Sensor Internet Share and Search

Enabling Collaboration of Citizen Scientists

Junghoo Cho†

Joint Work with Sasank Reddy*, Gong Chen*, Brian Fulkerson*,
Sung Jin Kim†, Unkyu Park‡, Nathan Yau*, Mark Hansen*, John Heidmann‡

*Center for Embedded Networked Sensing; University of California, Los Angeles
†Department of Computer Science; University of California, Los Angeles,
‡Information Sciences Institute; University of Southern California

Outline

Problem Definition

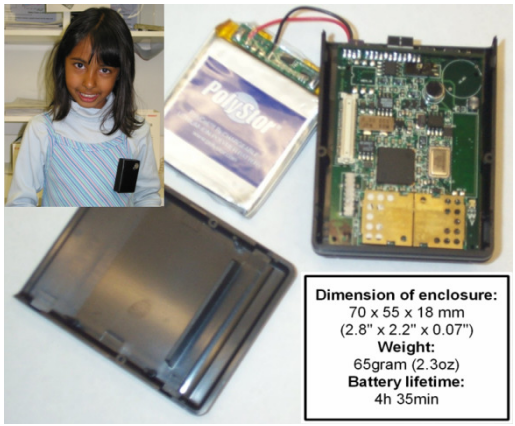
Architecture Overview

Research Challenges

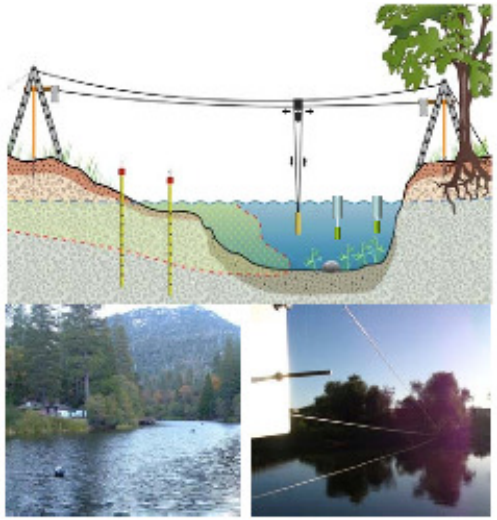
Prototypes

Current State of Sensor Networks

Sensor networks are deployed in various application areas ranging from education and science to military and industry.



Education: Smart Kindergarden



Science: Networked Info-mechanical Syste

Military: Shooter Localization

These systems are ***robust, capable, and widely adopted***.

Current State of Sensor Networks

Yet most sensor networks today:

- operate in ***isolated patches***,
- use ***different mechanisms to deliver data*** to their users,
- often have ***no formal methods to share*** data with others.

It is increasingly becoming important to enable a ***common means to share data***, preferably using the Internet as a communication platform.

Goals

The goal of our research is to build architectures and create protocols to enable users to:

- discover

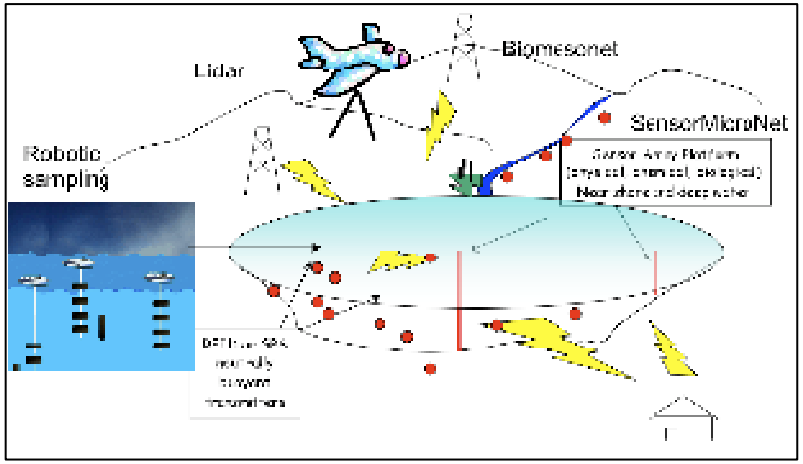
- process

- republish data

from thousands of independently operated sensors.

