Swiss Experiment:
From Wireless Sensor Networks to E-Science

Prof. Karl Aberer
EPFL, Lausanne

Distributed Information Systems Laboratoray and
National Center of Competence in Research on
Mobile Information and Communication Systems
www.mics.org

A research instrument of the
Swiss National Science Foundation

FNSNF
Fonds National Suisse
Schweizerischer Nationalfonds
Fondo Nazionale Svizzero
Swiss National Science Foundation

Overview

1. Motivation and Background
2. SensorScope:
   Out-of-the box wireless sensing
3. Global Sensor Networks:
   Weaving the Sensor Web
4. Swiss Experiment:
   Web 2.0 e-science
5. Conclusions

Motivation and Background

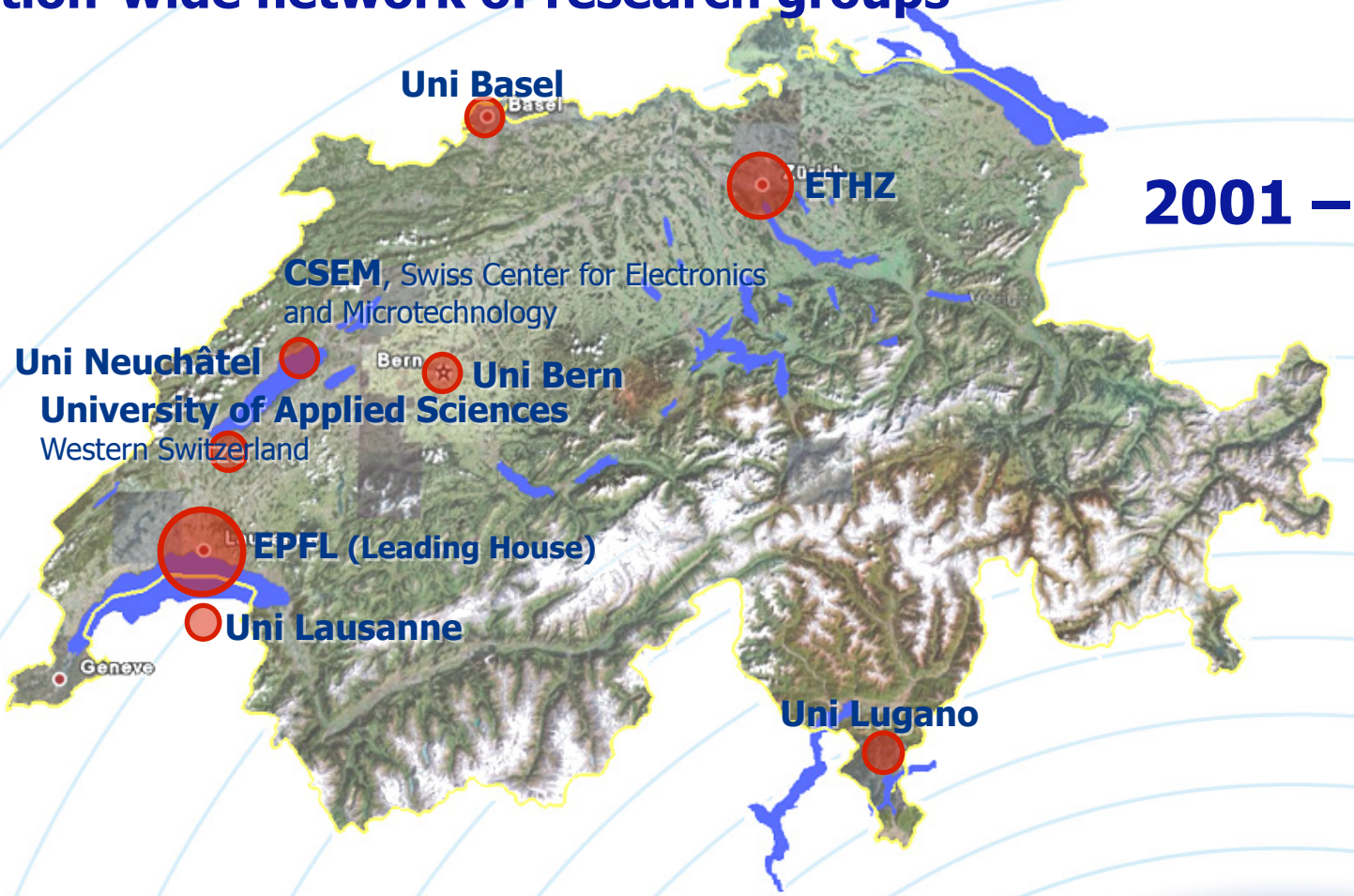
The Promise of Wireless Communication

Bringing the cyberspace to the physical world:
mobile phones, wireless Internet access, …

Bringing the physical world to the Internet:
RFID, wireless sensor networks, …

NCCR MICS: Excellence in Mobile Information and Communication Systems Research

30+ faculty members, 80+ researchers
a nation-wide network of research groups

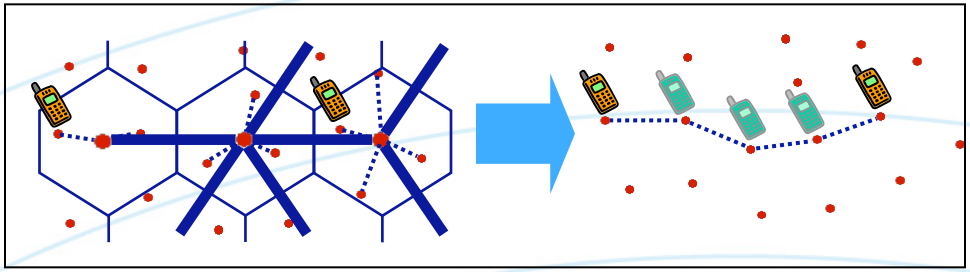


2001 – 2013

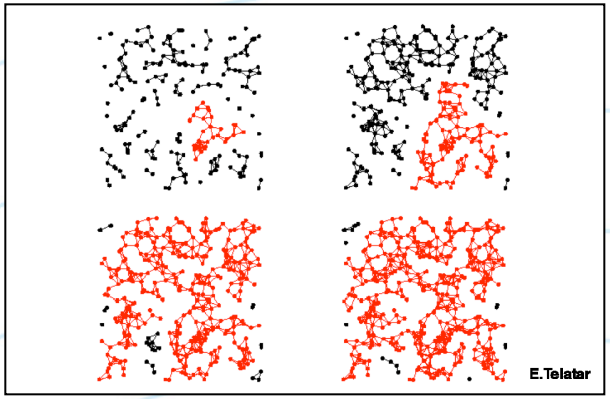
Supported by Swiss SNF

NCCR MICS – An evolving mission

2001
Mobile ad-hoc networks as alternative infrastructure

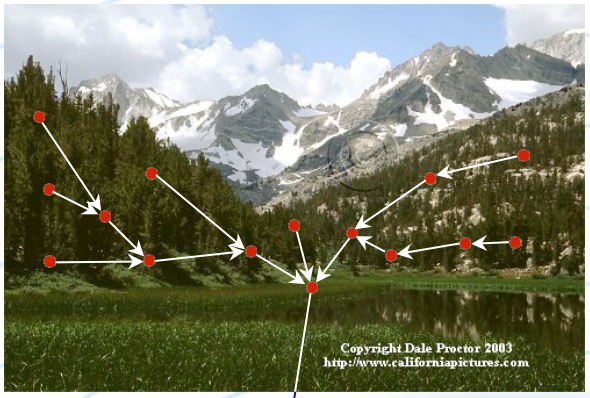


2001 – 2005
Self-organizing Networks: Theory and systems



2005 –

Wireless Sensor Networks for Environmental Monitoring



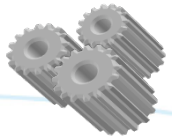
Wireless Sensor Networks

Wireless Sensor Networks

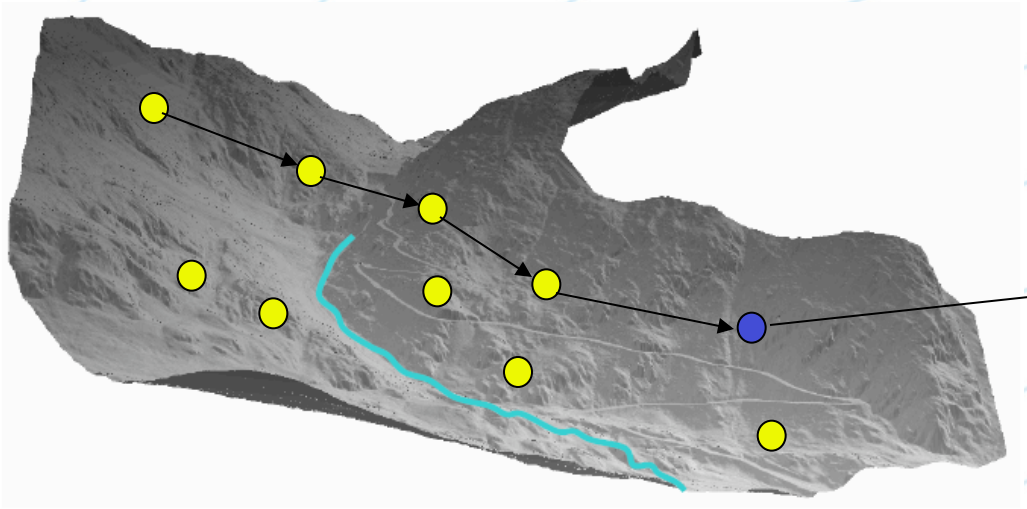
Sensor motes

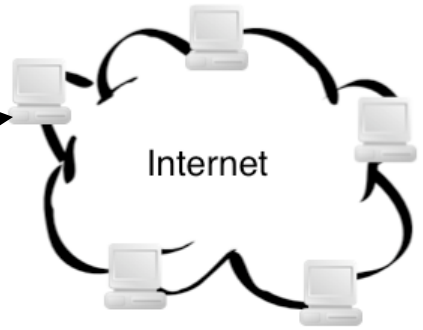
sense

compute



communicate

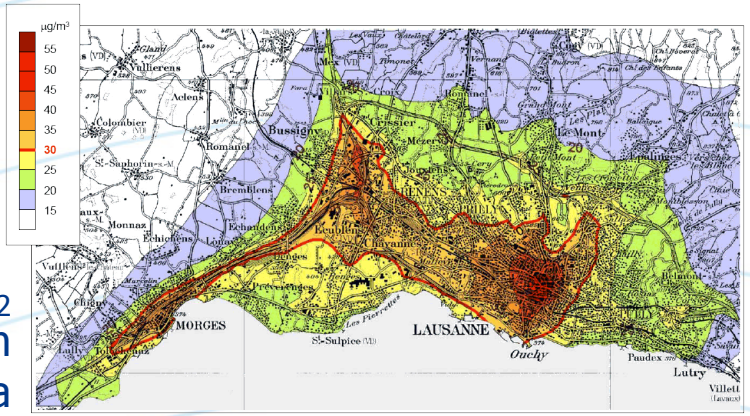




Wireless sensor networks change the way we can monitor our environment!

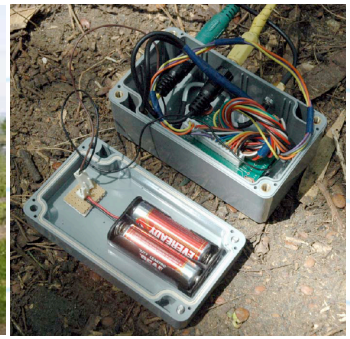
Application Domains

Societal concerns,
e.g. air pollution



Mean yearly NO_2 concentrations in the Lausanne area

Economic concerns,
e.g. agriculture



Water management for arid regions

Scientific concerns,
e.g. environmental engineering

Snow measurement

Environmental Engineering in Mountain Regions

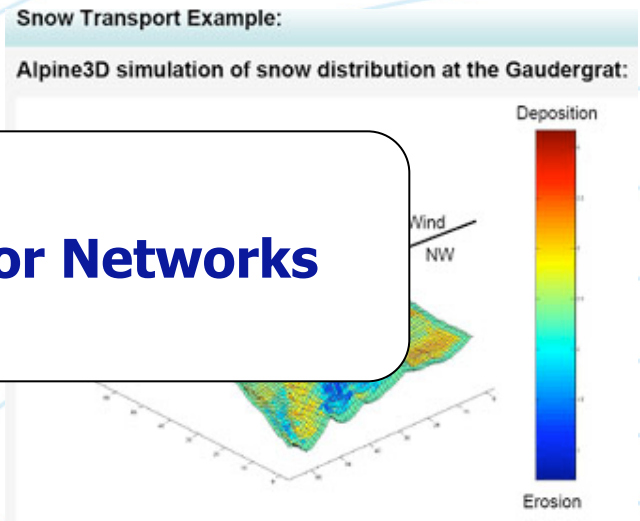
- **Global significance of mountain regions**
  - 10-25% of the human population live in mountain regions
  - 50% of the freshwater resources (water towers)
  - Tourism, biodiversity, natural hazards
- **Vulnerability**
  - Climate change most pronounced in mountain regions
  - Extreme events (flooding, avalanches, landslides)

1906

2003

Modelling Enviromental Processes

- **Today limited capabilities for gathering, analyzing and visualizing data to obtain reliable models**
  - Extreme mountain hydrologic and weather events essential to accuracy of hydrological forecasts
  - Large changes over short distances



