
PermaSense Data Management

System documentation and tutorial for online data access

April 9, 2014

Samuel Weber, Jan Beutel, Christoph Walser



THE PERMASENSE SYSTEM

This section gives an overview about:

- The PermaSense system architecture
- GSN-based data management:
 - data flow
 - servers and virtual sensors
 - web user interface
 - timing information

PermaSense System Architecture

Sensor Node (SN)

- Collects data from different sensor options

Wireless Sensor Network

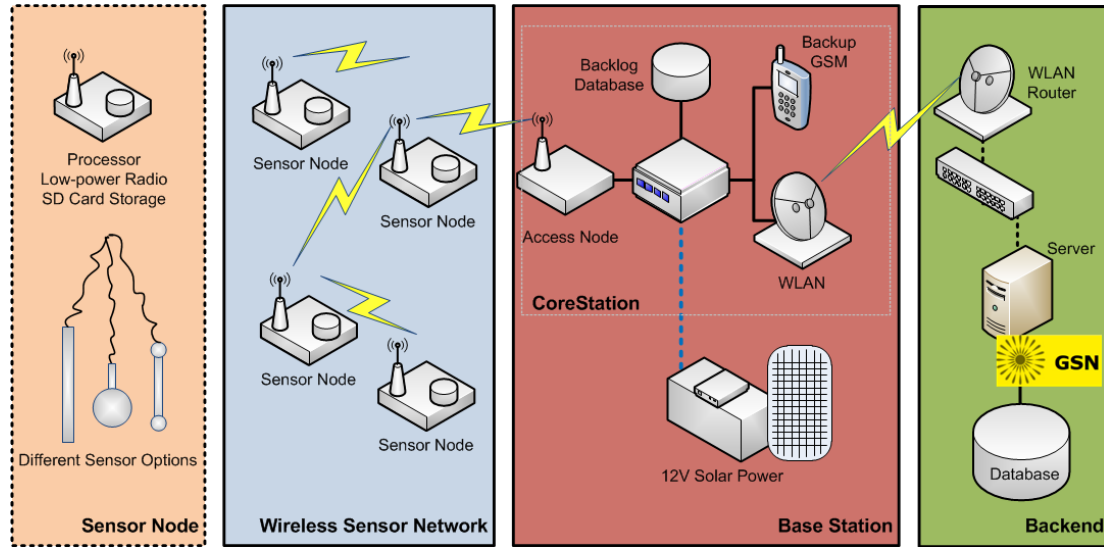
- Forwards the collected data over a 868 MHz wireless communication channel to the base station

Base station

- Sinks all data and forwards it to a central server over an IP network

Backend- and GSN-Server

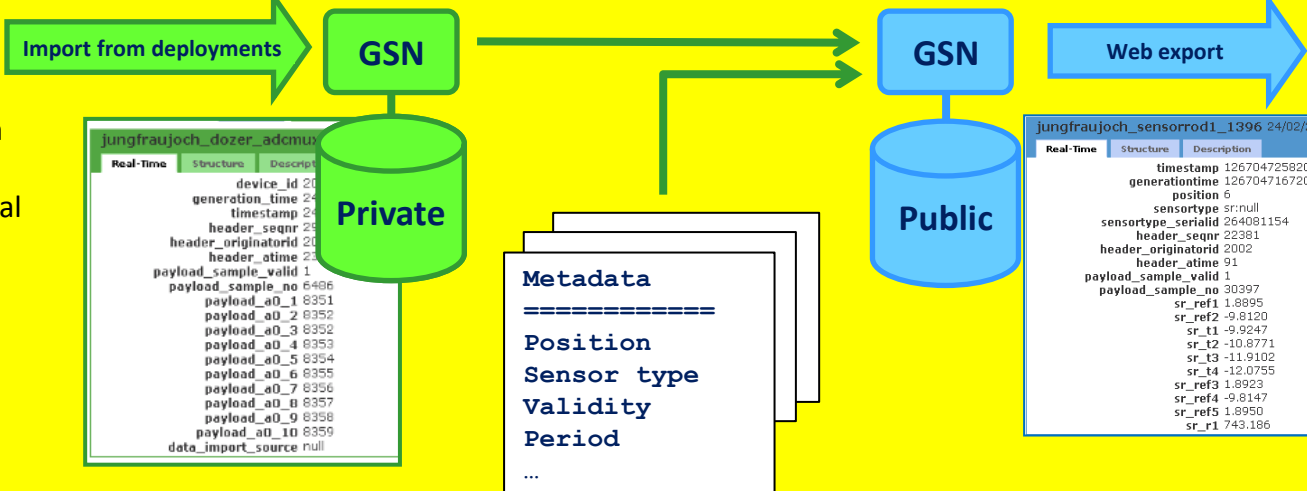
- Collects data from all deployments and implements a number of management and monitoring services