Goals

We see a future where there are both ***large sensornets*** that formed of multiple wireless patches ***bridged by the Internet***, and where ***individual sensors*** are deployed by ***casual users or citizen scientists***, with their data posted to sensor web logs or ***slogs***.



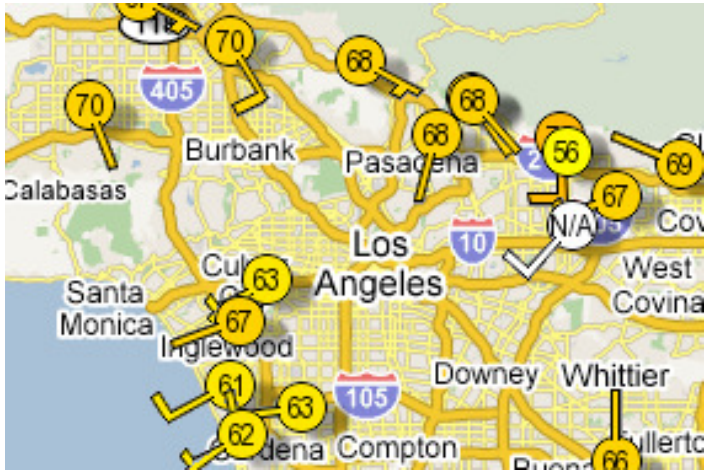
National Ecological Observatory Network (NEON)

Backyard Weather Stations (Weather Underground)

Goals

In both deployment settings (large sensornets and individual sensors), the idea of of having ***many collaborators*** exists with different scales.

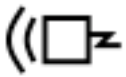
National Ecological Observatory Network (NEON)



Backyard Weather Stations (Weather Underground)

Architecture Components

sensors (clustered into sensornets)



sensor publishers that act as gateway

sensor stores

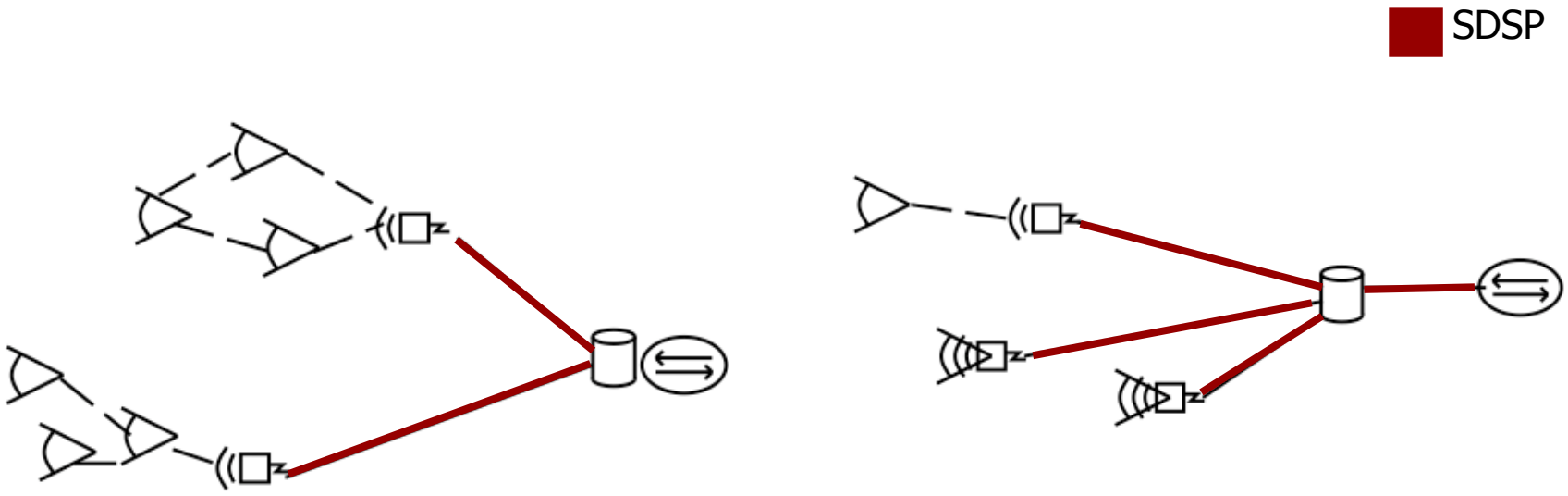
sensor search engines

Connected using a common protocol: ***Sensor Data Stream Protocol*** (***SDSP***).

