Research Project

# weObserve: Integrating Citizen Observers and High Throughput Sensing Devices for Big Data Collection, Integration, and Analysis

**Third-party funded project**

**Project title** weObserve: Integrating Citizen Observers and High Throughput Sensing Devices for Big Data Collection, Integration, and Analysis
**Principal Investigator(s)** Roth, Volker ;
**Co-Investigator(s)** Schuldt, Heiko ; Alewell, Christine ; Liechti, Felix Stephan ;
**Project Members** Samarin, Maxim ; Zweifel, Lauren ;
**Organisation / Research unit**
Departement Mathematik und Informatik / Biomedical Data Analysis (Roth)
Departement Mathematik und Informatik / Databases and Information Systems (Schuldt)
Departement Umweltwissenschaften / Umweltgeowissenschaften (Alewell)
**Department**
**Project start** 01.05.2017
**Probable end** 30.06.2021
**Status** Completed

Even though hypothesis driven research is the fundamental core of scientific advance, scientific progress by monitoring and subsequently analysing is crucial for certain phenomena of the real world. Specialized high throughput sensor devices can be used in controlled lab environments that provide very large data collections. These collections can be analysed to come up with new findings by verifying or falsifying concrete hypotheses. However, the majority of scientific domains is more complex and cannot rely on such rather simple data gathering and processing pipelines: first, the phenomena to be monitored in the real world are complex in their spatial and temporal dynamic, and confounding factors are typically of multi-causal origin. Thus, the phenomena can-not be isolated nor can natural environments be rebuilt in controlled lab environments, which was one of the lessons learned from the Biosphere II programme. Monitoring in the field is essential, but it can hardly be done with conventional sensor technology only at landscape scale, since there are always technical trade-offs between spatial resolution, coverage, temporal resolution and in-terpretability. Furthermore, even if resolution and coverage of sensors is satisfactory, it is not obvious where and when to deploy these high precision/throughput sensors to capture relevant phenomena. In the weObserve project, we will rely on citizen observers to provide semantically rich information directly from the field to complement and enrich existing sensor data. In addition, monitoring data from citizen observers will be used to anticipate where interesting phenomena are supposed to take place, to use societal knowledge and judgement on relevance of phenomena and to deploy high resolution sensing devices in these areas.

From a technical point of view, weObserve will address the collection, integration, and processing of het-erogeneous data in applica-tions which cannot rely on off-the-shelf sensing devices for moni-monitoring purposes. Heterogeneity includes the volume of data provided via different channels, the precision, the coverage in time and location, and also the predictability. Data collection will there-fore seamlessly combine several types of data sources: (i) high throughput sensing devices which produce very large volumes of data and cover large areas, but with rather low resolution, (ii) Citizen Observers which pro-vide semantically rich data, but with varying levels of precision and substantial sampling bias, and (iii) specific high resolution sensing devices that need to be manually deployed. Data integration will deal with such heterogeneous data. For subsequent analysis, the origin (provenance) and uncer-tainty of in-dividual data items needs to be kept in an integrated data set. Data analysis will detect hidden patterns

and the main explanatory factors in data collections. Integration of domain knowledge into the analysis process will be essential for detecting sampling biases and confound-ing factors. Specific emphasis needs to be put on analysis models that can deal with multiple data queues varying in size, reliability, representation and resolution. A further important aspect con-cerns visualization of detected patterns as a means for improving communication between those using the data for scientific purposes and those collecting it. WeObserve will design, implement, integrate, and evaluate the individual parts of the data collection/integration/analysis pipeline in two selected applications, namely i.) monitoring of soil degrada-tion and landslides, and ii.) monitoring of bird migration, with complementary requirements and different ways to gather data.

**Financed by**
Swiss National Science Foundation (SNSF)

**Add publication**

**Add documents**

**Specify cooperation partners**

weObserve: Integrating Citizen Observers and High Throughput Sensing Devices for Big Data Collection, Integration, and Analysis