PermaSense GSN Architecture: Hierarchical Online Data Processing

Global Sensor Network (GSN)

- Data streaming framework from EPFL
- Organized in “virtual sensors”, i.e. data types/semantics
- Hierarchies and concatenation of virtual sensors enable on-line processing
- Dual architecture translates data from machine representation to SI values, adds metadata



Real-Time	Structure	Description
	device_id	20
	generation_time	24
	timestamp	24
	header_seqnr	29
	header_originatorid	20
	header_atime	23
	payload_sample_valid	1
	payload_sample_no	6406
	payload_a0_1	8351
	payload_a0_2	8352
	payload_a0_3	8352
	payload_a0_4	8355
	payload_a0_5	8354
	payload_a0_6	8355
	payload_a0_7	8356
	payload_a0_8	8357
	payload_a0_9	8358
	payload_a0_10	8359
	data_import_source	null

Metadata

=====

Position

Sensor type

Validity

Period

...

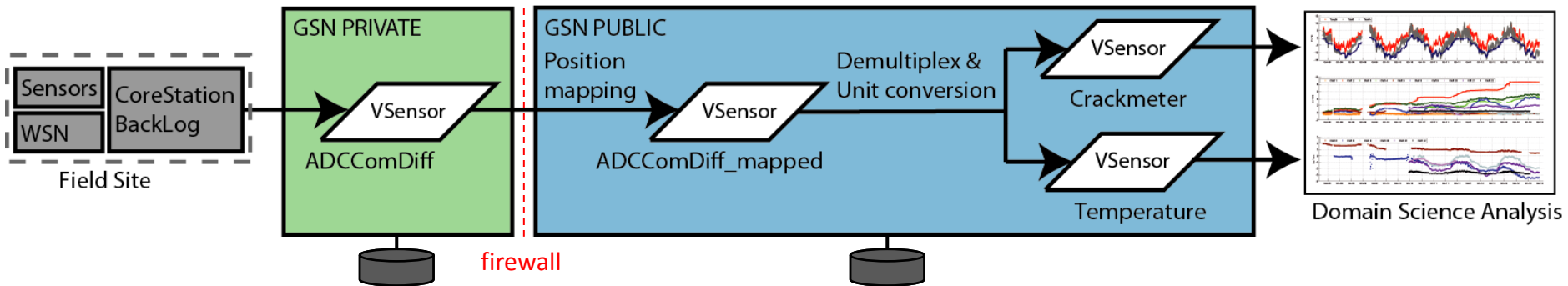
Real-Time	Structure	Description
	timestamp	1267047258207
	generationtime	1267047167207
	position	6
	sensortype	sr:null
	sensortype_serialid	264081154
	header_seqnr	22381
	header_originatorid	2002
	header_atime	91
	payload_sample_valid	1
	payload_sample_no	30397
	sr_ref1	1.8895
	sr_ref2	-9.8120
	sr_t1	-9.9247
	sr_t2	-10.8771
	sr_t3	-11.9102
	sr_t4	-12.0755
	sr_ref3	1.8923
	sr_ref4	-9.8147
	sr_ref5	1.8950
	sr_r1	743.186

GSN Server Data Flow

Data from field site is received by the private GSN server “as is” and stored in a primary database.

Data is passed on to a public GSN server where it is mapped to positions, sensor types and converted to convenient data formats.

Data is downloaded, analyzed and plotted using external tools.



- All data from sensor nodes to backend is transmitted in packets
The length of the sensor network data packets is limited to 23 bytes. In cases where sensor data is larger, several consecutive packets are generated.
- Data is received and sorted according to data types

Structure of a Virtual Sensor

- Data in database (GSN) is organized as virtual sensors (VS) per deployment
If there are multiple sensors yielding the same data types, this data is multiplexed into the same VS.
- Each VS has a unique **name**: *<deployment>_<sensor type>_<processing step>*
Processing steps: *raw* → *mapped* → *converted*

- Each VS has **three tabs** with different information

- Each VS contains a **header**. The header for sensor network specific data contains:

- *position*: number of physical locations
- *device_id*: mapped device id at this position
- *timed/generation_time/timestamp*: there are **several time formats** that are described later
- *sensortype*: sensor types and calibration constants connected at a given position
- *sensortype_serialid*:
- *header_seqnr*: sequence number denoting successively generated sensor network packets
- *header_originatorid*: the same as device id
- *header_atime*: packet transfer time in seconds, used to calculate generation time
- *payload_sample_valid*: flag which specifies data acquisition errors
- *payload_sample_no*: sample number denoting packets originating from same data sampling period (typ. period is 120 sec)

```
malterhorn dozer adcmux1_mapped 05/06/2013 15:49:29.032 CEST timed
Real-Time Structure Description
position 2
device_id 2114
generation_time 05/06/2013 15:49:11.919 CEST
timestamp 05/06/2013 15:49:28.919 CEST
sensortype nctt:1 tc:null sht11:1
sensortype_serialid 264310762
header_seqnr 53435
header_originatorid 2114
header_atime 17s
payload_sample_valid 1
payload_sample_no 11914
payload_a0_1 31887
payload_a0_2 40504
payload_a0_3 32342
payload_a0_4 33408
payload_a0_5 5923
payload_a0_6 6814
payload_a0_7 31886
payload_a0_8 40504
payload_a0_9 27207
payload_a0_10 3556
data_import_source null
```

GSN Web User Interface

<http://data.permasense.ch>

PermaSense :: GSN - Public

- HOME
- DATA
- NETWORK TOPOLOGY
- LOGS
- BACKEND
- SENSOR NETWORK
- SCIENCE
- ON-SITE WEATHER
- PERMASENSE HOME

HOME

All virtual sensors are listed and sorted by deployment, data type, sensor type and process step

DATA

Manual download of data

NETWORK TOPOLOGY

Topology and table with network information sorted by deployment

LOGS

Log files from core stations

BACKEND

Backend, MySQL and Zabbix monitoring server specific plots

SENSOR NETWORK

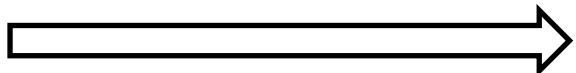
Sensor network specific plots

SCIENCE

Collection of scientific plots

ON-SITE WEATHER

Weather information and webcams



The screenshot displays the 'NETWORK TOPOLOGY' view of the PermaSense web interface. At the top, there are tabs for different locations: ADM, DIRRUHORN, MATTERHORN, JUNGFRAUJOCH, THUR, and GLACIERS. Below the tabs, a network topology map shows various sensor nodes (represented by green circles with IDs) connected by lines. The nodes are distributed across the map, with some clusters. Below the map, a table lists sensor data for 31 positions. The table columns include Position, ID, Pkt Count, Vsjs, VsDs, Temp, Hum, Flash Count, Uptime, Generation Time, Timestamp, and Packetrate. The data is sorted by Position (0 to 31). The table shows various sensor readings and status indicators (green bars for Vsjs and VsDs).