Sensors and Publishers

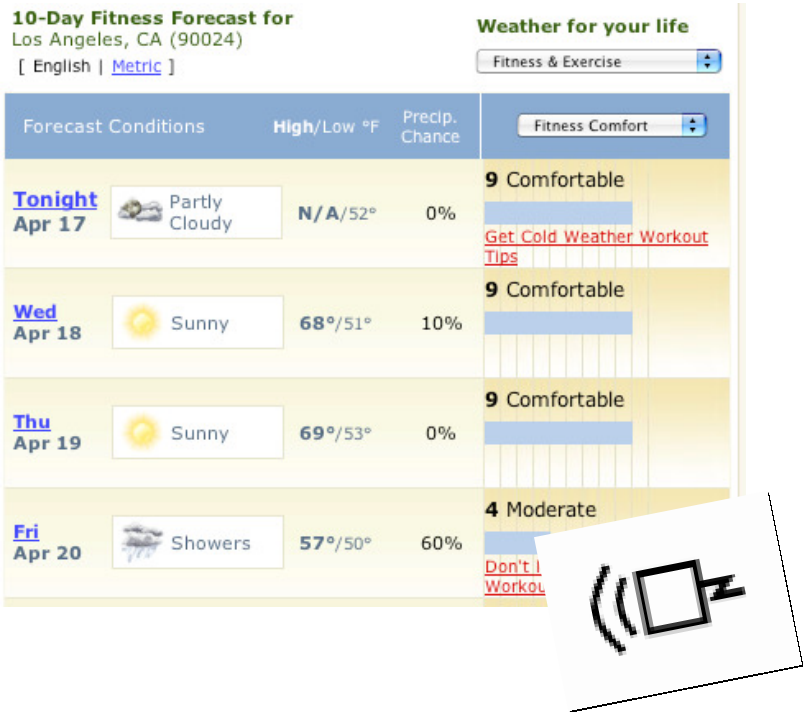
Sensors are gathered into sensornets, clusters of sensors that share a wireless network.

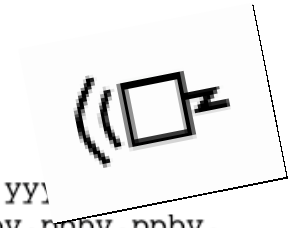
Sensor publishers ***link sensornets to the Internet***, providing a gateway from non-IP sensornet protocols to Internet standards. We expect sensor-side of the publisher to be optimized for the sensors, but the Internet side to use ***SDSP***.



Sensors and Publishers

Virtual publishers also can exist that scrape websites or other sources of sensing data and publish the data to a sensor store in a standard format.