Wireless Sensor Networks

SensorScope:
Out-of-the box wireless sensing

Work by: M. Parlange, M. Vetterli, G. Barrenetxea,
O. Couach, M. Krichane, V. Luyet, F. Ingelrest, G. Schaefer

NCCR MICS
National Competence Center In Research
Mobile Information and Communication Systems

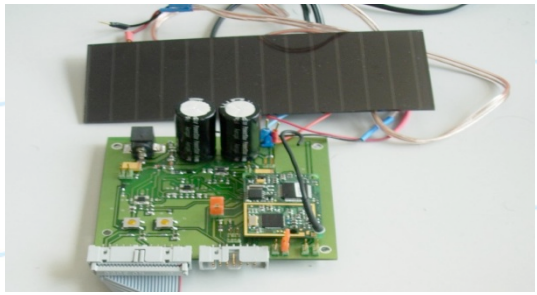
FNSNF
Fonds National Suisse
Schweizerischer Nationalfonds
Fondo Nazionale Svizzero
Swiss National Science Foundation

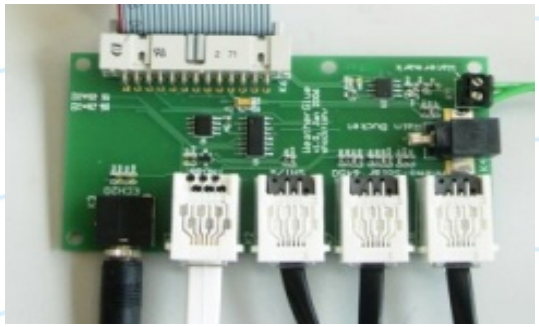
Sensorscope

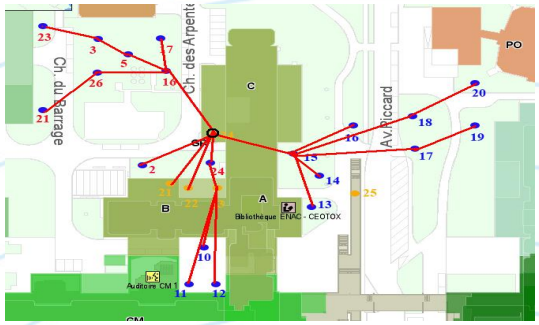
- **State of the art in environmental monitoring**
  - Expensive instruments and data loggers
- **Replace by wireless sensor networks**
  - low cost, real-time measurement, high resolution
- **Collaboration between hardware/software engineers and environmental scientists**
  - gather environmental data for modeling the energy exchanges at the earth-atmosphere boundary

SensorScope System Components

- **Hardware: data acquisition module**
  - Communication and processing unit
    - Solar powered
    - Low power consumption
  - Sensors and interface board
- **Software: networking**
  - Multiple sink topology
  - Over-the-air reprogramming
  - Reliable data collection
    - packet combining
    - routing without routes







SensorScope Weather Station

- **Properties**
  - Easy to deploy (plug and play)
  - Low-cost (< 500$ each)
  - Autonomous
  - Flexible configuration

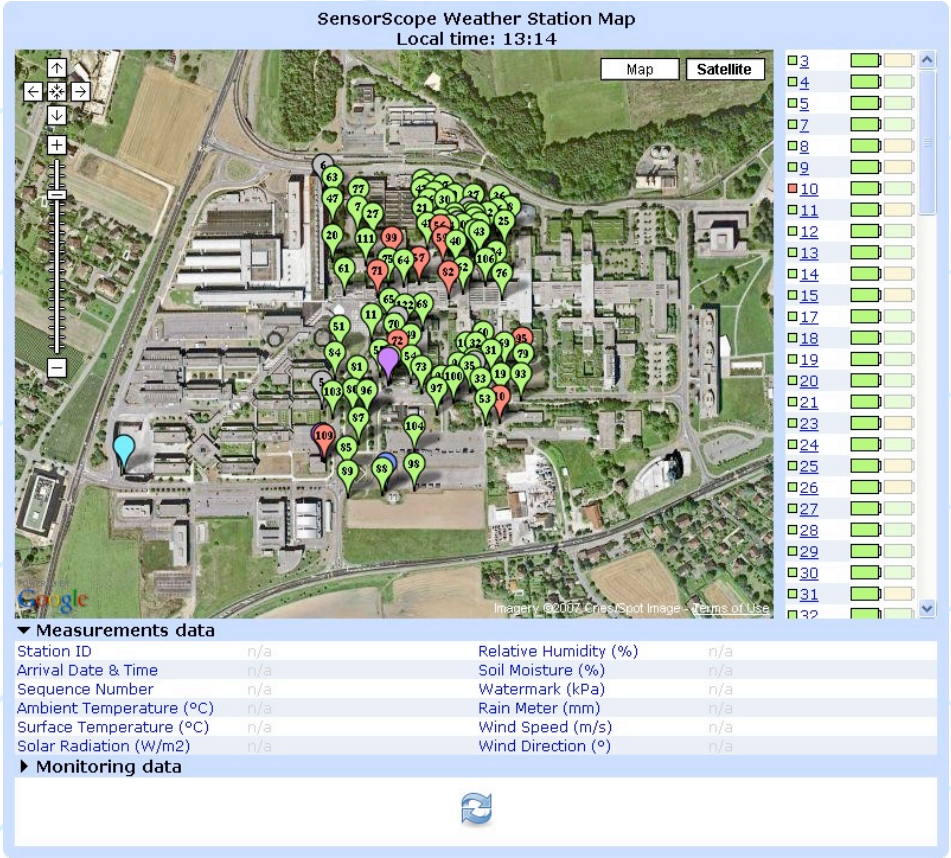
- **Measurements**
  - Ambient temperature and humidity
  - Surface temperature
  - Wind speed and direction
  - Soil moisture
  - Soil water pressure
  - Solar radiation
  - Precipitation

- **One regular station = 100 SensorScope weather stations**

SensorScope Front-End

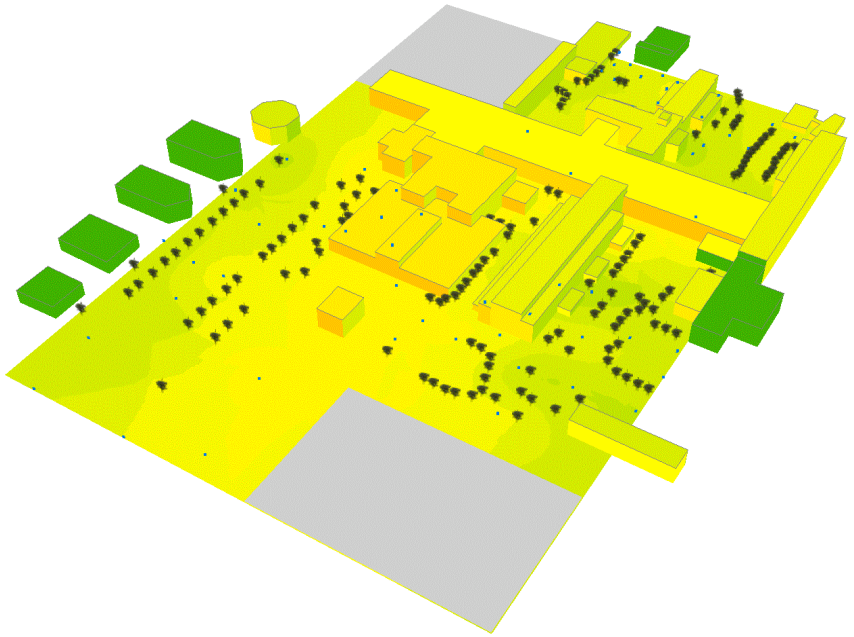
- **Properties**
- Data available online in real-time
- User-friendly interface
- User can plot/download data

- **Administration**
- Station activity monitoring
- Battery level monitoring
- Change settings
  - Location
  - Installed sensors
  - Photo

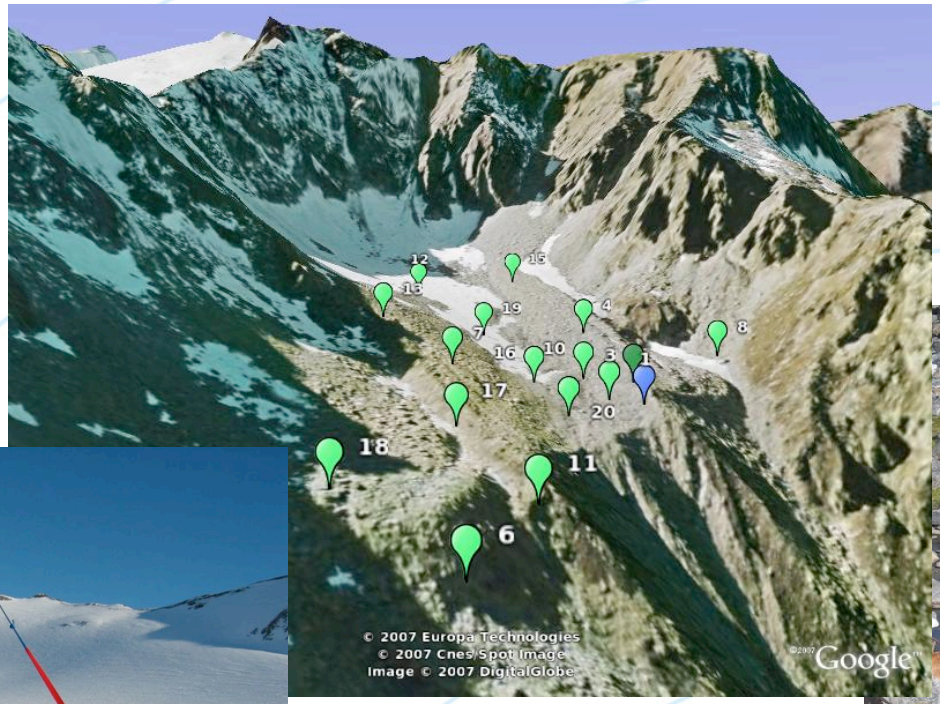


Deployments – EPFL Campus

- **LUCE deployment (July 2006 – May 2007)**
  - Lausanne Urban Canopy Experiment
  - 110 stations measure every 30 s to achieve high temporal and spatial density measures
- **Objectives**
  - Study urban micro-meteorology on the EPFL campus
  - Validate simulations of air flow over the campus

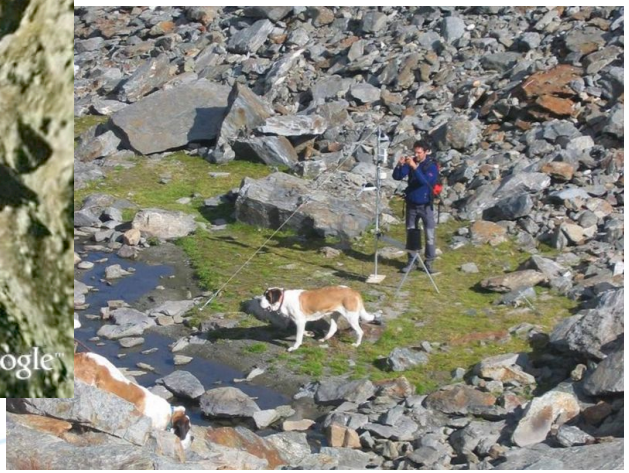


Deployments – Alps



Le Genepi

Plaine Morte

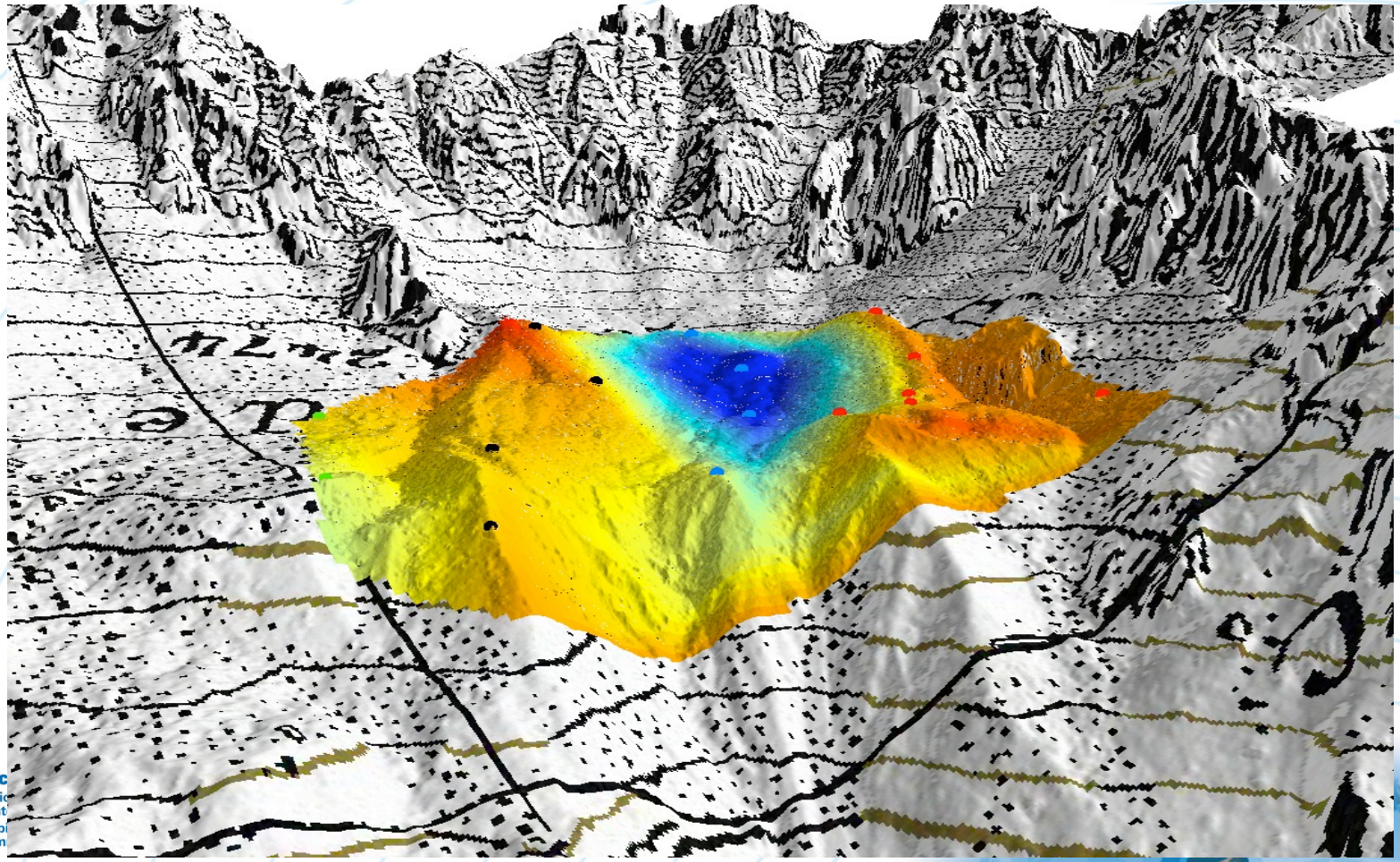


St. Bernard

Sample results

3d Digital Elevation Model and air temperature distribution [°C]