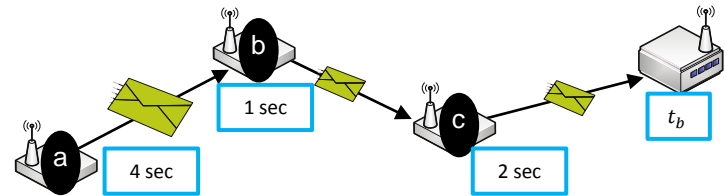
Position	ID	Pkt Count	Vsjs	VsDs	Temp	Hum	Flash Count	Uptime	Generation Time	Timestamp	Packetrate		
0	7	412519	13.32	13.32	25.18	25.83	0	4488171	2013-06-06 11:22:19 +2:00	2013-06-06 11:22:19 +2:00	0.00		
1	2088	520276	3.63	11.63	0.49	15.15	60.52	0	9324506	2013-06-06 11:21:06 +2:00	2013-06-06 11:21:36 +2:00	0.00	
2	2114	1261864	3.63	11.49	0.60	13.52	54.18	0	8357055	2013-06-06 11:21:45 +2:00	2013-06-06 11:22:06 +2:00	0.00	
3	2074	521098	3.62	11.57	0.47	-1.59	74.17	0	10965970	2013-06-06 11:19:59 +2:00	2013-06-06 11:20:26 +2:00	0.00	
4	2055	521685	3.62	11.55	0.50	-5.24	41.67	0	11037142	2013-06-06 11:19:45 +2:00	2013-06-06 11:20:24 +2:00	0.00	
5	2053	605594	3.59	11.55	2.85	8.75	54.47	0	12779185	2013-06-06 11:20:18 +2:00	2013-06-06 11:20:40 +2:00	0.00	
6	2055	346941	3.63	11.49	0.83	10.23	54.27	0	12779315	2013-06-06 11:21:21 +2:00	2013-06-06 11:21:52 +2:00	0.00	
7	2095	607375	3.53	11.54	3.04	3.06	52.15	0	12776383	2013-06-06 11:20:43 +2:00	2013-06-06 11:21:25 +2:00	0.00	
8	2093	521052	3.62	11.58	0.77	5.21	52.78	0	12776368	2013-06-06 11:19:52 +2:00	2013-06-06 11:20:37 +2:00	0.00	
9	2055	347205	3.64	11.49	0.32	10.05	65.66	0	3296156	2013-06-06 11:21:16 +2:00	2013-06-06 11:21:56 +2:00	0.00	
10	2104	521445	3.63	11.51	0.81	8.03	40.61	0	13975952	2013-06-06 11:21:11 +2:00	2013-06-06 11:22:02 +2:00	0.00	
13	2059	519516	3.61	11.45	0.24	9.01	74.97	0	10963985	2013-06-06 11:21:35 +2:00	2013-06-06 11:21:56 +2:00	0.00	
14	2020	260977	3.65	13.91	1.96	20.89	45.80	0	8969828	2013-06-06 11:20:38 +2:00	2013-06-06 11:21:27 +2:00	0.00	
17	2099	260771	3.59	11.57	0.23	7.17	70.64	0	10963630	2013-06-06 11:16:17 +2:00	2013-06-06 11:22:06 +2:00	0.00	
19	5015	963621	13.99	13.99	25.20	20.55	0	789280	2013-06-06 11:20:37 +2:00	2013-06-06 11:21:23 +2:00	0.00		
20	2115	1263721	3.61	11.47	0.36	-3.90	64.56	0	8357059	2013-06-06 11:21:39 +2:00	2013-06-06 11:21:53 +2:00	0.00	
21	2094	346506	3.64	11.56	0.36	7.38	76.57	0	10963980	2013-06-06 11:19:26 +2:00	2013-06-06 11:20:41 +2:00	0.00	
23	6000	347413	3.68	13.86	6.63	49.89	0	6761593	2013-06-06 11:20:55 +2:00	2013-06-06 11:21:32 +2:00	0.00		
24	2052	261434	3.65	11.58	0.21	7.84	67.31	0	3298264	2013-06-06 11:19:45 +2:00	2013-06-06 11:20:27 +2:00	0.00	
25	2127	521292	3.63	13.51	0.36	11.28	42.29	0	8357996	2013-06-06 11:21:27 +2:00	2013-06-06 11:21:57 +2:00	0.00	
26	6001	347752	3.65	13.60	11.13	25.13	0	12192288	2013-06-06 11:20:38 +2:00	2013-06-06 11:21:28 +2:00	0.00		
30	2081	522122	3.61	11.60	0.27	1.04	60.41	0	7571311	2013-06-06 11:20:15 +2:00	2013-06-06 11:21:29 +2:00	0.00	
31	2025	261567	3.61	11.60	0.01	1.68	1.16	74.37	0	12790674	2013-06-06 11:21:11 +2:00	2013-06-06 11:21:54 +2:00	0.00

GSN Multiple Time Bases

- Global reference time (UTC) is often not available

→ Solution: Elapsed time on arrival

- Sensor nodes measure/accumulate sojourn time
- Base station annotates data with arrival time (e.g. UTC)
- Generation time is calculated as difference between arrival time and sojourn time



$$\tilde{t}_b = t_b - \tilde{t}_b = 2013/06/06\ 17:47:11 - (4+1+2) = 2013/06/06\ 17:47:04$$

- All data carries multiple timestamps

- *generation_time* depicts time when data is sampled
- *timestamp* denotes the time when data reaches a UTC synchronized time base for the first time (e.g. base station)
- *timed* is database time, i.e. the time data is inserted into the database

- GSN time

- Is in UTC (Switzerland: winter +1h (CET), summer +2h (CEST))
- Is unix-time in millisecond!
 - *Unix time stamp* is merely the number of *seconds* between a particular date and the Unix Epoch (January 1, 1970).
 - *GSN time stamp* is merely the number of *milliseconds* between a particular date and the Unix Epoch.

- Example for time conversion in Matlab:

```
function mt=time_gsn2matlab(gt)
% converts times in gsn format to time in matlab datenum
% format, UTC time GMT+2
mt=datenum(1970,1,1) +gt./(3600*24*1000) +datenum(0,0,0,2,0,0);
return;
```



Metadata for Deployments

- PermaSense maintains four deployments, Matterhorn (MH), Jungfrauoch (JJ), Dirruhorn (DH), Aiguille du Midi (ADM)
- Metadata for each deployment is described in the *nodeposition.xls* file
 - **Name** of the deployment
 - **Position**: physical location of a sensor type (shown in an additional map or overview picture)
 - **Sensor type** used at the position
 - **Coordinates** of the position
 - **Device ID**: unique ID identifying a piece of hardware

