

From Crowd to Cloud: Apply Named Data Networking Principles to Manage Air Sensor Data

Yifang Zhu, UCLA School of Public Health

Alex Afanasyev & Lixia Zhang, UCLA Computer
Science

New opportunities

- ◇ Technology advances in low-cost and portable sensors
- ◇ An increasing number of communities and individuals engaging in collecting and sharing sensor data

Crowdsensing by citizen scientists

New problems

As of today: (manual) data uploading into the cloud

- ◇ Data discovery
 - Data commonly named by ad hoc local conventions, incompatible data naming schemes
- ◇ Data provenance and authenticity
 - Not associated with data itself
- ◇ Data quality
 - Many different Sensor types
 - Untrained collector community

The Need

Developing user-friendly tools to engage researchers, communities, and citizen scientists to make such data “Findable, Accessible, Interoperable, and Reusable (FAIR)”.

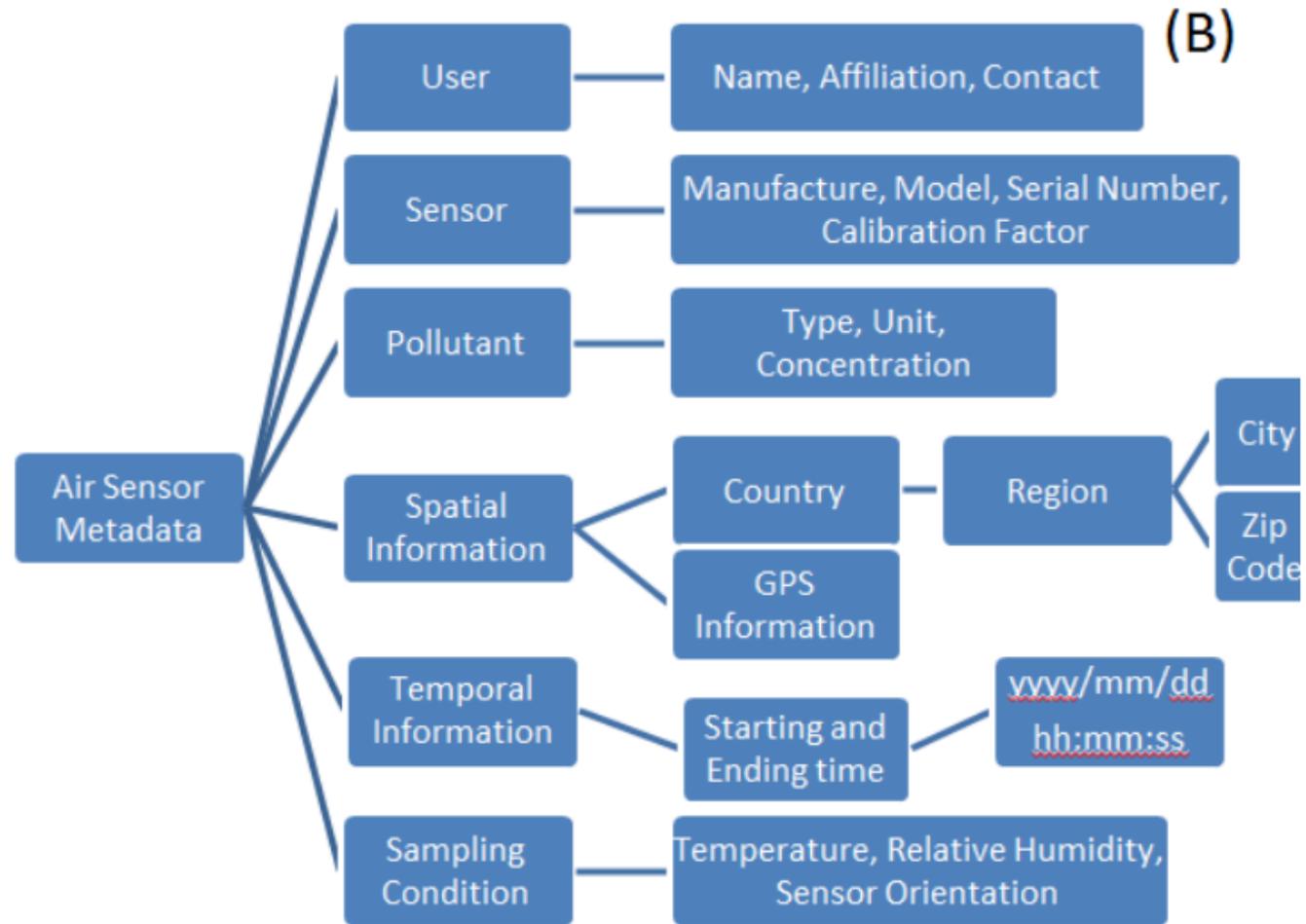
Proposed Solutions

- ◇ Establishing *systematic naming conventions* for sensing data
 - Time, location
 - Sensor type
 - Sampling condition
 - User community
- ◇ Securing data directly
- ◇ Rendezvousing collection and usage via standard data naming assisted by
 - Use of metadata
 - catalog lookup

Dylos Logger v 1.6.0.0 (A)
Unit: DC1700 v 2.08
Date/Time: 08/17/15 14:40

Particles per cubic foot

Date/Time, Small, Large
08/10/15 13:19, 83100, 3300
08/11/15 10:26, 127000, 2800
08/11/15 10:27, 124000, 1500
08/11/15 10:28, 118200, 1500
08/11/15 10:29, 108400, 1400
