The PermaSense Mountain Lab – Technology and Infrastructure for an Open Permafrost Data Repository

Jan Beutel[1], Samuel Weber[2], & Tonio Gsell[1]

[1]*ETH Zurich, Switzerland*
[2]*University of Zurich*

Starting in 2006 as a single research project PermaSense has pioneered experimental permafrost research in alpine environments using spatially distributed online wireless sensors. Conceived as a focused investigation to understand the utility of many miniature battery powered wireless sensors to be used in high-alpine permafrost research has since developed into a full-fledged multi-site research infrastructure spanning across the western part of the European Alps. This distributed sensing infrastructure is today serving both campaign-based experimental permafrost research as well as long-term monitoring and natural hazard warning needs. Driven by the main property of the core building blocks, namely miniature, battery powered wireless sensors, online and open data access was part of the concept from day one. The focus on a powerful, standardized and open data infrastructure has proven to be very successful enabling a multitude of data access needs simultaneously based on one common data infrastructure. A web frontend allows users to access both primary sensor data as well as data products incorporating metadata or value conversions, i.e. from machine to SI units as well as processed data products. All data on the PermaSense data repository can be accessed in real-time streaming formats as well as using full historic data digests. It was initially suggested that the data analysis would benefit significantly from low-latency online stream processing. This vision has however only been exploited once in the case of post-processing of RAW L1-GPS in the form of a research-grade prototype real-time kinematic processing. The other parts of the data collected by our online sensor systems is today still analyzed using traditional batch data processing. A benefit that the data streaming interface on a publicly accessible web platform does provide is that all interested parties can observe the evolution of a specific experiment in real-time and at their own pace without subscribing to and waiting for e.g. monthly data digests. Therefor this kind of online and open access experimentation facilitates to plan ahead, discuss actions and intervene at a much earlier time than with traditional data logging or closed data gathering systems.

The core functionality is wireless access to all sensors. This wireless access can be either realized using custom, ultra-low power wireless sensors for e.g. thermistors and crackmeters or commodity solutions like 802.11x Wireless LAN that can transfer much higher data rates e.g. from webcams or InSAR radar. We have gained very good experience using Wireless LAN as a high throughput data backbone on a valley scale and serving multiple field sites across the Mattertal Valley in Switzerland or the Aig. Du Midi, Chamonix, France. However, connectivity can fail intermittently and without notice. Additionally, explicit duty-cycling of devices, e.g. the wireless radio, while the sensor is kept operating is a common means for the reduction of power consumption. Therefore, all wireless devices require a storage layer to deal with disconnected operation, handle and synchronize delayed data. On the server side a scalable data management solution is required that on the one hand meticulously preserves all primary data in its original state and on the other hand allows to grow according to the user's needs and/or the evolution of new technologies and analysis methods. We have therefore partitioned the data management into an instance termed "private" acting as data ingress point and holding all incoming primary data in an original, unmodified state. A successive instance termed "public" is a structurally identically system with the added functionality of metadata mapping, value conversion. While the "private" server can be operated in a protected, firewalled environment ensuring long-term data consistency the "public" server also serves a public web frontend and data warehouse for different data aggregates. Upon incorporation of a new data analysis or the change of a data conversion function, metadata etc. data on the "public" server can be replayed, meaning data is deleted, fetched again from the "private" server and the updated data are computed and stored again on the "public" instance.

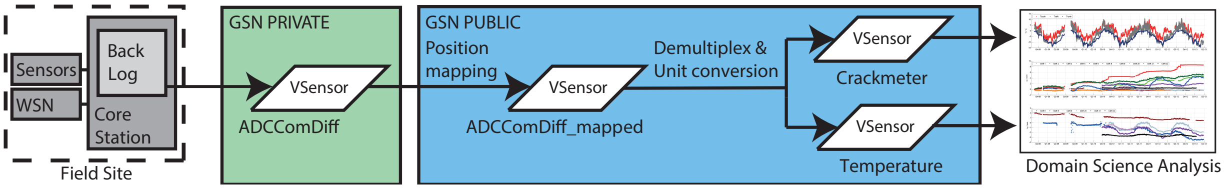

Figure 1: Data from the field sites is stored in unmodified form in the primary instance of the GSN data management server. After mapping of metadata and value conversion it is stored as secondary data in the public instance of GSN from where it is accessible by various tools.

We have gathered and maintained over 5,835,038,653 of data amounting to 1,352.6 GB of data and originating from over 8 field sites over the past ten years. The main users of the data are scientists although there is an increase of users from the government, local authorities, businesses and contractors w.r.t. natural hazard mitigation. While the science users are often searching for the unknown, frequently developing an analysis method while doing so and as such require frequent full access to all data, natural hazard experts prefer consolidated views and reports with a clear structure, low-latency and high update frequency. Hiding complexity on the one side while exposing all possibilities on the other is not possible. Therefore, the frontend data services provided are in a gradual but constant process of change, addition and refinement while the core of the data management system is remaining largely stable. The PermaSense data repository is accessible at http://data.permasense.ch.

Andreas Hasler, Igor Talzi, Jan Beutel, Christian Tschudin and Stephan Gruber: Wireless Sensor Networks in Permafrost Research – Concept, Requirements, Implementation and Challenges

Proc. 9th International Conference on Permafrost (NICOP), June 2008.

Jan Beutel, Stephan Gruber, Andreas Hasler, Roman Lim, Andreas Meier, Christian Plessl, Igor Talzi, Lothar Thiele, Christian Tschudin, Matthias Woehrle and Mustafa Yuecel: PermaDAQ: A Scientific Instrument for Precision Sensing and Data Recovery under Extreme Conditions. Proceedings of the 8^{th} International Conference on Information Processing in Sensor Networks (IPSN), p. 265-276, April 2009.

Matthias Keller, Jan Beutel, Olga Saukh and Lothar Thiele: Visualizing Large Sensor Network Data Sets in Space and Time with Vizzly. Proceedings of the Seventh IEEE International Workshop on Practical Issues in Building Sensor Network Applications (SenseApp 2012), Clearwater, FL, USA, p. 963-971, October 2012.