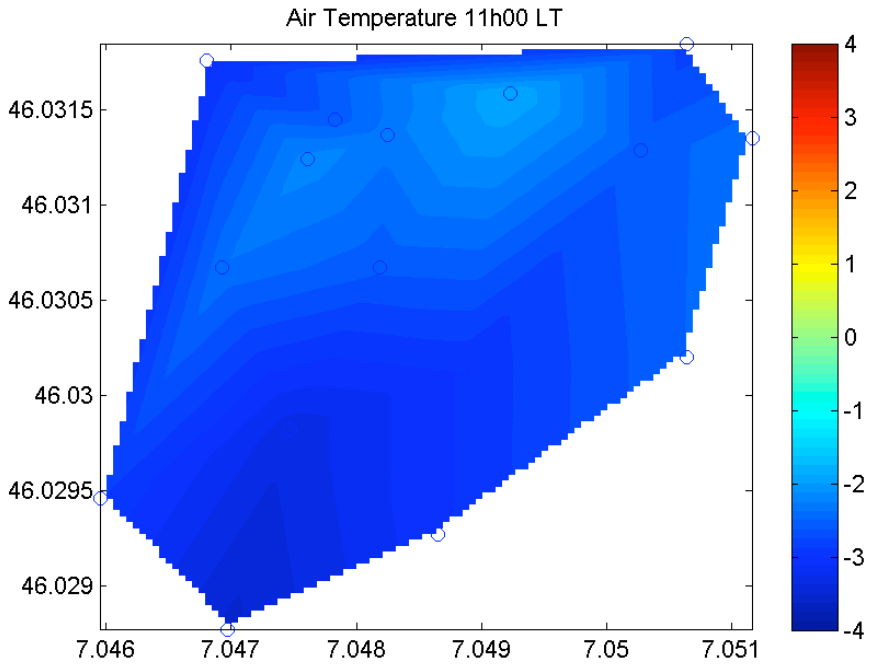
Period 22 to 23/10/2007

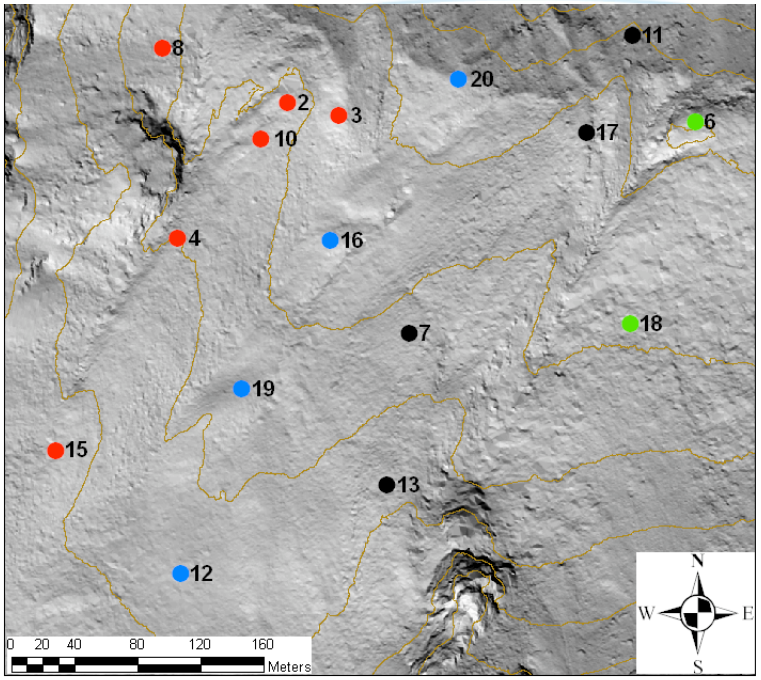


Sample results

Spatial air temperature distribution [°C] over the Génépi rock glacier

Period 28 to 30/10/2007





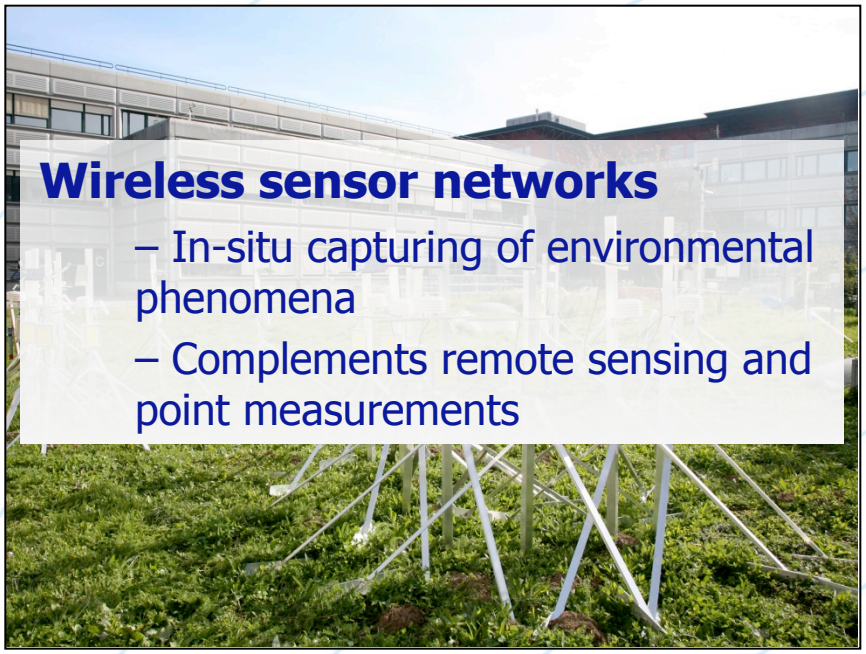
Experience

- Happy scientists
  - New data, new insights
  - More deployments, more instruments, more data
  - Faster understanding the impact of climate change

Swiss Experiment

www.swiss-experiment.ch

Environmental Science = Measuring + Modelling

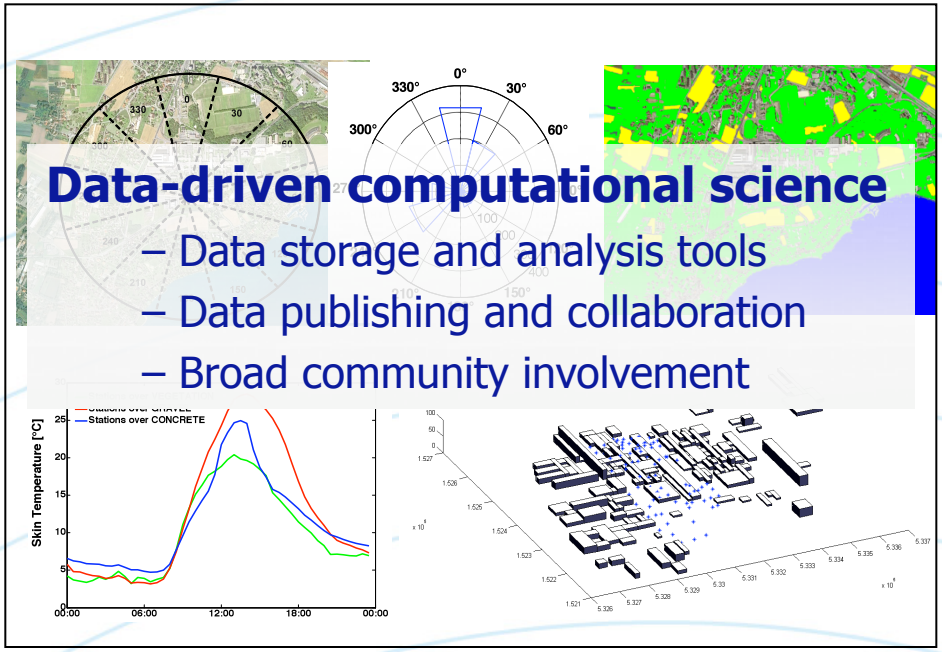


Wireless sensor networks

– In-situ capturing of environmental phenomena

– Complements remote sensing and point measurements

+



Data-driven computational science

– Data storage and analysis tools

– Data publishing and collaboration

– Broad community involvement

Swiss Experiment – the idea

- **Environmental computational science**
  - –Real-time capturing and control of environmental monitoring experiments
  - –Collaborative capturing, sharing and analysis of real-time sensor data

Global Sensor Networks:
Weaving the Sensor Web

Work by: K. Aberer, M. Hauswirth, S. Michel, A. Salehi, Y. Zhou, A. Ouksel, O. Jurca

NCCR MICS
National Competence Center In Research
Mobile Information and Communication Systems

FNSNF
Fonds National Suisse
Schweizerischer Nationalfonds
Fondo Nazionale Svizzero
Swiss National Science Foundation

Data Management Challenges

Similar requirements
- Data browsing and querying
- Map-based interface
- Data visualization
- Data storage
- Sensor access and control

Software,
Data management

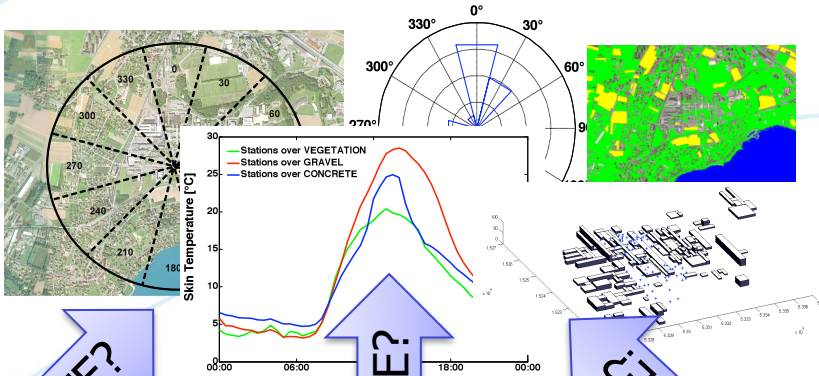
Hardware,
communication

Data Management Challenges

Analysis,
Modeling

Software,
Data management

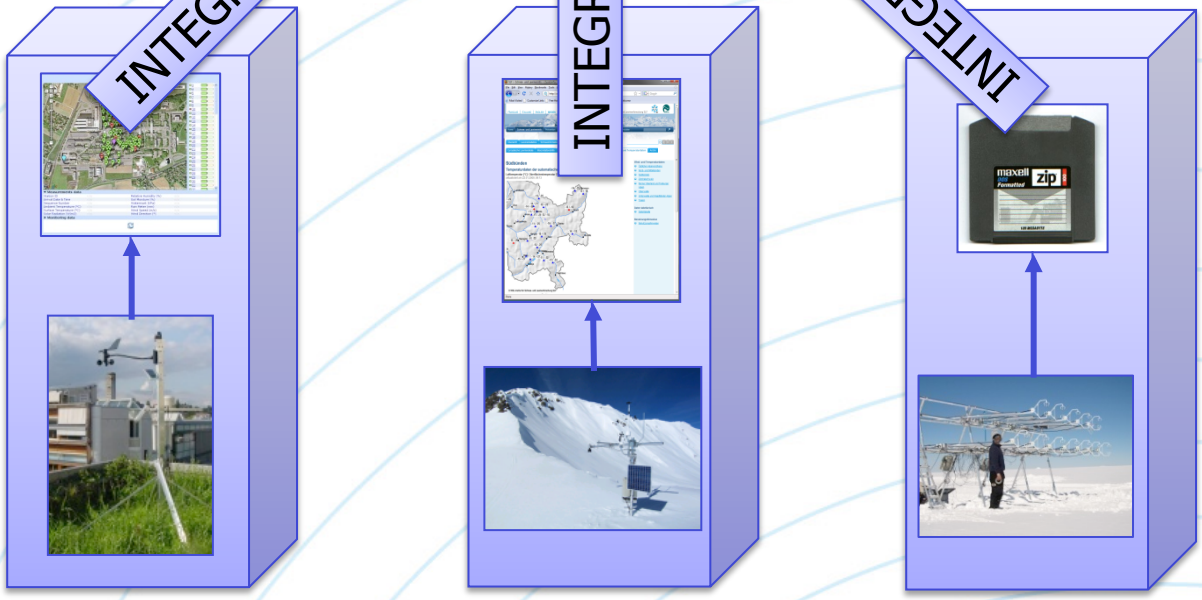
Hardware,
communication



INTEGRATE?