Deployment Overview: Aiguille Du Midi											Date	31.05.2013	06.05.2013	08.07.2012
		green = new installation		radiochannel = 12							# Devices	18	17	12
Position	Label	Name	Comment	Host Name	Sensor ID	Sensor Type	CR Length	GPS X	GPS Y	GPS Alt				
0		Base Station		permasense-caa-adm-1				6.887489	45.879446	3750		29	29	29
1		West Face Piton Central Top Crack			SN 10270716	NCTN	100	6.88762	45.878993	3750		2121	2121	2121
2		West Face Piton Central Bottom Crack				NCNN	150	6.887391	45.879083	3750		2105	2105	2105
3		West Face Piton Nord				NCNN	200	6.886823	45.879243	3750		2110	2110	2110
4		South Face Piton Nord				NCTN	150	6.886902	45.879011	3750		2109	2109	2109
5		East Face Piton Nord				NCNN	100	6.887519	45.87931	3750		2108	2108	2108
6		North Face Piton Nord				NCTN	200	6.887177	45.879532	3750		2111	2111	2111
7		Borehole South				CR1000		6.887523	45.878458	3750		2131		
8		Borehole North				CR1000		6.886813	45.878913	3750		2130	2130	
9		BoreHole East				CR1000		6.888173	45.878611	3754		2129	2129	
10		Relay South Piton Central						6.887625	45.878244	3741		2120	2120	2120
11		Argentière Glacier Webcam				Webcam						6000	6000	6000
12		Aiguille du Midi Webcam				Webcam						6001	6001	6001
13		Arrete Aiguille du Midi Webcam				Webcam						6002	6002	6002
14		Mont Blanc Webcam				Webcam						6003	6003	6003
15		Relay North Top Piton Central						6.887902	45.878943	3825		2066	2066	
16		Relay West Terrace Piton Nord						6.886505	45.878946	3723		2054	2054	
17		Relay Cable Car						6.887404	45.879565	3710		2057	2057	

- If an additional sensor gets installed, a new position is created
- If a sensor gets replaced, a new data column is added with the new device ID



DATA ACCESS

This sections gives an overview about data access using external tools.

Accessing Data from the GSN Server

- There are two different methods to get data from GSN using *http queries*:
 1. **One-shot query approach**: each request returns data based on the database state at the time of the query. This allows one to quickly obtain aggregated data from a virtual sensor and export this data in convenient formats, e.g. CSV or XML.
 2. **Streaming approach**: requested data is continuously streamed to the user in real-time until the connection to the server is closed.
- Example for a simple *one-shot query* without aggregation/conditions
 - Suppose, you want to query all fields of the *matterhorn_crackmeter__tctc* virtual sensor (2 crackmeters, 2 thermistors) between 25/08/2012 and 13/06/2013 (UTC):
 1. Open in browser:
[http://data.permasense.ch/multidata?vs\[0\]=matterhorn_crackmeter_tctc&time_format=iso&field\[0\]=All&from=25/08/2012+00:00:00&to=13/06/2013+00:00:00](http://data.permasense.ch/multidata?vs[0]=matterhorn_crackmeter_tctc&time_format=iso&field[0]=All&from=25/08/2012+00:00:00&to=13/06/2013+00:00:00)

where: vs[0]: name of virtual sensor
time_format: time format of returned data
field[0]: list of data fields to return
from, to: time limits of data request in UTC

For further options and syntax information please consult:
<https://github.com/LSIR/gsn/wiki/Web-Interface#multidata>
 2. You will get a CSV-formatted file with the requested data.
 - A complete example using Matlab can be downloaded



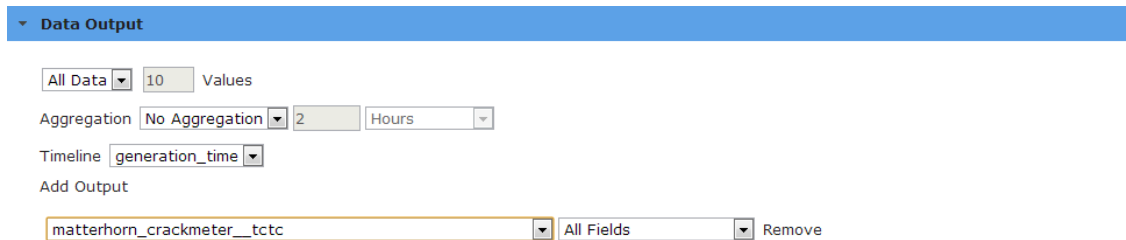
Accessing Data over the *data.permasense.ch* Web Interface