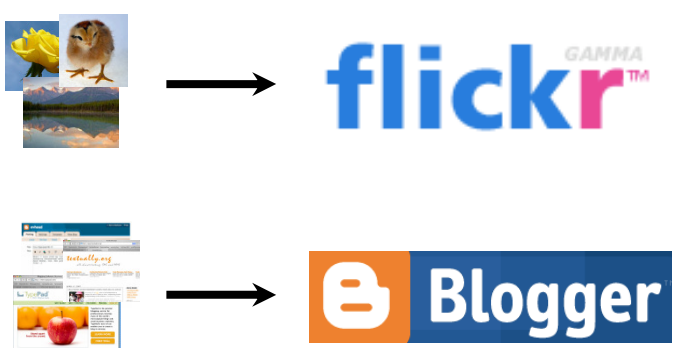


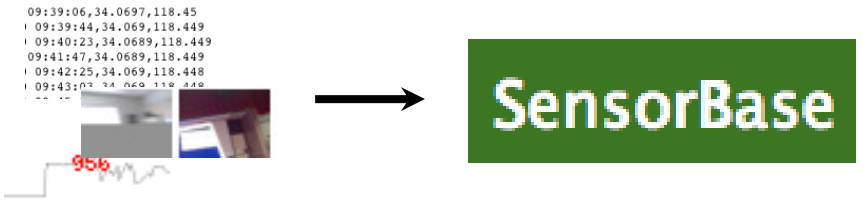


```
*TABLE COLUMN UNITS,none,none,yy
dd,hh:mm:ss,ppbv,ppbv,ppbv,ppbv,ppbv,ppbv,
*TABLE BEGINS,,,,,,,,,,,,,,,,,,,,,,,,
,LaPorte Airport,
1,2000/08/19,22:18:00,-999.999,0.183,0.441
,0.000,0.000,,,,,,,,,
,LaPorte Airport,
1,2000/08/19,22:51:00,-999.999,0.160,0.354
,0.000,0.000,,,,,,,,,
,LaPorte Airport,
1,2000/08/19,23:24:00,-999.999,0.137,0.179
,0.000,0.000,,,,,,,,,
,LaPorte Airport,
1,2000/08/19,23:57:00,-999.999,0.088,0.137
,0.000,0.000,,,,,,,,,
,LaPorte Airport,
```

Sensor Stores

Similar to how people store and share images, documents, and blogs on the Internet using services such as FlickR, GoogleBase, and Blogger, we envision a similar repositories for sensor data.





Sensor Stores

We expect to have a ***variety*** of sensor stores ranging from ***large sensor stores*** that are well connected and managed to ***small stores*** that are run by hobbyists or individual research groups.

Also, the ***price, availability, and sharing policies*** will vary.

Republishers

One advantage of sharing sensor data on the Internet with a common protocol is that it becomes easy to add value to data by republishing it.

We imagine having republishers that perform tasks including: ***data aggregation, filtering, statistical estimation, vetting, error suppression***.

The new, republished data would be available on a sensor store for others to build upon and links would exist to the original streams to enable users to investigate the republisher's work.

Sensor Search Engines

Sensor search engines will ***index and search*** sensor data.

We expect a few, large (Internet-wide) sensor search engines, augmented by local, laboratory-specific engines.

Sensor search engines will facilitate the process of new sensor ***discovery*** and ***exploration*** by general users based on the their particular need expressed by queries.

Users

Users will interact with this architecture in several ways.

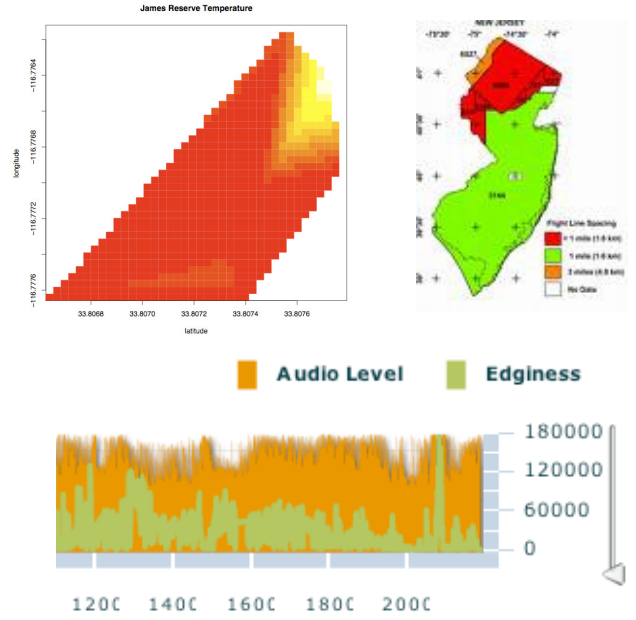
- Search engines to help users discover new data streams.
- Tools that interface Excel, Matlab, and R to sensor data.

We expect to have new approaches for sensor visualization.

- Visualize temporal and spatial trends will be important.

- Dynamically updated charts and graphs for realtime views of data.

- Data overlays on maps to analyze data that has spatial variations.