INTEGRATE?

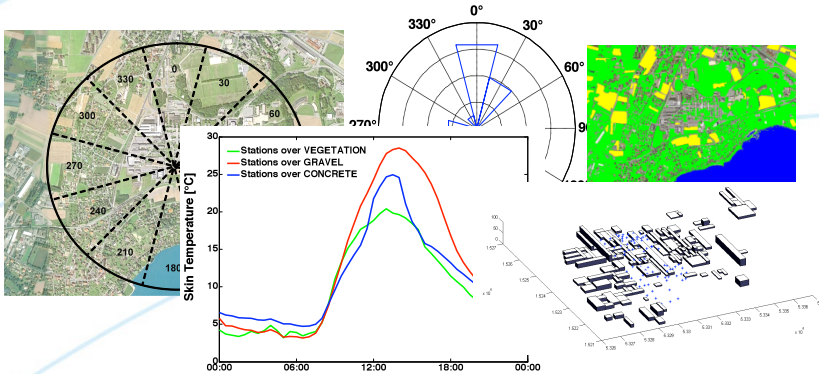
INTEGRATE?



SensorScope **SLF** **Complex weather station**

Data Management Challenges

Analysis, Modeling



Data management middleware

Hardware, communication

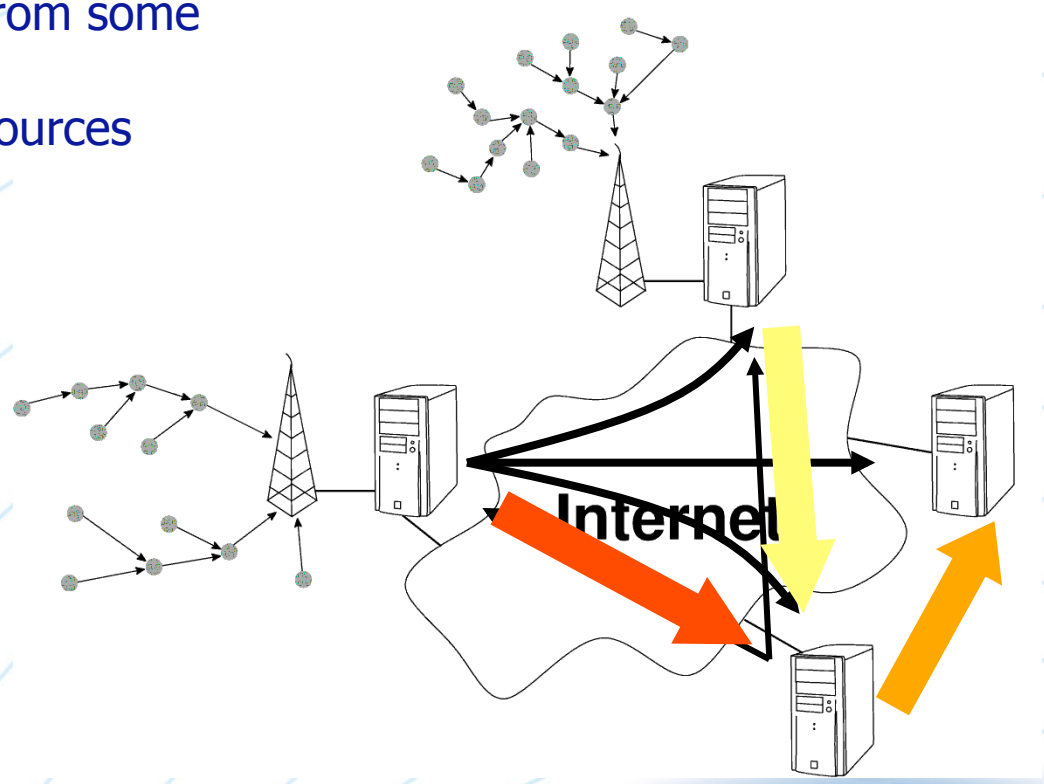
SensorScope **SLF** **Complex weather station**

Global Sensor Networks [VLDB 2006, MDM 2007, MDM 2009]

- **GSN = Internet-scale infrastructure for rapid deployment and integration of heterogeneous sensor networks**
- **Goal: fast deployment, simple publishing, simple sharing**
  - … for data stream sources
  - … based on heterogeneous sensing platforms
  - … supporting data sharing and correlation
- **Design principles**
  - Any Internet node can be producer or consumer of sensor data streams
  - Any data stream, be it raw or processed, logically be treated equally
  - Processing sensor data (filtering, correlation) should be specified declaratively
  - Sensor data streams can be deployed or removed at any time in runtime
- **Light-weight, portable and efficient implementation**

Sensor Data Sharing

- **Peer-to-peer streaming architecture**
  1. Peers announce availability of sensor data
  2. Peers request data from some sensor data sources
  3. Data streams from sources to the peers



One common abstraction: Virtual Sensor

- **Virtual sensor**
  - Separates low level implementation from application logic
  - Determines structure of data stream and metadata
  - Wrappers encapsulate different types of stream data sources
  - Both raw data and processed data streams are virtual sensors
  - Light-weight specification
- **Data stream structure**
  - Relational tuple streams with timestamps
  - Time attributes generated and maintained by default at each processing step – no specific temporal model supported
  - Declarative specification of data processing in SQL
- **Metadata**
  - Deployment and discovery

Example: a Virtual Sensor Specification

This is all you ever have to specify as GSN user!

Structure of data stream

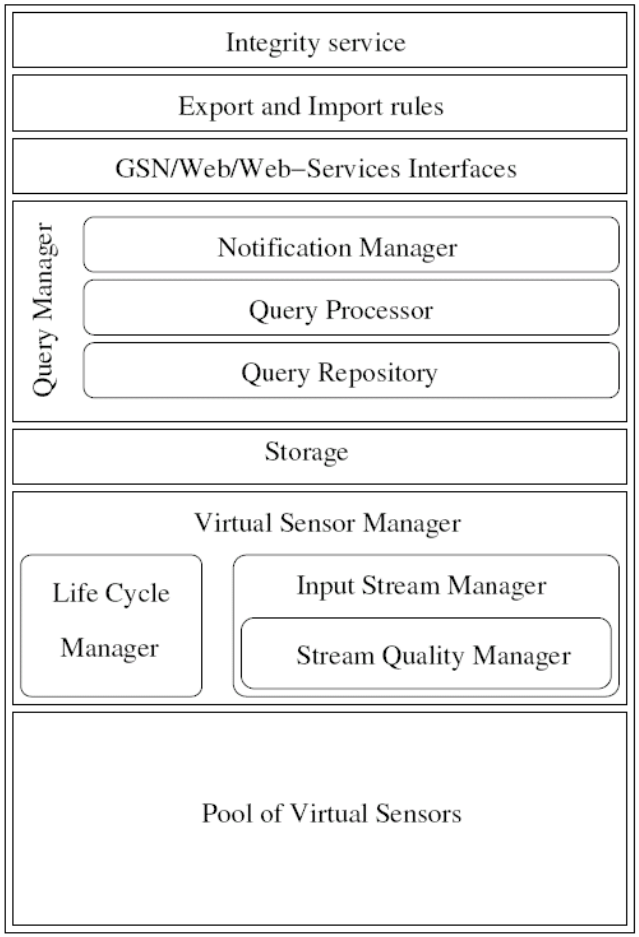
```
<virtual-sensor name="room-monitor" priority="11">
  <addressing>
    <predicate key="geographical">BC143</predicate>
    <predicate key="usage">room monitoring</predicate>
  </addressing>
  <life-cycle pool-size="10" />
  <storage permanent="true" history-size="10h" />
<output-structure>
    <field name="image" type="binary:jpeg" />
    <field name="temp"  type="int" />
  </output-structure>
<input-streams>
    <input-stream name="camera">
      <stream-source alias="camera"  storage-size="1"
                     disconnect-buffer-size="10">
        <address wrapper="remote">
          <predicate key="geographical">BC143</predicate>
          <predicate key="type">Camera</predicate>
        </address>
        <query>select * from WRAPPER</query>
      </stream-source>
      <stream-source alias="temperature1"
              storage-size="1m"
              disconnect-buffer-size="10">
        <address wrapper="remote">
          <predicate key="type">temperature</predicate>
          <predicate key="geographical">BC143-N </predicate>
        </address>
        <query>
          select AVG(temp1) as T1 from WRAPPER
        </query>
      </stream-source>
      <stream-source alias="temperature2"  storage-size="1m"
              disconnect-buffer-size="10">
        <address wrapper="remote">
          <predicate key="type">temperature</predicate>
          <predicate key="geographical">BC143-S</predicate>
        </address>
        <query>
          select AVG(temp2) as T2 from WRAPPER
        </query>
      </stream-source>
      <query>
        select camera.picture as image, temperature.T1 as temp
        from   camera, temperature1 where  temperature1.T1 > 30
         AND  temperature1.T1 = temperature2.T2
      </query>
    </input-stream>
  </input-streams>
</virtual-sensor>
```

Binding to physical device

Combining data from multiple stream sources

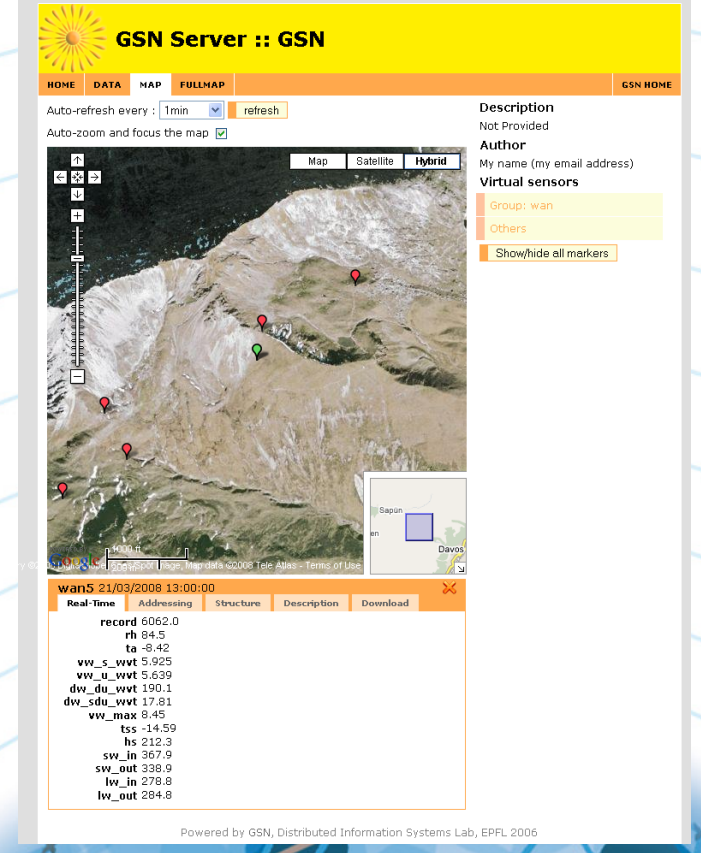
Managing Virtual Sensors in Runtime

- **GSN container**
  - Software that hosts several virtual sensors
  - Design inspired by application servers
- **Life cycle management**
  - Instantiates and deletes VS in runtime
  - Manages resources for VS and wrappers
  - Handles failures
- **Data stream management**
  - Manages data consumed and produced
  - Processing of virtual sensor queries
  - Interface to persistent storage
  - Handling notifications