<http://data.permasense.ch>

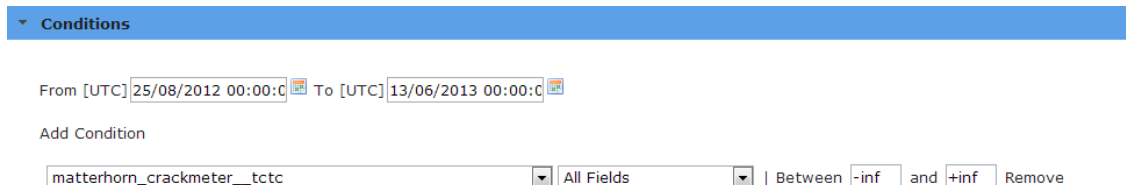
- Manual access to all virtual sensors and data fields over web interface
- Example (same as for [one-shot query](#)):
 - Get all fields of the *matterhorn_crackmeter__tctc* VS between 25/08/2012 and 13/06/2013:

1. Go to <http://data.permasense.ch>, open tab 'DATA'

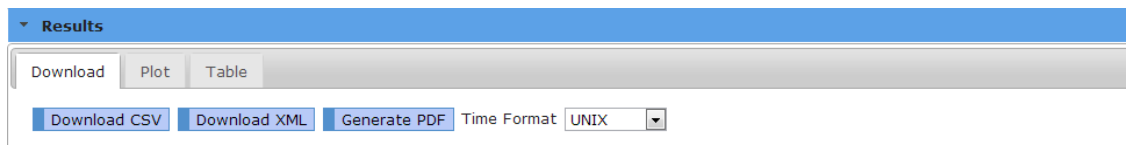
2. In sub-tab 'Data Output', select 'All Data' from the first drop down menu and the according virtual sensor:



3. Click on sub-tab 'Conditions' and set the time limits:



4. Click on sub-tab 'Results' and access the data in one of the offered formats:

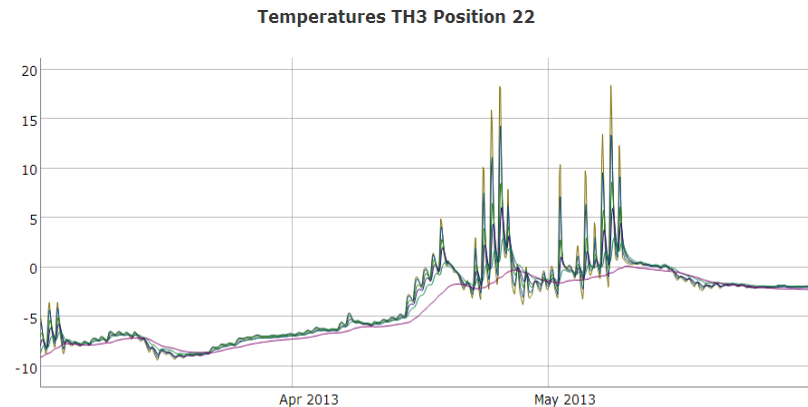
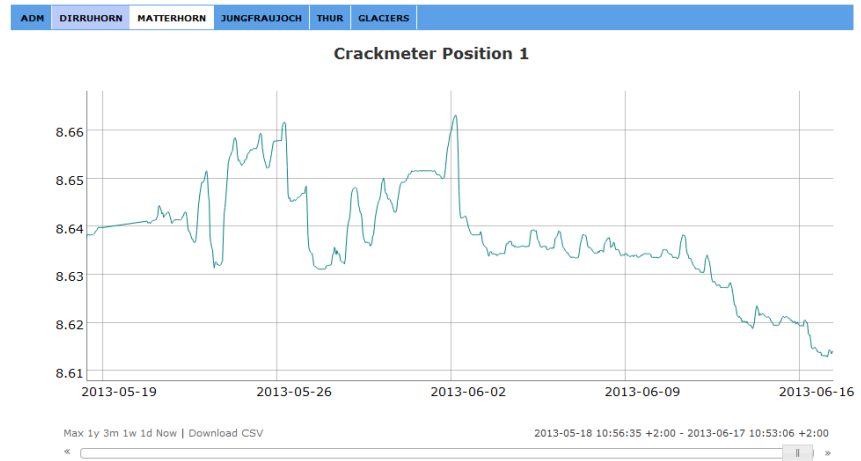


Note: The plot is interactive and can be used to easily preview the resulting data.



Accessing Data over the *data.permasense.ch* Web Interface

- A selection of predefined plots can be found in the tabs *SCIENCE* and *SENSOR NETWORK*
- Plots are generated using the *Vizzly* framework
- All plots are interactive:
 - **Time**: time range to be shown
 - **Pos**: data fields to be plotted
 - **Update**: has to be clicked after ticking
 - **Download CSV** download aggregated of currently shown plot