08/11/15 10:30, 105300, 2100
08/11/15 10:31, 100400, 1700
08/11/15 10:32, 102300, 1700
08/11/15 10:33, 103400, 1400
08/11/15 10:34, 102000, 1500
08/11/15 10:35, 102300, 1400
08/11/15 10:36, 98900, 1100
08/11/15 10:37, 101100, 1200
08/11/15 10:38, 102800, 800



Air-Data.net Cloud

[https://air-data.net/upload?
/Air/US/CA/Dylos/PM
/AQMD/historical
/2016-01-01T19:14:01](https://air-data.net/upload?/Air/US/CA/Dylos/PM/AQMD/historical/2016-01-01T19:14:01)



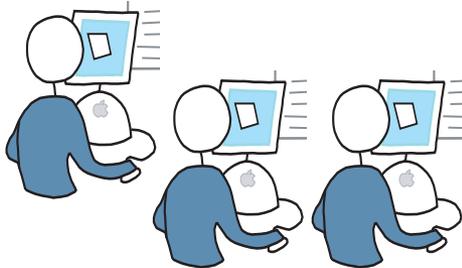
Catalog



Storage

Upload data

Name, assign metadata,
and sign data



Citizen scientists crowd

The screenshot shows a web interface for uploading data. At the top, there are tabs for 'Find', 'Upload', and 'Account'. The 'Name' field contains the URL: `/Air/US/CA/<SENSOR>/PM/<TYPE>/historical/2016-01-01T19:14:01`. Below this, there are two dropdown menus: 'SENSOR' and 'TYPE'. The 'SENSOR' dropdown is currently open, showing options: 'S-300', 'S-300E', 'T-110', and 'T-110-3V'. To the right of these dropdowns is a 'Captured By' field with a placeholder icon of a person. At the bottom, there is a 'Meta Data' section with a grid of 18 checkboxes arranged in two rows of nine.

Air-Data.net Cloud



Catalog



Storage

[https://air-data.net/upload?
/Air/US/CA/Dylos/PM
/AQMD/historical
/2016-01-01T19:14:01](https://air-data.net/upload?/Air/US/CA/Dylos/PM/AQMD/historical/2016-01-01T19:14:01)

[https://air-data.net/discover?
q=/Air/US/CA
&type=CO2&year=2016&....](https://air-data.net/discover?q=/Air/US/CA&type=CO2&year=2016&...)