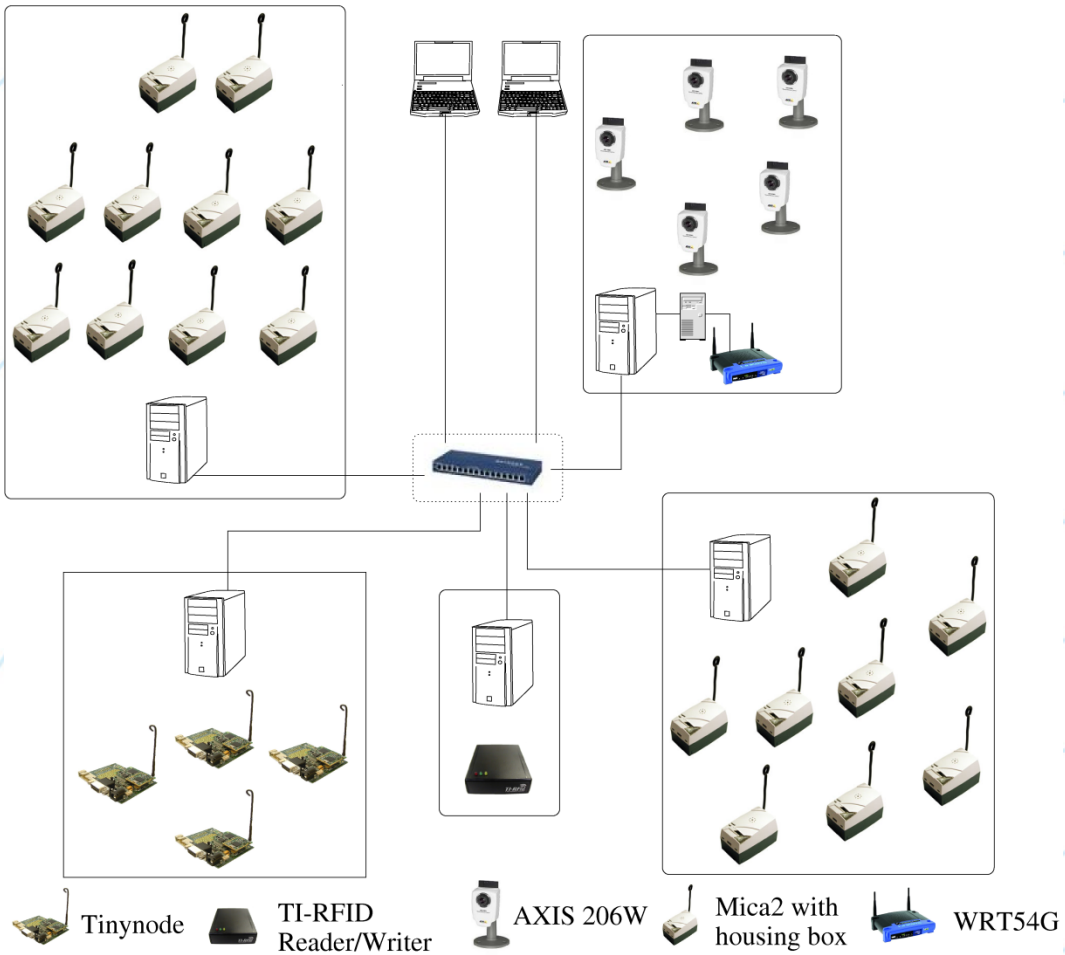


GSN System

- **Implementation**
  - Java
  - Standard Web and Web Service Interfaces
- **Various Services**
  - Notification, Data replay, Safe Storage, R integration, Visualization, …
- **Support for data stream sources**
  - Sensor motes, cameras, RFID, etc.
- **Deployments**
  - Used in various Sensorscope deployments
  - Permanent deployment in Davos
- **Release**
  - available on Sourceforge (http://gsn.sourceforge.net/)
  - >2000 downloads
  - Used by several other research groups in various application domains

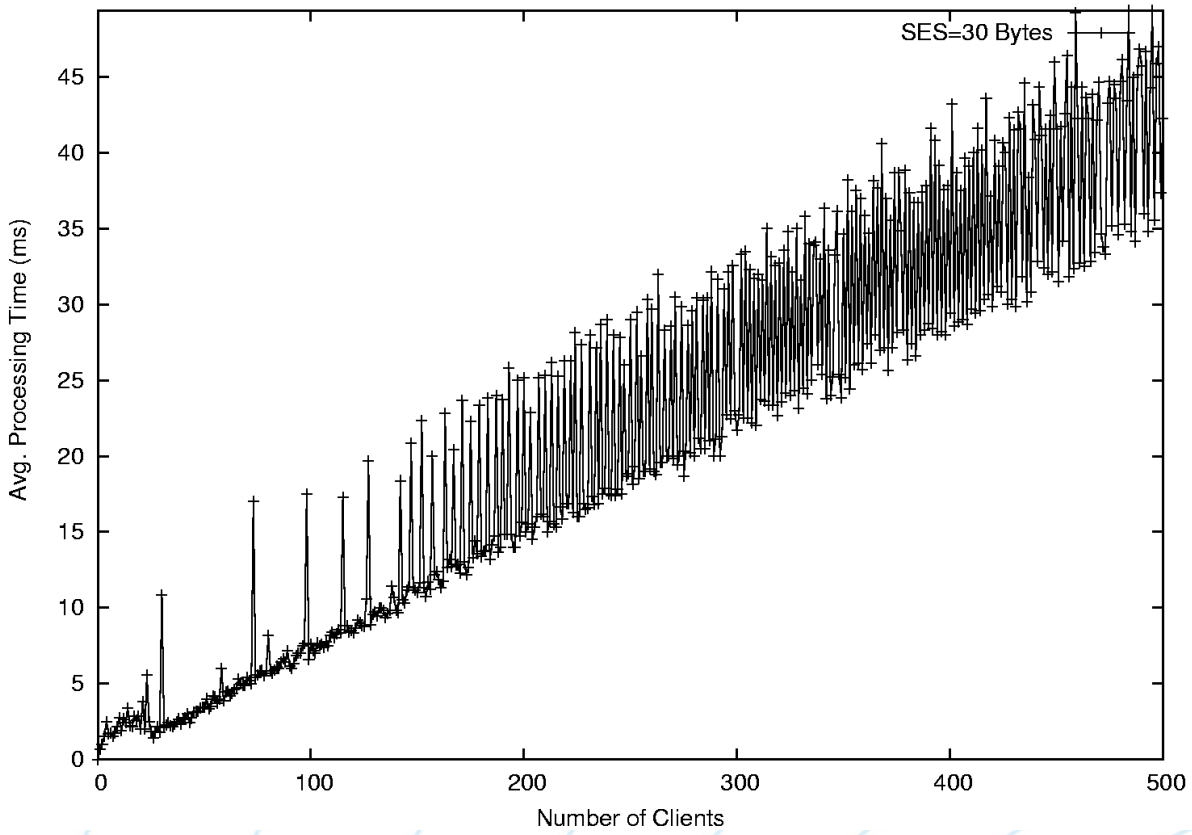


Experimental Evaluation



- 5 desktop PCs
  - Pentium 4, 3.2GHz with 1MB cache, 1GB memory, 100Mbit Ethernet, Debian 3.1
  - Linux kernel 2.4.27, MySQL 5.18
- **SN-1:** 10 Mica2 motes with light and temperature sensors, packet size 15 Bytes, TinyOS
- **SN-2:** 8 Mica2 motes with light, temperature, acceleration, and sound sensors, packet size 100 Bytes, TinyOS
- **SN-3:** 4 Shockfish Tiny-Nodes with a light and two temperature sensors packet size 29 Bytes, TinyOS
- **SN-4:** 15 wireless 8002.11b cameras (AXIS 206W), 640x480 JPEG, 5 with 16kB average image size, 5 with 32kB, 5 with 75kB
- **SN-5:** TI Series 6000 S6700 multi-protocol RFID reader with three different kind of tags (up to 8KB of data)
- 2 laptops as observers

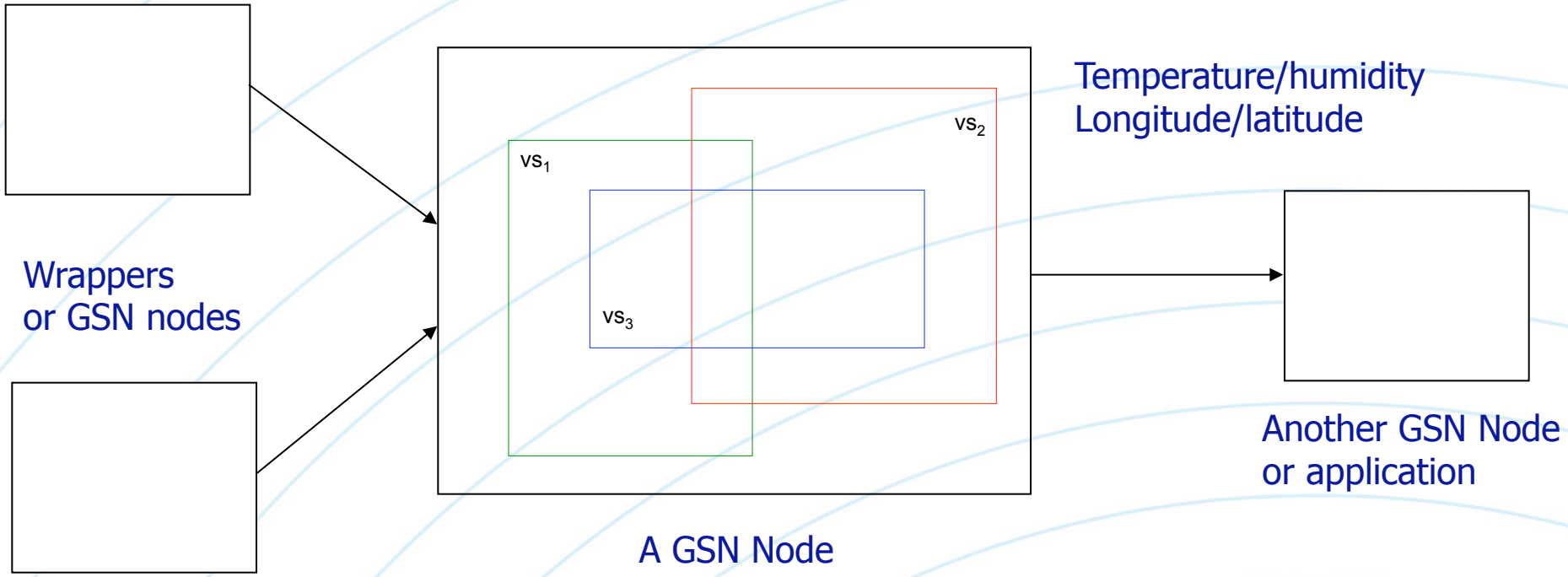
Scalability in number of clients



- 2 1.8 GHz Centrino laptops with 1GB memory as observers
- **Each ran up to 250 lightweight GSN instances**.
- Each instance produced random queries with varying table names, varying filtering condition complexity, and varying configuration parameters
- 3 filtering predicates in the where clause on average, using random history sizes from 1 second up to 30 minutes and uniformly distributed random sampling rates (seconds) [0.01, 1]
- Motes produce random bursts (1-100 data items) with 25% probability

total processing time per client is about 0.1 millisecond!

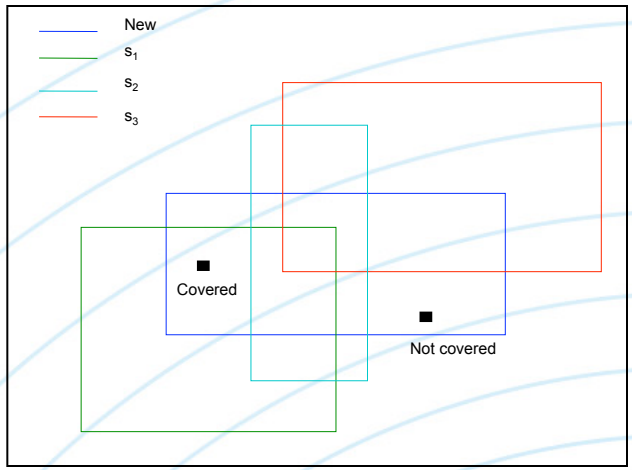
Optimizing Data Stream Processing



- **Eliminate redundant virtual sensors**
  - Reduction of traffic
  - Reduction of load for brokering nodes and sensor network
- **Problems**
  - Checking set coverage is NP complete!
  - Virtual sensors aggregate data over time windows

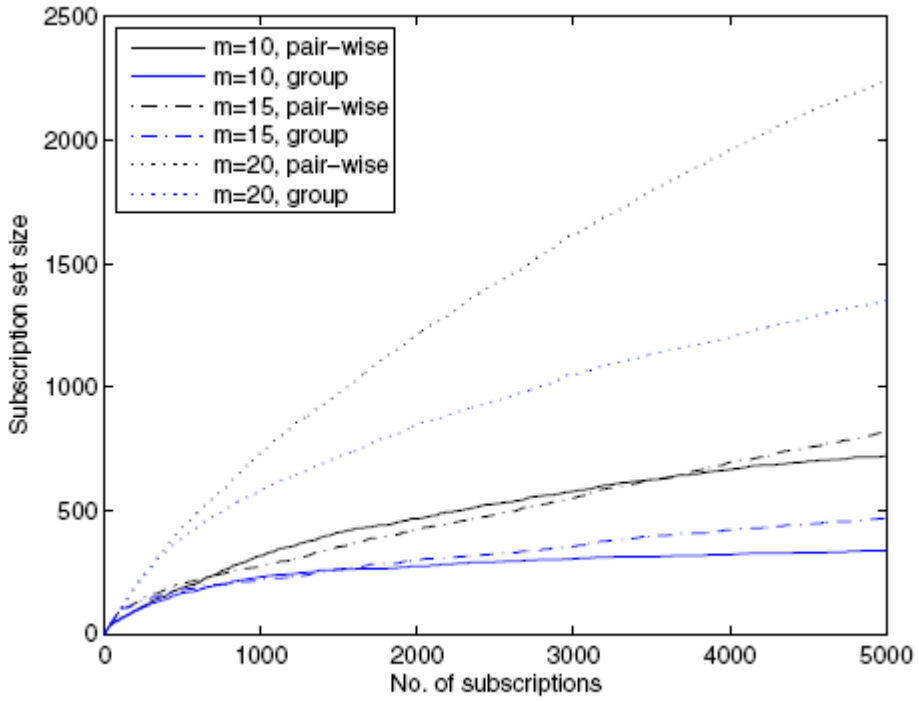
Probabilistic Subsumption Checking [Middleware 2006]

- **When new virtual sensor arrives**
  - Fast deterministic checking of specific overlaps (pairwise coverage, no coverage)
    - Produces a minimized cover set for further checking
  - Perform probabilistic checking (Monte Carlo method)
    - Generate random points inside the new subscription
    - Check each point against the minimized cover set
    - Number of trials depends on accepted required error probability