Sensor Data Streaming Protocol

Cutting across all aspects of our work is the definition of the Sensor Data Stream Protocol (SDSP), the protocol that will link publishers to sensor stores and search engines.

We also see it as being used by user programs that directly access data, complementing interactive, web access.

The protocol will use ***XML*** as the markup language and support ***bi-directional*** communication to allow:

- data to be transferred
- metadata retrieval
- sensor parameter adjustment
- event-based triggers

Research Challenges

In order to enable a marketplace for sensor data exchange and support flexible and efficient sharing, there are a few areas of interest to explore further:

- meta-data definition and sharing

- sensor search

- sensor stores

Research Challenges

Meta Data Definition and Sharing

- How do you enable a variety of users to give metadata easily.
- Raw sensor data is meaningless unless annotated by metadata.
- Type, units, calibration, and context are important.

Solutions:

- Providing a pre-packaged set of metadata for common configurations and sensor types.
- Enable casual annotations of sensors (tags and photos).
- Create enriched sensor services, where third parties will annotate, filter, and republish data including sensor confidence ratings, popularity, and reputation with trace back mechanisms.

Research Challenges

Sensor Search

Should support at least two search mechanisms: ***exploratory and analytic queries***.

Users will often have only a vague informational need that is difficult to formulate exactly (temperature and Los Angeles).

Web-based GUI to let users issue queries based on keywords.
Search engine returns a ranked list of relevant sensors.
Users manually look at the sensor values and metadata.

Solutions:
- Ranking based on popularity, reliability, meta-data richness.
- Able to translate metadata types and fill in missing metadata.

Research Challenges

When the user's need can be expressed precisely (e.g., average yearly rainfall in LA), an analytical query language is effective.

Modeling each sensor as a relational table and supporting SQL queries can be a reasonable approach.

But user is unlikely to know in advance all relevant sensors.

Solution:

- Important to extend the SQL to let users declaratively specify tables of interest like they can do with tuples in normal SQL.
- Quality and confidence values associated with sensor readings important to enable users to find source of errors.

Research Challenges

Sensor Stores

Need to be able to support a range of data disclosure models so users are comfortable sharing their data widely.

Provide a simple, yet robust interface to enable both users and other components in the architecture to publish and obtain data.

Solution:

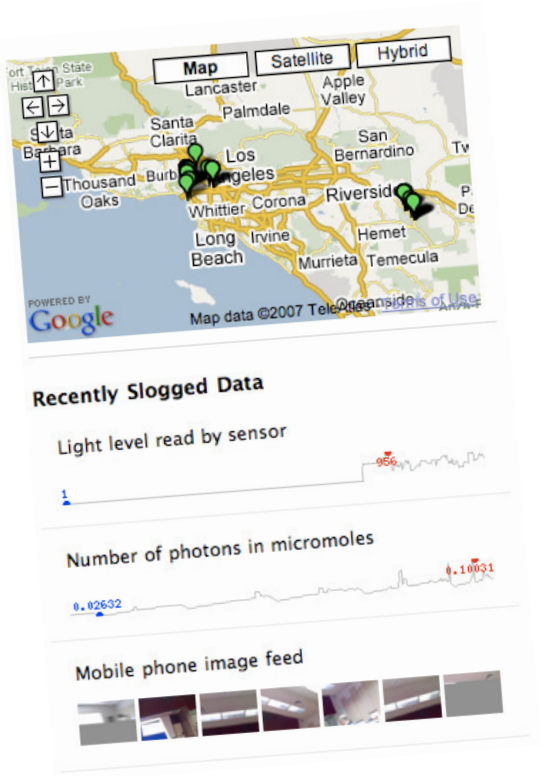
- Enable project/table/row level sharing model for data stored.
- Summaries or snapshots of actual data and general statistics about a project kept. Enable search indexing on the stores.

