

## BigGIS - Stream Enrichment With Temperature Data

*Wednesday, October 11, 2017 4:00 PM (30 minutes)*

The BigGIS Project aims to develop a new GIS to deal with the challenges of Big Data in the field of Geo Information Systems (GIS). Historically, data in this field is available in pre-defined data formats, has a high data quality and the data is not in motion. The data volume was, apart from few exceptions, relatively small. In the last years this has changed. Based on many new data gathering devices such as smart phones or volunteered geographic information (VGI) with e.g. citizen weather stations, data is now available in a size never before seen. But this new data presents its problems especially with the velocity, the variety and the veracity of the data.

In contrast to older data gathering methods, new sensors create data every second instead of days. Smart phones alone leave a steady stream of data. The data sources themselves are saved in many different formats. From polygons and vector based formats to raster files, from yearly aggregates to second measurements and any combinations, from single point measurements to video files, geo data assume a high variety of different formats. Lastly, the quality of the data is not clear anymore. In the past, data was gathered via expensive methods and guidelines as well as methodical planning went into the guarantee of the quality of data. Today, a researcher or practitioner cannot be sure about the quality and has to filter or develop methods how to deal with these problems.

In BigGIS, we introduce a novel continuous refinement model to deal particular with veracity and uncertainty in spatio-temporal big data by an integrated data processing pipeline that leverages big data analytics frameworks, semantic web technologies and visual analytics methodologies. By using the well-established pipes and filters architectural pattern and incorporating uncertainty in our statistical modeling approaches, we address the

challenges of modern GIS. In this presentation, we will give an overview of how data streams can be enriched with additional information from other data sources in a stream enrichment pipeline in BigGIS

We present our approach and the underlying architecture in the field of temperature prediction. Based on existing measurements from satellites, VGI and official weather stations we combine different spatio-temporal information to create an enriched output vector upon which we can reliably predict temperatures at any point given reference point in time. We include the uncertainty by use of Bayesian Hierarchical Modeling and enhance the accuracy by including the knowledge of the user by using visual analytics.

The benefits of our approach are shown in two scenarios: smart city and the detection of invasive species. In the case of a smart city, accurate temperature distribution maps as well as their underlying reasons can be used to measure and mitigate the impact of heat on the human health as well as reduce the energy costs derived from temperature. In the case of invasive species, the knowledge of the temperature is essential to detect their habitats as well as potential breeding grounds.

We demonstrate that the stream enrichment pipeline in BigGIS is efficient in the processing of data and generating valuable insight. In cooperation with the smart data innovation lab (SDIL) we show the potential use of our stream enrichment approach for data cleaning in order to guarantee certain data quality levels.

### Track

BDAHM

**Author:** Mr BRUNS, Julian (FZI Forschungszentrum Informatik)

**Co-authors:** Prof. NIMIS, Jens (Hochschule Karlsruhe); Mr FRANK, Matthias (FZI Forschungszentrum Informatik); Mr WIENER, Patrick (Hochschule Karlsruhe); Prof. SETZER, Thomas (FZI Forschungszentrum Informatik); Dr SIMKO, Viliam (FZI Forschungszentrum Informatik)

**Presenter:** Mr BRUNS, Julian (FZI Forschungszentrum Informatik)

**Session Classification:** Use-Cases