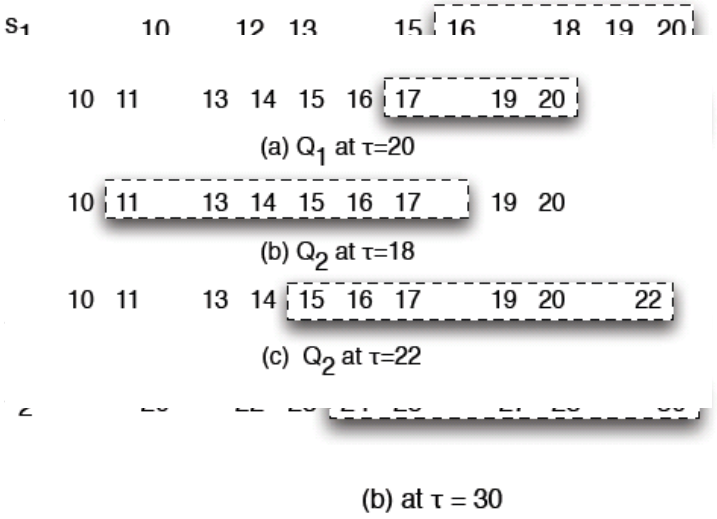
Performance Evaluation

- **Pairwise vs probabilistic subsumption checking**
  - m number of attributes



Subsumption Checking with Time Windows [VLDB 2009]

- **When new virtual sensor arrives**
  - Assume query VS1 subsumed by VS2 when ignoring time window
  - Window predicates specify length of window available at time t and sliding step s (distance among windows)
  - We provided a complete characterization for subsumption checking with time windows

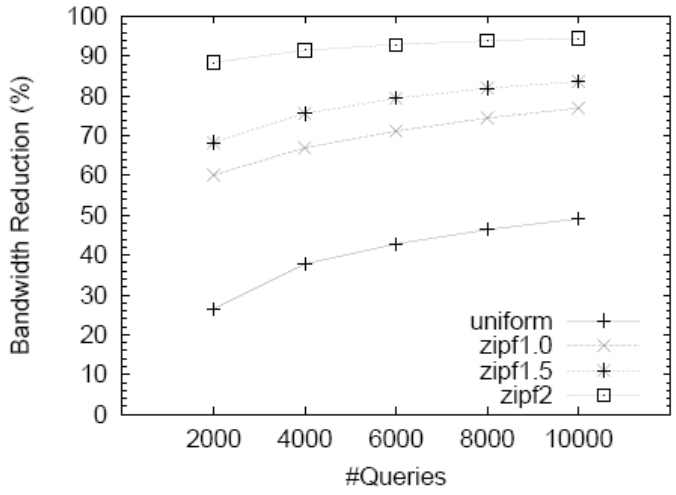


(b) at τ = 30

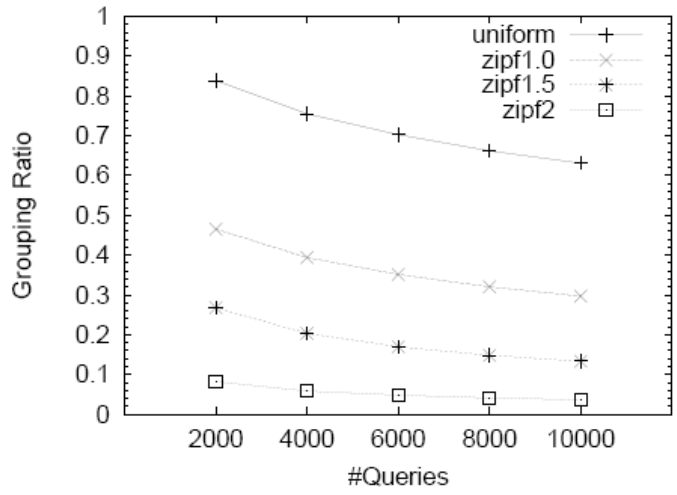
Figure 3: Temporal relations defined by two example queries on the same stream. The parameters are: $interval(Q_1) = 4, interval(Q_2) = 8, slide(Q_2) = 4$

Performance Evaluation

- **Cost savings by grouping overlapping queries into query groups with a single representative**



(a) Bandwidth Reduction



(b) Grouping Ratio

Swiss Experiment:
Web 2.0 e-science

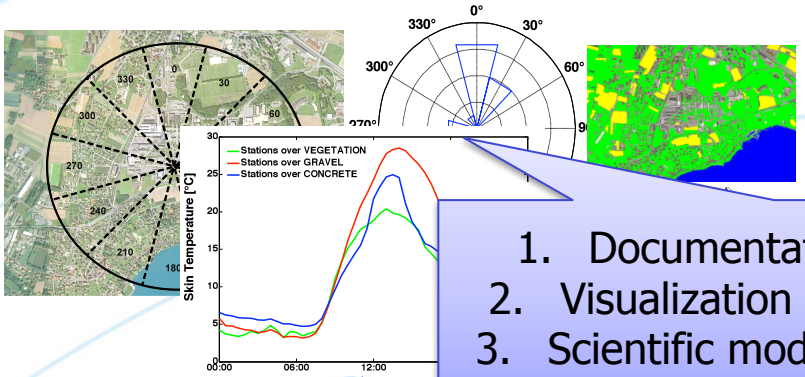
Work by: all mentioned before plus
Michi Lehning, Feng Zhao, Matthias Bavay, Nick Davies, Sebastian Michel, Ashwin Kumar, Liqian Luo, Wu Ji, Andreas Wombacher

NCCR MICS
National Competence Center In Research
Mobile Information and Communication Systems

FNSNF
Fonds National Suisse
Schweizerischer Nationalfonds
Fondo Nazionale Svizzero
Swiss National Science Foundation

Information Management Challenges

Analysis,
Modeling



1. Documentation of experiments
2. Visualization of experimental data
3. Scientific models and analysis tools

Software,
Data management

Hardware,
communication

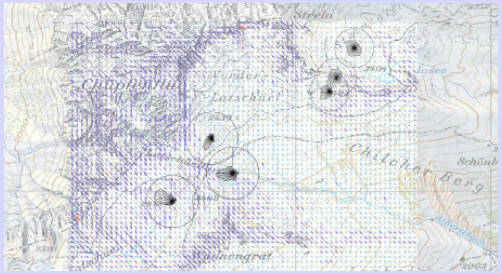
SensorScope **SLF** **Complex weather station**

Swiss Experiment Platform – Science Wikipedia

- **Science Wikipedia**
  - collaborative platform to collect and share scientific knowledge and support scientific work

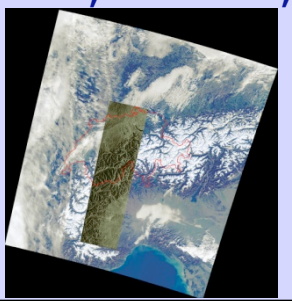
Data captured through wireless sensor networks

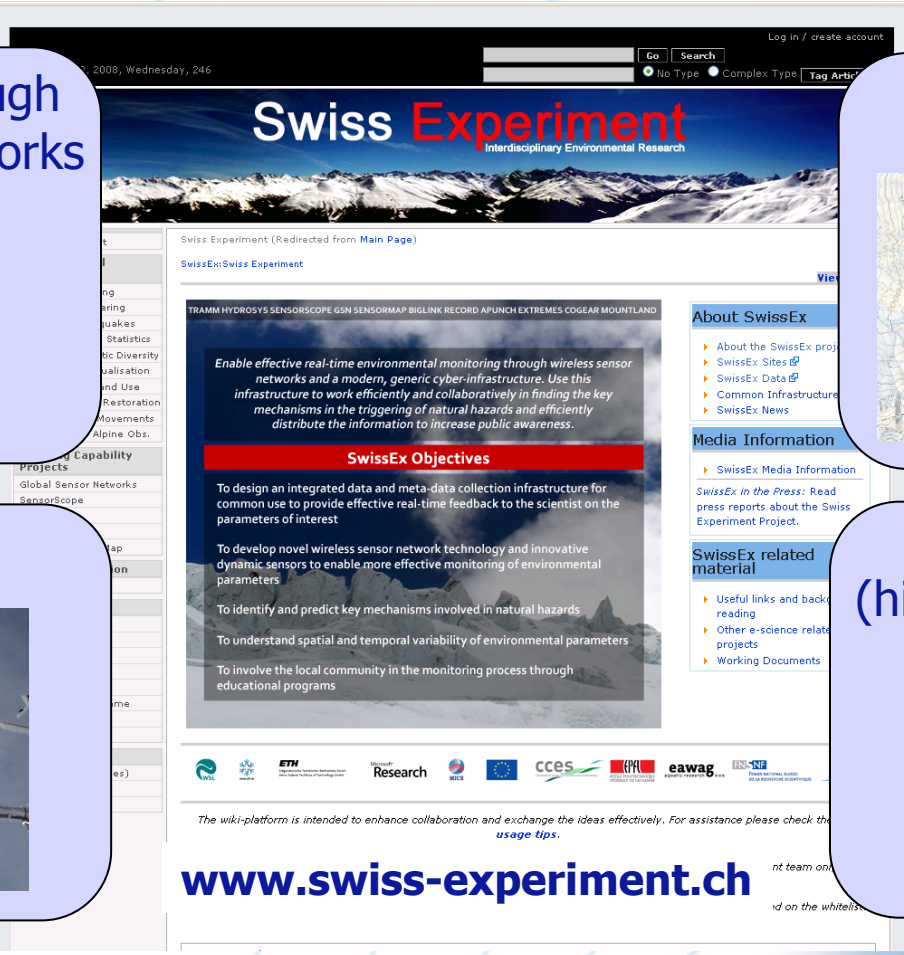
Models and simulations



Experiments

Other scientific data (historical, satellite, etc.)

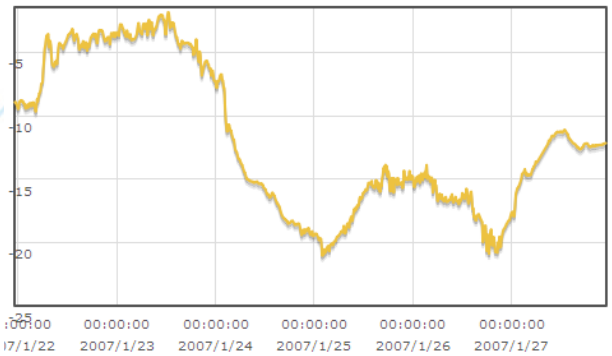




Information Management in Environmental Science

[e-science 2008, ICDE 2009]

- **Scientific work requires structured and unstructured information**
  - Unstructured: text, images (e.g. experiment documentation)
  - Structured: properties of sensors, experiments, measurements



Temperatures from Wannengrat, Davos 22/01/07 – 28/01/07

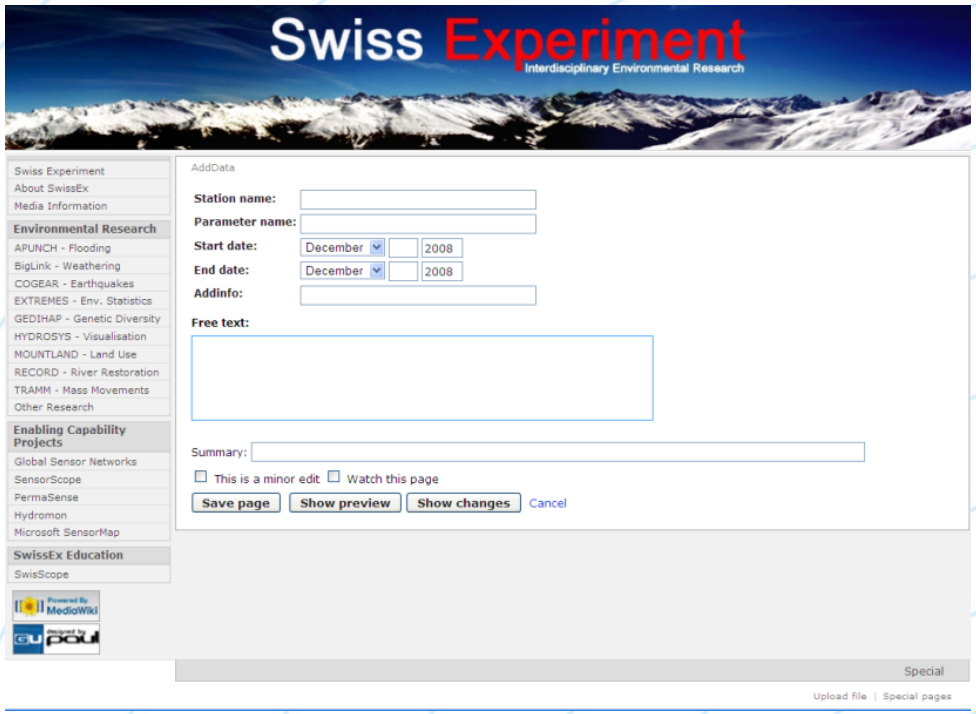
Seem legitimate for 2400m altitude in winter?

Sensor opening covered in reim

The data from this sensor during the preceding period is unusable, but indistinguishable from the data after this photo was taken.