Selected time: 2013-03-02 14:40:00 +1:00

Position 22 temperature_5cm

Pos 22: -5.04 °C

Position 22 temperature_10cm

Pos 22: -5.79 °C

Position 22 temperature_20cm

Pos 22: -7.06 °C

Position 22 temperature_30cm

Pos 22: -8.01 °C

Position 22 temperature_50cm

Pos 22: -8.71 °C

Position 22 temperature_100cm

Pos 22: -9.05 °C

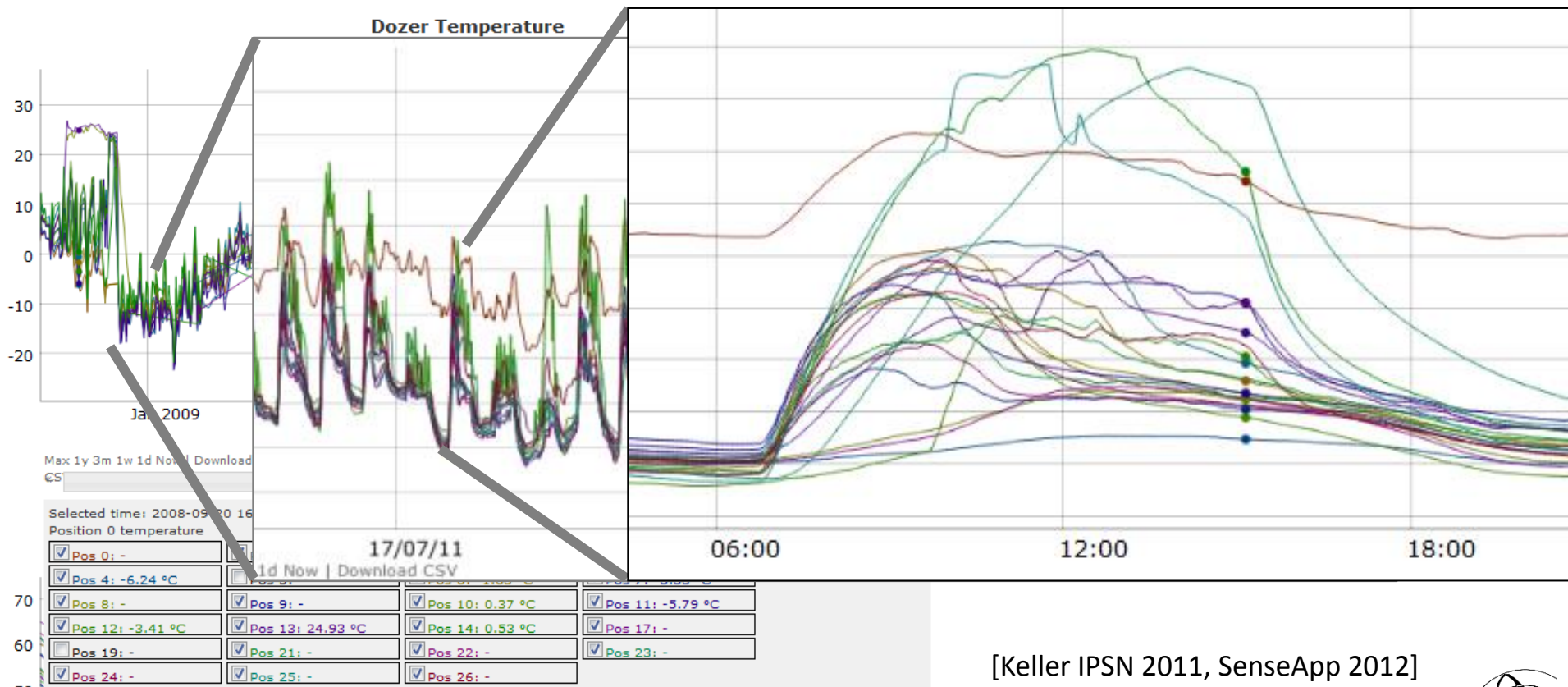


BACKGROUND ON GSN AND SYSTEM ARCHITECTURE

This sections gives an overview about:

Vizzly: Visualization of Large Data

- Fast access to millions of data samples
- Pan, zoom, channel selection
- Combination of historic and real-time data



[Keller IPSN 2011, SenseApp 2012]



WSN On-Node Storage Layer

- On-node flash based storage (SD-Card)
 - Integrated with Dozer queuing mechanism (beacon traces & per-link ack's with backpressure)
 - All generated packets are stored on local flash memory
 - Packets not yet sent are flagged for sending later
 - Bulk access optimized for flash memory (no single packet transfers)
- Enables both delayed sending (disruptions) and post-deployment validation

