysis and knowledge extraction. Traditional data warehouses and OLAP systems are dedicated mainly for processing of transactional data represented as sets of independent data. A typical example of an OLAP analysis on transactional data is a query informing about the sales of various types of AGD devices (refrigerators, washing machines, vacuum cleaners) in each quarter of a given year in every province.

However, nowadays, more and more systems and applications generate and process data characterized mainly by the order in which they emerge. Such a data type is called sequential data and, as already mentioned, is currently one of the most prominent data types in many IT systems: ticket systems (analyzing travelling patterns in public transportation), workflow management systems, medical applications, systems analyzing the behavior of users in the World Wide Web, intelligent transportation systems based on RFID technology, infrastructure management systems (e.g., intelligent buildings, systems remotely measuring the consumption of energy, gas or water), bioinformatics (DNA sequence analysis). Sequential data processing still remains a very important research area in fields of database systems, workflow systems, and bioinformatics. There have been several important findings in this research area regarding: transactional processing of sequential data, execution plan optimization, novel indexing structures enhancing various types of sequential queries. Unfortunately, far less has been done in terms of analytical processing (OLAP) and storing sequential data in data warehouses. As mentioned above, OLAP systems and data warehouses available today have been designed with transactional processing in mind. As a result, they do not support analytical or exploratory processing of sequential data. In recent years, with the increasing popularity of sequential data, there has been a rapid growth in interest in the field of sequential on-line analytical processing – SOLAP.

Analytical processing of sequential data is a difficult task due to a variety of sequential data types (discrete, interval, continuous) as well as high complexity of aggregating sequential data. Sequential data can be aggregated based on: values of the attributes describing the elements of sequences, sequences, or subsequences contained within the sequences. Additionally, SOLAP systems allow for clustering of sequences based on frequent patterns appearing within them (e.g., what is the number of passengers using a public transportation according to the following pattern: home-work-work-home – XYYX). Such types of queries are commonly referred to as pattern-based queries.

The aim of this project is to develop new solutions for advanced analytical processing and data mining analysis of sequential data in data warehouses. Two main types of sequential data will be analyzed: sequences of categorical attributes (called also time-point-based sequential data or discrete data sequences) and sequences of interval attributes (called also interval-based sequential data). In particular, the project will focus on developing solutions to the following research problems: (1) new data models representing time-point-based and interval-based sequential data, (2) a query language for sequential data OLAP processing, (3) data storage architectures suitable for OLAP processing of sequential data, (4) new indexing structures and techniques supporting analytical processing of sequential data, (5) new algorithms for mining sequential data (particularly: classification and clustering algorithms), and (6) implementation of a sequential data warehouse and OLAP system available as a public service according to the SaaS (software as a service) architecture.