Semantic WIKI

- **Collaborative data management through WIKI platform**
  - Data stored in RDF and queried in SPARQL
  - Both predefined and ad-hoc (semistructured) data structures
  - Querying across experiments and data exchange



Data entry

Sensor Database Parameter

DATABASE PARAMETER INFORMATION	
STATION NAME	wan1
DATABASE PARAMETER NAME	sw_in
PARAMETER ID	11
SENSOR SERIALNO	970047/8(-24)
MEASURED QUANTITY	Incoming Shortwave Radiation
UNITS	W/(m^2)
SAMPLING FREQUENCY	1Hz averaged over 10mins
MEASUREMENT ACCURACY	unk

EDIT MODEL INFORMATION	
EDIT LINK	Special:EditData/Mediawiki:DParameter /Wannengrat:232f688/

Invalid data and observations

	ACTION	START DATE	END DATE
2f07ab5	Invalid data	18 August 2008	4 September 2008
4e56f65	Invalid data	13 December 2007	13 December 2007
937ea62	Invalid data	12 December 2007	12 December 2007
B42d022	Invalid data	23 March 2008	24 March 2008
C770345	Invalid data	6 February 2008	6 February 2008
D5ed480	Invalid data	23 November 2007	27 November 2007
E34cc04	Invalid data	24 June 2008	8 August 2008
F683be1	Invalid data	21 March 2008	21 March 2008
F8a1bf2	Invalid data	12 January 2007	15 March 2007

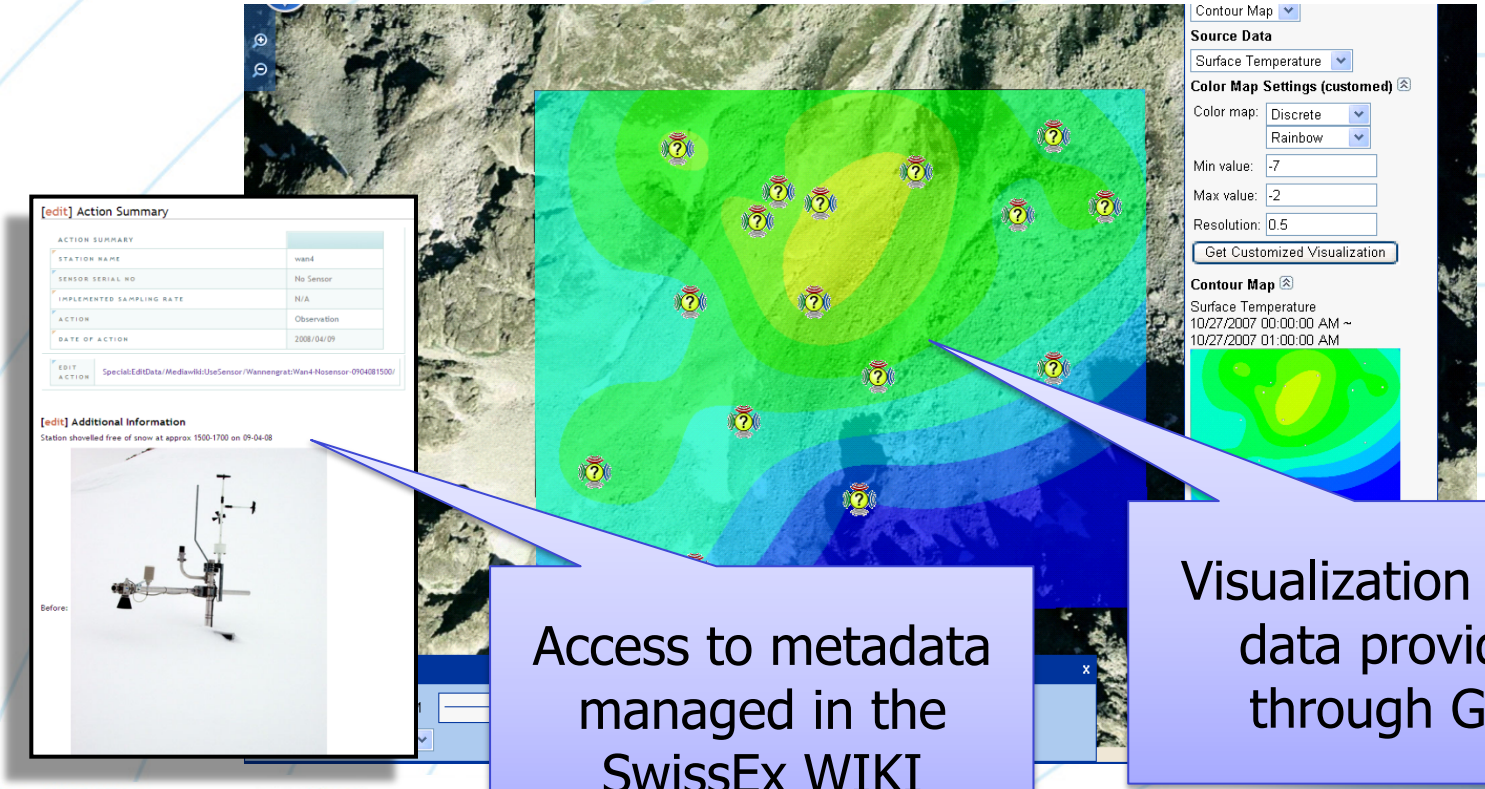
Click here to register invalid data

Click here to add an observation affecting this parameter

Data browsing

Visualization and Internet Sensor Data Publishing

- **Support for Sensor Data Management and scientific work**
- **SensorMap: collaboration with MS Research Redmond**
  - Web-based visualization of sensors and measurements
  - Spatial distributions and changes over time
  - Web Service based architecture



Application Scenarios

Marc, a hydrologist, utilizes SensorMap during

Experiment Planning

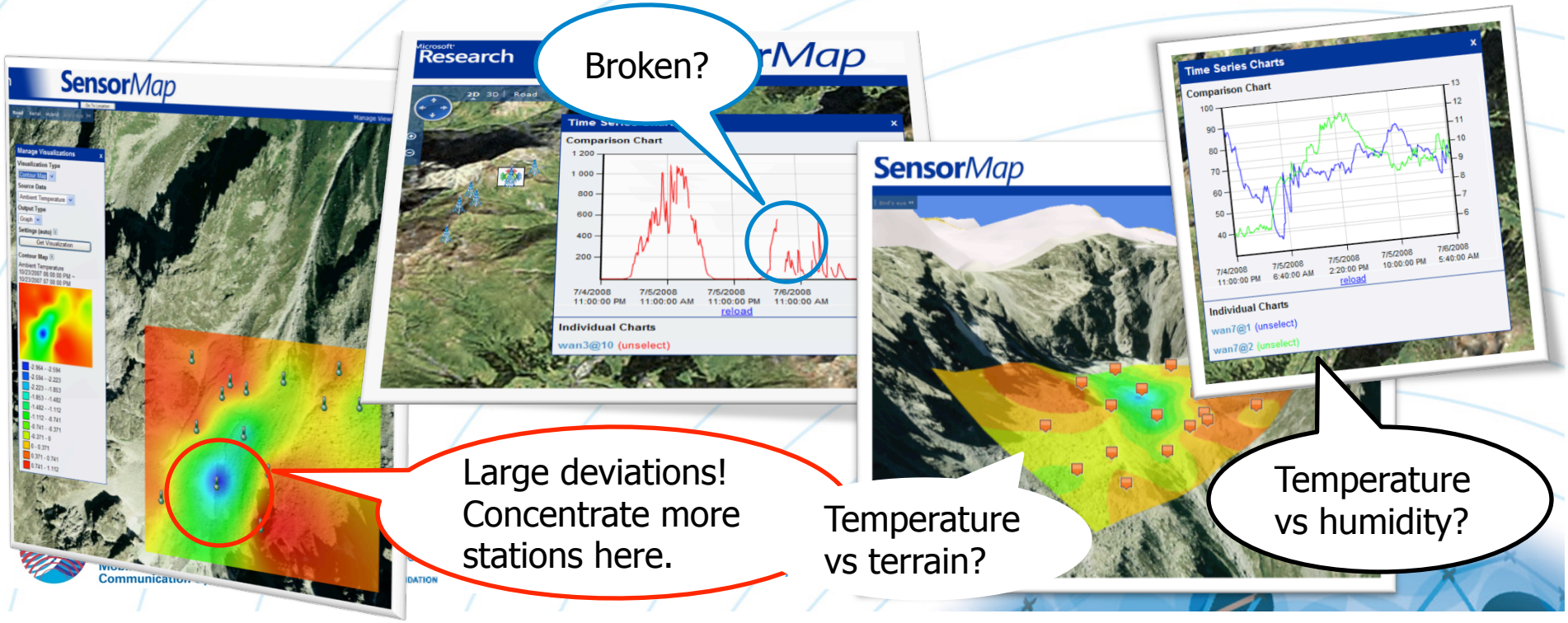
To view sensor layout and visualize measurements in real-time to decide the placement of sensors

Deployment Monitoring

To inspect real-time output of sensors, and to discover and fix broken sensors

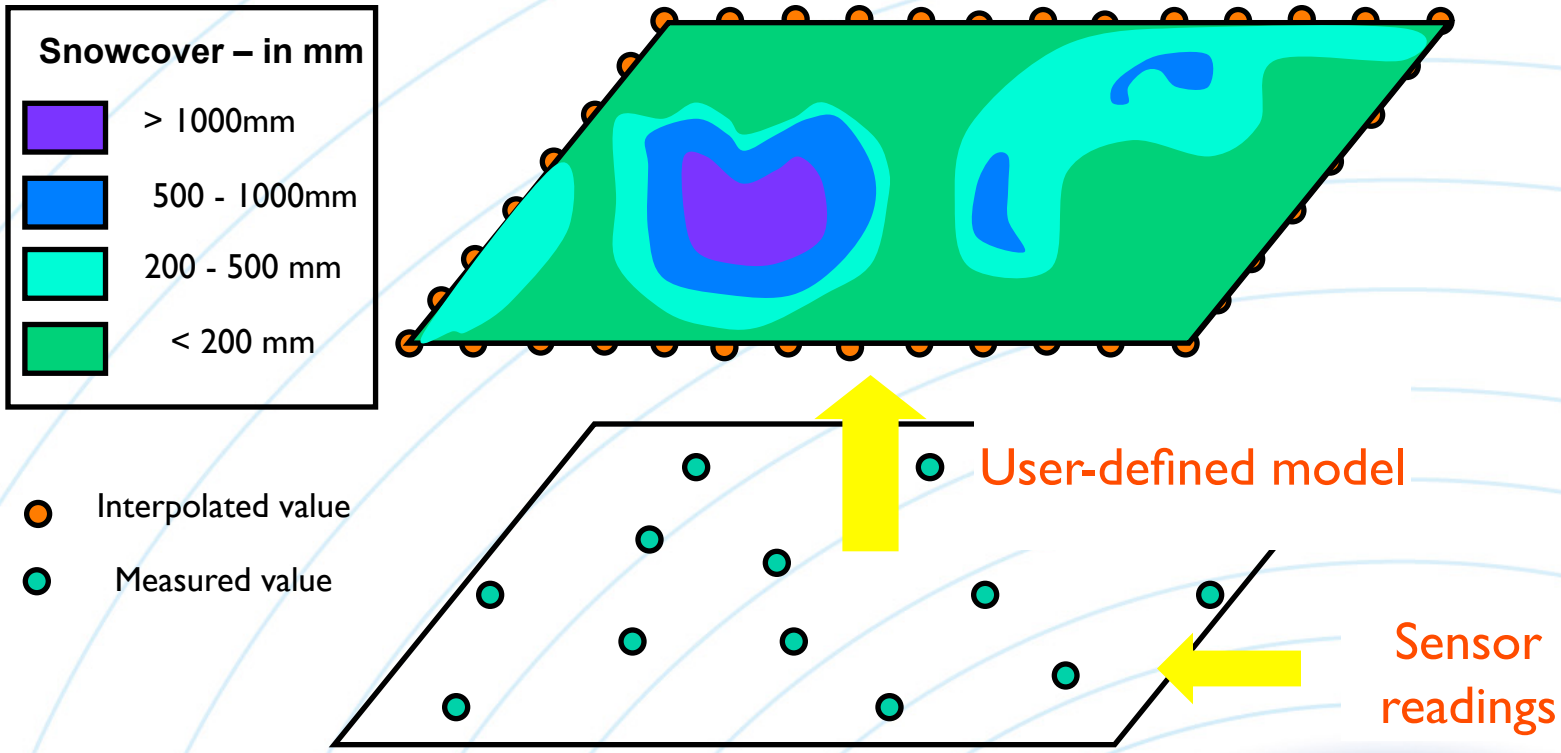
Data Analysis

To visualize dependencies among different measurements and correlations with topological terrains