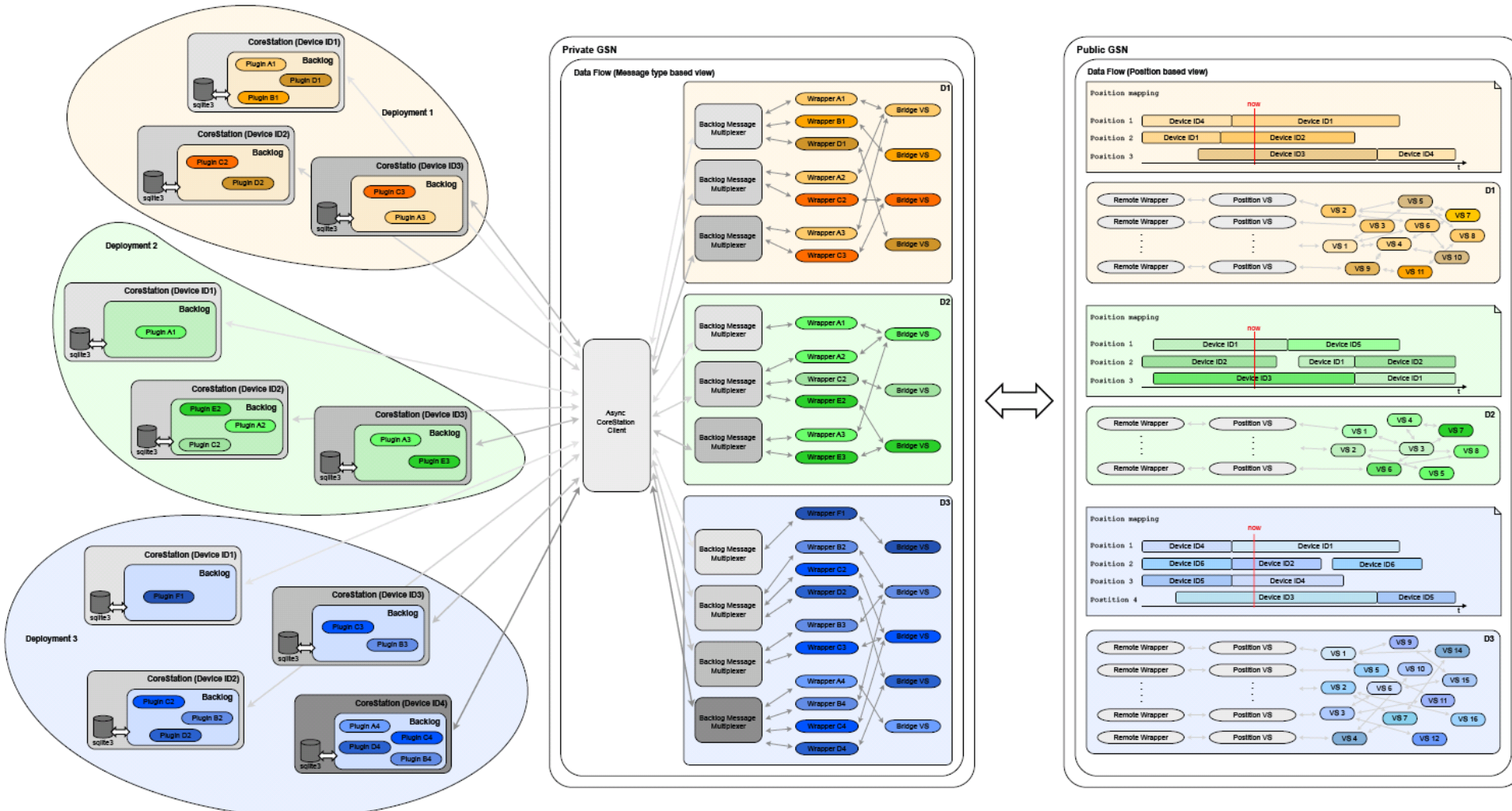


Mitigating Post WSN Data Loss

- BackLog = Auxiliary data aggregation layer at device level
 - Remote storage and synchronization layer for Linux systems
 - Python based, designed for PermaSense CoreStation
 - Plugin architecture for extension to custom data sources
 - Data multiplex from plugin to GSN wrapper over one socket
- Reliable (flow controlled) synchronization
- Schedulable plugin/script execution, remote controlled



Multi-Site, Multi-Station, Multi-Revision Data...



Metadata Mapping Architecture

- Based on 2 GSN instances
 - Separation of load/concern across two machines
 - “Private” GSN instance, raw data, protected, high availability
 - “Public” GSN instance, mapped and converted data, open, non-critical
- Metadata stored in version control system (CSV, SVN)
- Mapping of
 - Positions, coordinates, sensor types, conversion functions, sensor calibration...
- Conversion of
 - Time formats, raw to SI values...
- Replay of metadata/mapping possible, e.g. on errors
- Change management

Metadata Change Management

- Allows simple exchange of sensor hard-/software at runtime
- Post-deployment annotation
 - Stop GSN– deployment change – annotate metadata – restart GSN
- Automatic synchronization with 1 day change boundaries