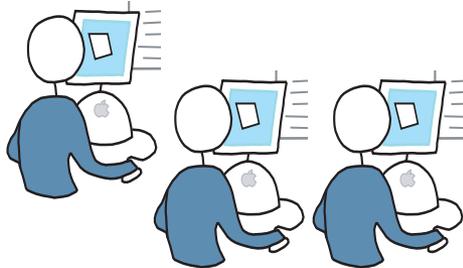
[https://air-data.net/
/Air/US/CA/Dylos/PM
/AQMD/historical
/2016-01-01T19:14:01](https://air-data.net/Air/US/CA/Dylos/PM/AQMD/historical/2016-01-01T19:14:01)

Upload data

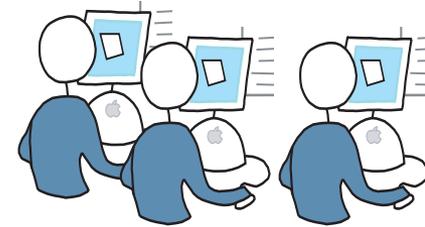
Discover available data

Name, assign metadata,
and sign data

Download and authenticate data



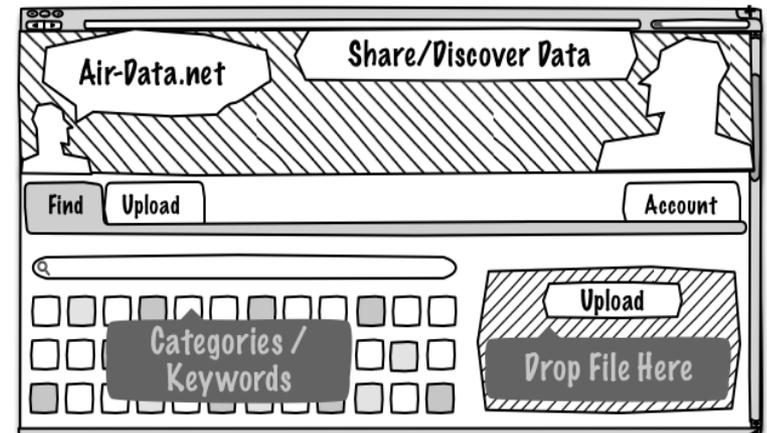
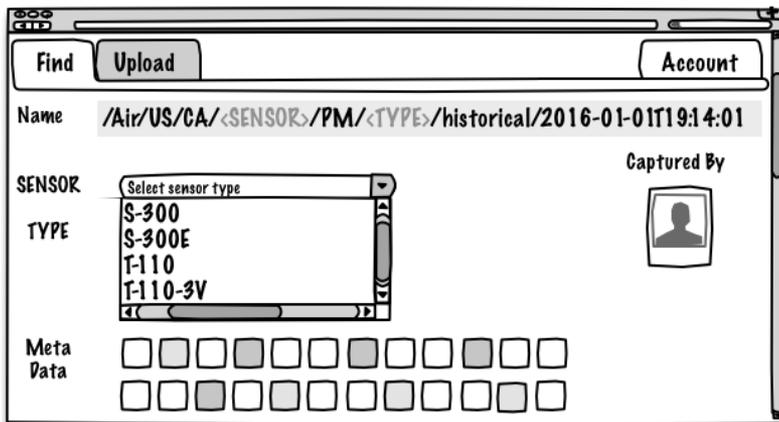
Citizen scientists crowd



Citizen scientists crowd



Data analysis



A quick starting point

- ◇ Use existing networking stack
- ◇ Data byte piling → Web-based pub/sub

Challenges

- ◇ Naming / naming conventions
 - What goes to names / metadata
 - Working with domain experts to develop
- ◇ Identity management
- ◇ Usable security

Tasks

- ◇ Data naming toolkit for user devices
- ◇ Web app development

Future steps: moving “named data” down the protocol stack

- ◇ NDN into user devices
- ◇ Distributed storage/rendezvous
 - Networking: naming data makes network transparent
 - Transport: requires distributed dataset synchronization

Take Away

- ◇ Data generation by community at large, at unprecedented scale
 - a paradigm shifting into future environmental health research
- ◇ Establishing systematic naming convention: a critical step towards data intensive applications
- ◇ Making security an integral ingredient into data