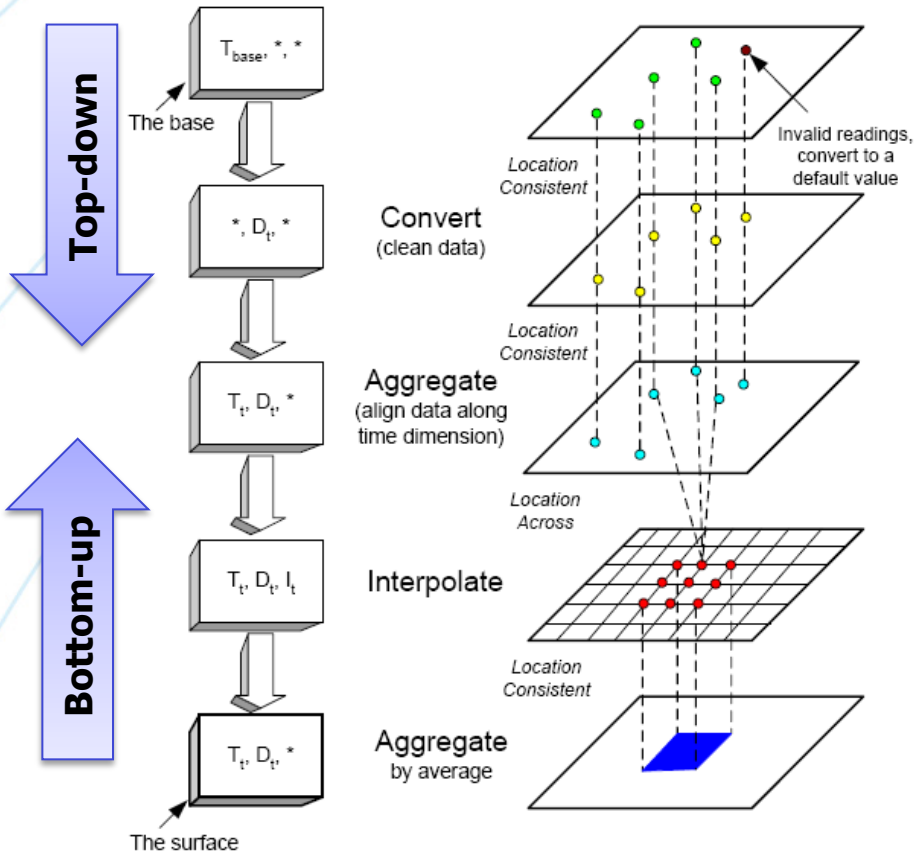


Support for Scientific Models

- **Example – Snow Cover Distribution**

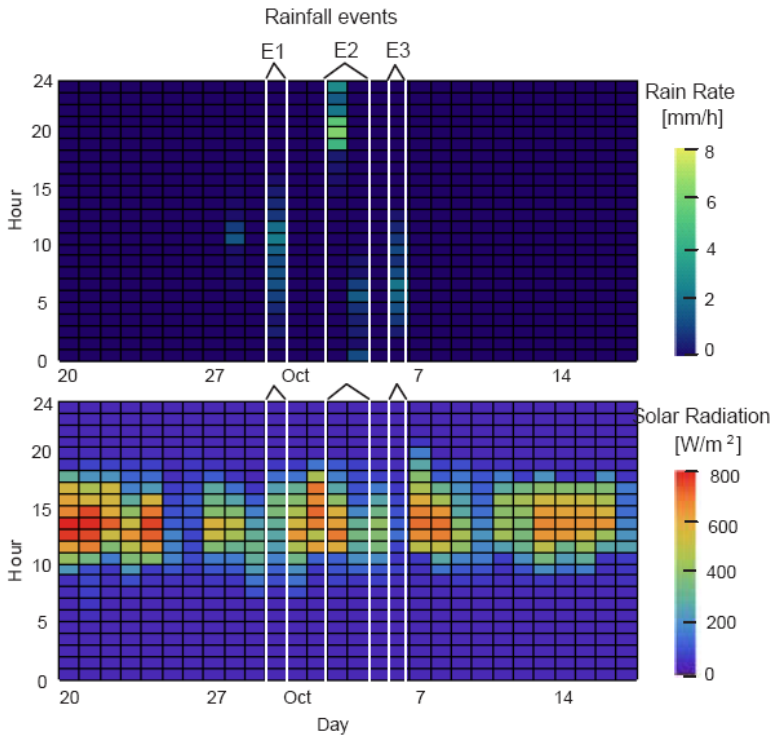


Model-based Views and Optimization [EDBT 2009]

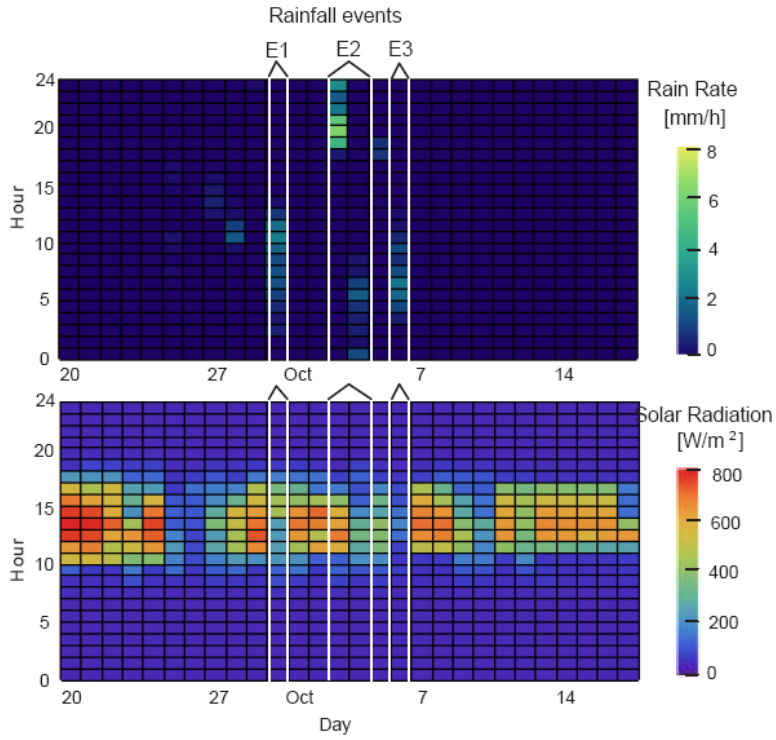


- **Top-down vs bottom-up evaluation**
- **Bottom-up**
  - Incremental results: early visualization
  - Redundant and unnecessary computation: caching
  - For rapid visualization compute first regions of interest (high variability)

Incremental Evaluation



15% of results



100% results

Conclusion

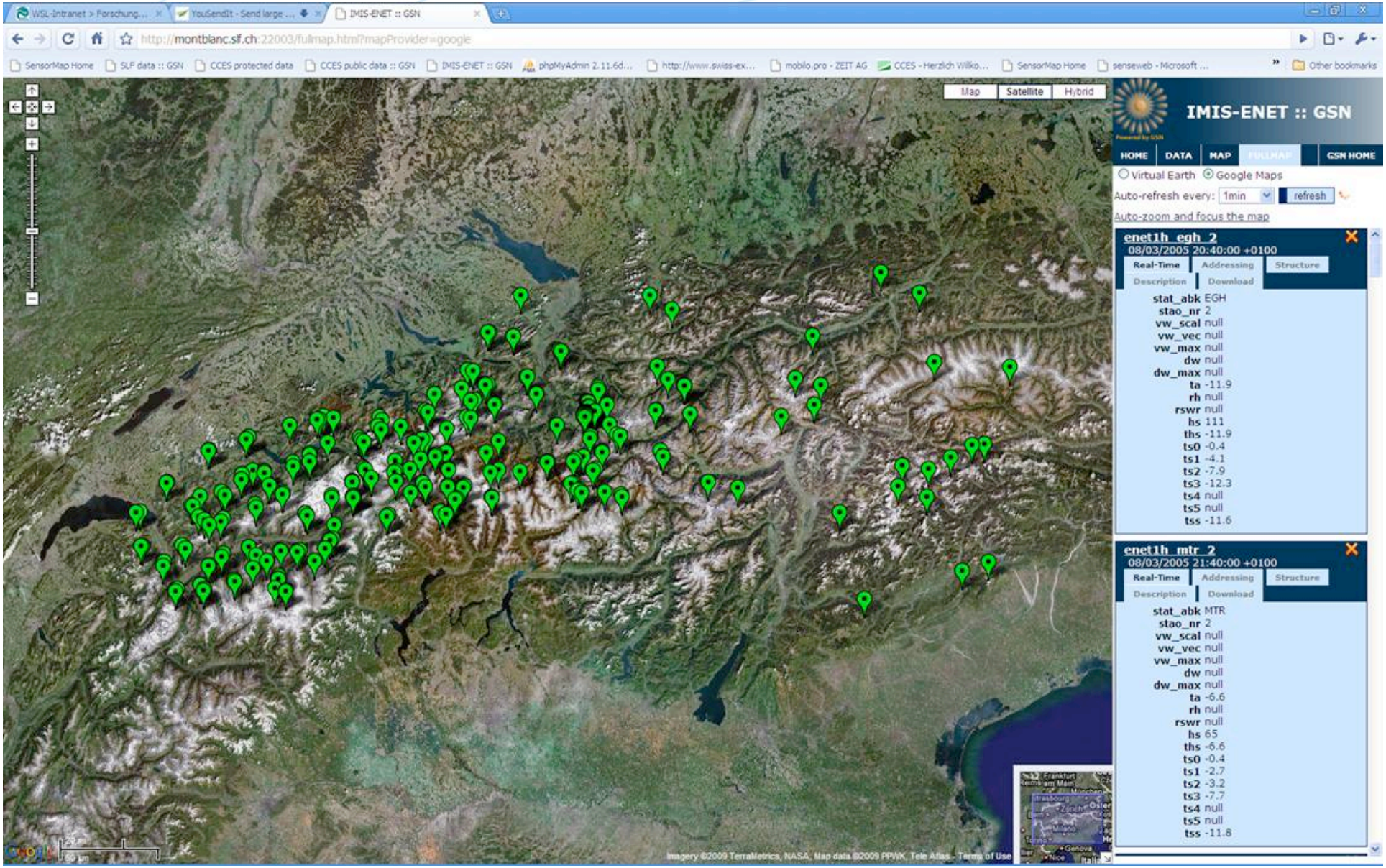
Conclusion

- **Challenge**
  - Today limited capabilities for gathering, analyzing and visualizing data to obtain reliable models
- **Wireless sensor networks**
  - In-situ capturing environmental data
  - Complements remote sensing and point measurements
- **Data-driven computational science**
  - Data publishing, integration and collaboration tools
  - Broad community involvement
- **Environmental computational science**
  - Collaborative capturing, sharing and analyzing of real-time sensor data
  - New (and often unexpected) data management challenges

Conclusion

- **Swiss Experiment is an example of a new type of computing system**
  - *Complex phenomena* in the environment monitored in real-time
  - *Interpretation* of data integrating models, simulations, data repositories and human interaction
  - *Control actions* influence environment, data capturing and interpretation

SwissEx Coverage



References

[Middleware 2006] A. M. Ouksel, O. Jurca, I. Podnar, K. Aberer, "Efficient Probabilistic Subsumption Checking for Content-Based Publish/Subscribe Systems", Middleware 2006: 121-140.

[VLDB 2006] K. Aberer, M. Hauswirth, A. Salehi, "A middleware for fast and flexible sensor network deployment", Demonstration, VLDB 2006, Seoul, Korea, September 2006.

[MDM 2007] K. Aberer, M. Hauswirth, A. Salehi, "Infrastructure for data processing in large-scale interconnected sensor networks", Mobile Data Management (MDM'07), Mannheim, Germany, May 7 - 11, 2007.

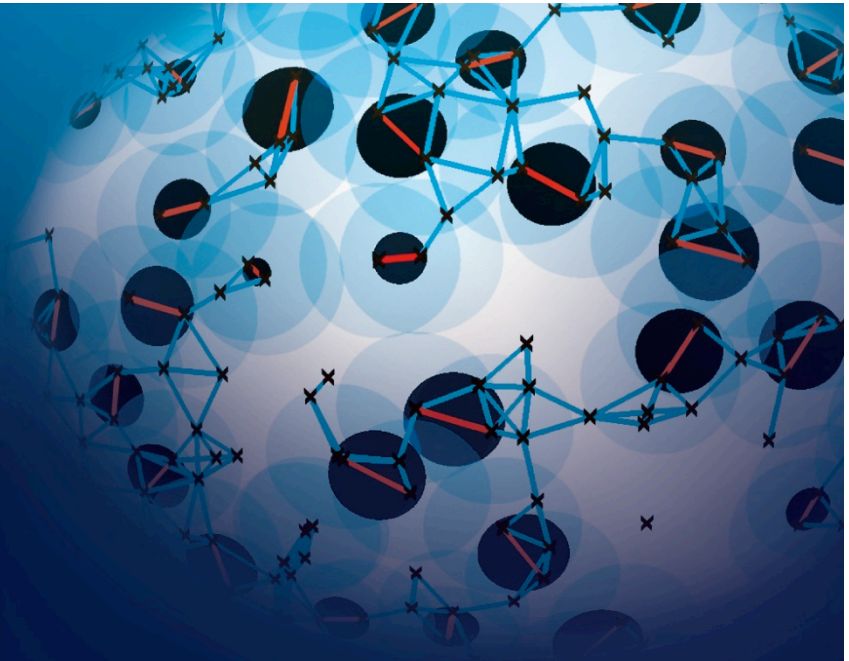
[netDB 2008] Y. Zhou, K. Aberer, A. Salehi and K.-L. Tan, "Rethinking the Design of Distributed Stream Processing Systems", Proceedings of the 24th International Conference on Data Engineering Workshops, 2008.

[e-science 2008] N. Dawes, A. Kumar, S. Michel, K. Aberer, M. Lehning: " Metadata Management and its Application in Large-Scale Environmental Research", 4th IEEE International Conference on e-Science, 2008.

[ICDE 2009] Michel S., Salehi A., Luo L., Dawes N., Aberer K., Barrenetxea G., Bavay M., Kansal A., Kumar K.A., Nath S., Parlange M., Tansley S., Van Ingen C., Zhao F., Zhou Y. Environmental Monitoring 2.0, ICDE 2009.

[EDBT 2009] J. Wu, Y. Zhou, K. Aberer, K.L. Tan, "Towards Integrated and Efficient Scientific Sensor Data Processing: A Database Approach", EDBT 2009, March 23-26 2009, Saint-Petersburg, Russia, 2009.

[VLDB 2009] Yongluan Zhou, Ali Salehi, Karl Aberer, « Scalable Delivery of Stream Query Results", VLDB 2009, Lyon, France, 2009.



thank you for your attention