Prototypes

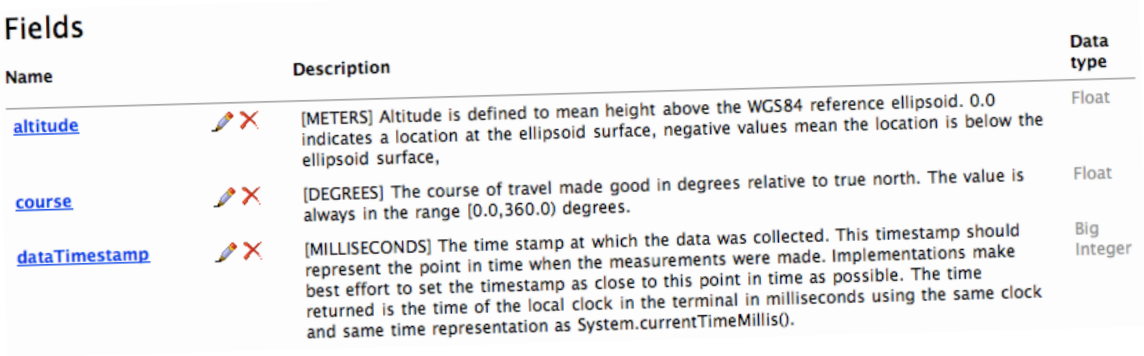
SensorBase (sensorbase.org)

Example of sensor store that is used by over 50 projects.

Provides API to post data, pull data, RSS feeds for notifications, a UI to browse/create/manage projects, etc...



		sbid	altitude	course	dataTimestamp	date	horizontalAccuracy
☐	✎	36712	0	0	0	2007-00-22 14:07:37	0
☐	✎	36713	0	0	0	2007-00-22 14:08:20	0



Prototypes

Republishers : Missing Data

Incoming data stream is temperature readings obtained from images of temperature displays. Sometimes the image processing leads to missing values for temperature value.

Republisher repairs the data stream by interpolating missing information from prior values and also using nearby sensor values with a confidence rating associated with the values as well.

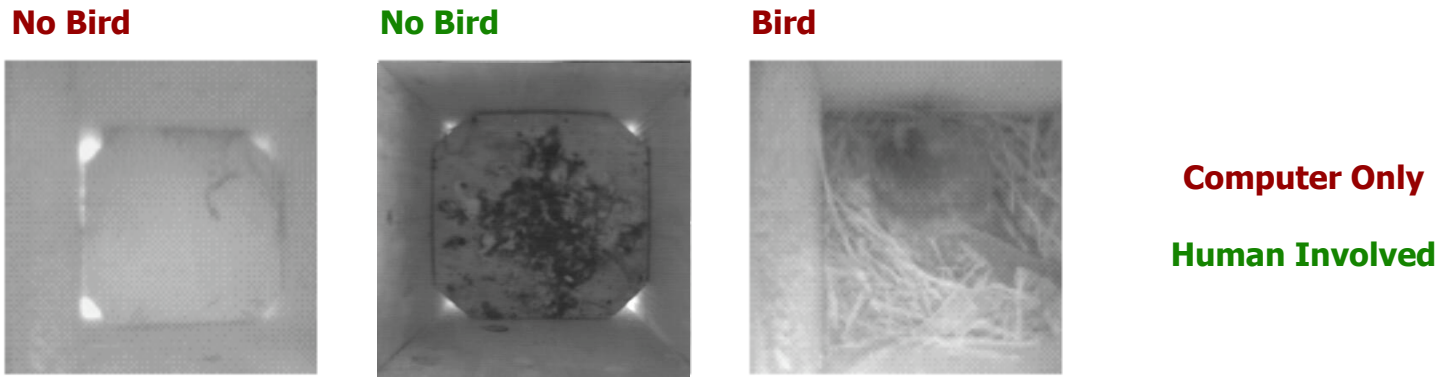
raw	corrected
87.1	87.1 ±0.1
?7.1	87.1 ±0.1
87.?	87.5 ±0.5

Prototypes

Republishers : Data Annotation

Images of nest boxes coming from James Reserve every 5 minutes.

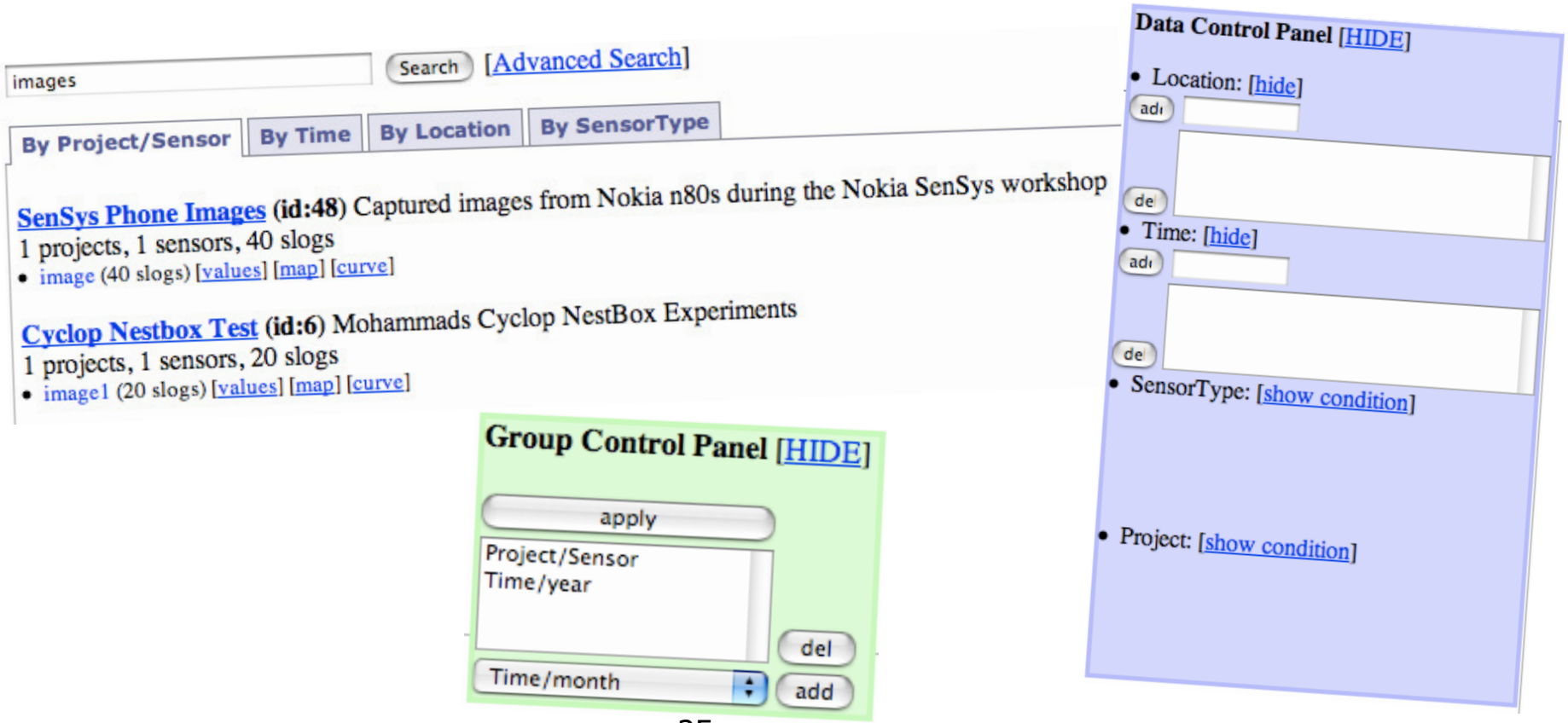
Republisher adds annotation information regarding the images in regards to whether a bird is present or not. Semi-automatic process where human is involved in ambiguous cases.



Prototypes

Sensor Search

Experimenting with providing a sensor search for SensorBase. Currently support primitive keyword based searching and exploratory search of data through UI features.



Conclusion

There is a need in the sensor network community for a ***common means to share sensor data***.

We are starting to explore this space and have come up with architecture primitives that would help in this process.

Components such as ***publishers, sensor stores, sensor search engines, and republishers*** connected using a ***common protocol*** are involved.

Numerous research challenges still exist, but if we provide a ***sound solution*** for these problems than there would be a ***great benefit*** for all users of sensor networks.

Questions?