Sensing-as-a-service and Big Data

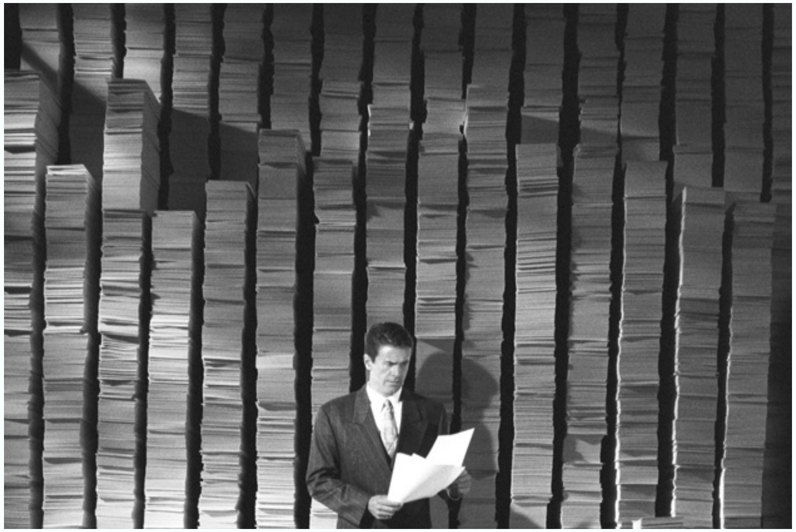
Arkady Zaslavsky, Charith Perera,
Dimitrios Georgakopoulos
CSIRO, Australia

www.csiro.au

Advances in Cloud Computing, Bangalore, 28 July, 2012

Outline

1. IoT
2. Sensors, sensor networks, sensing-as-a-service
3. Big Data
4. EU FP7 OpenIoT
5. CSIRO projects
6. Conclusion



CSIRO today: a snapshot

Australia's national science agency

One of the largest & most diverse in the world

6500^{+} staff over 55 locations

Ranked in top 1% in 14 research fields

150^{+} spin-offs based on our IP & expertise

170^{+} active licences of CSIRO innovation

Building national prosperity and wellbeing

National Research Flagships

Climate Adaptation

Light Metals

Sustainable Agriculture

Energy Transformed

Minerals Down Under

Water for a Healthy Country

Food Futures

Preventative Health

Wealth from Oceans

Future Manufacturing

Future Internet

Internet of Things IoT

A dynamic global network infrastructure with self configuring capabilities

based on standard & interoperable communication protocols

where physical & virtual “things” have identities, physical attributes,

virtual personalities, use intelligent interfaces, and

are seamlessly integrated into the information network.

Internet of Things

Imagine a world where:

- your car knows where the traffic jams / road anomalies are
- your fridge knows how long before the milk expires
- you know the areas with the less pollution where you can jog freely
- energy resources can be managed efficiently

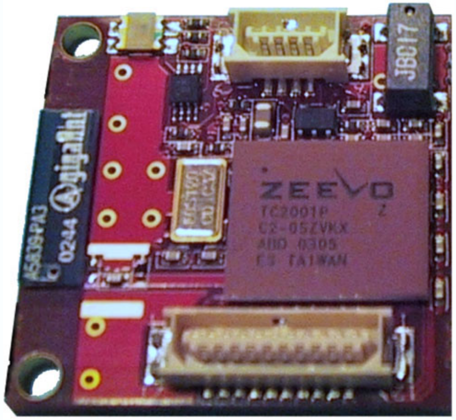
And the list goes on forever!

Existing WSN Devices

Wikipedia lists 31 Sensor Network devices

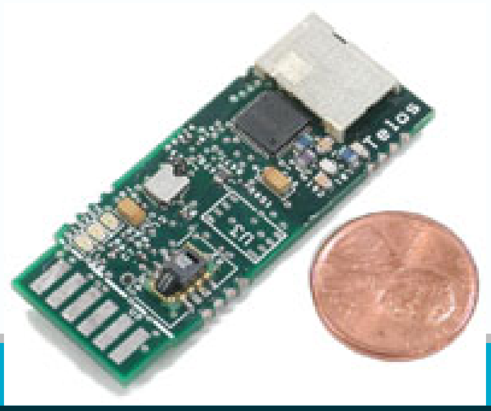
Sensor Network Museum lists 30 devices
http://www.snm.ethz.ch/Main/HomePage

IMote 2.0:

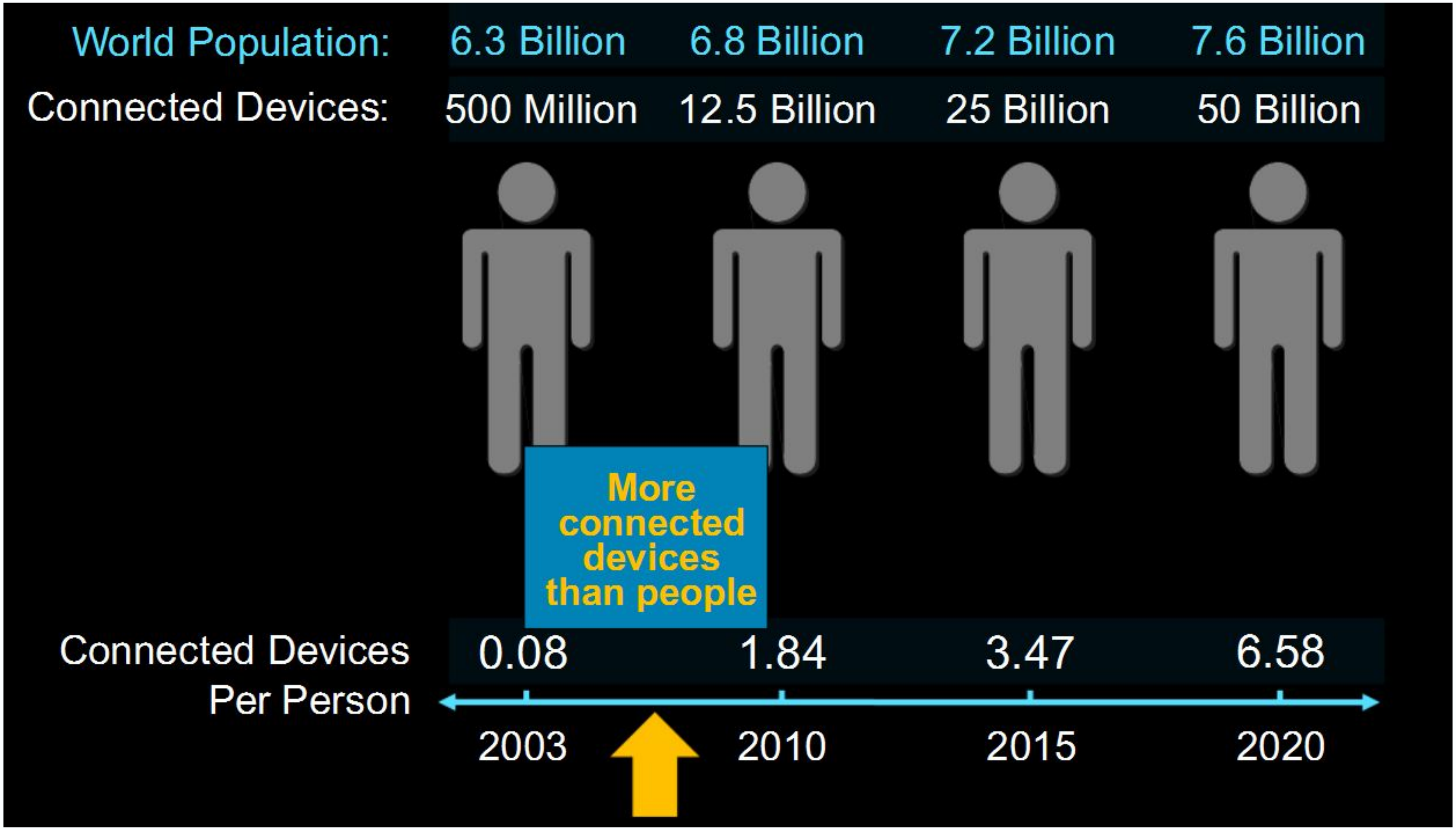


MicaZ:

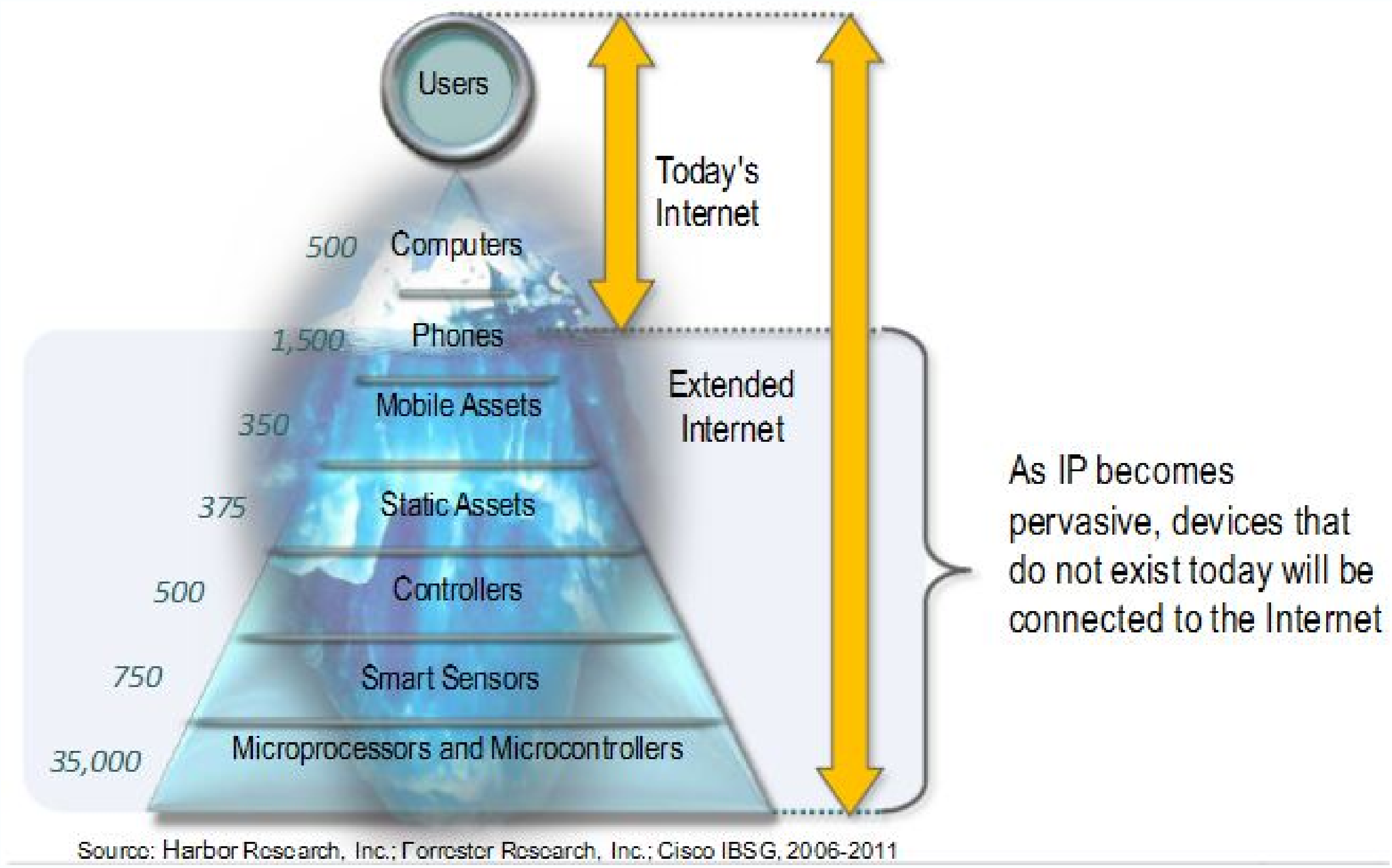
TelosB:



IoT is Already Here



The Internet Will Extend to Billions of Devices



Wireless and Sensors Will Be Everywhere

90% of world's population has wireless connectivity

100,000 wireless network masts are erected annually

1 billion electronic devices equipped with WiFi shipped in 2012

Billions of smart sensors

- Cisco2009: PlanetarySkin—integrate sensors on land, in sea, in air, and in space to help make it possible to see the whole picture when it comes to the effects to and changes in the environment
- HP2010: Central Nervous System for the Earth (CeNSE) – 10-year mission to embed up to a trillion push-pin-sized sensors and actuators around the globe

Source: Cisco IBSG, UN: International Telecommunications Union, Real-Aliens. Com 2006-2011

The “Zettaflood” is just the Beginning of the IoT Traffic

Total IP Traffic on the global Internet:

- 2003-1.8 Petabytes
- 2007- 161 Exabytes
- 2009- 487 Exabytes
- 2010- ½Zettabyte
- 2011- 1 ZettaByte (540,000 X increase from 2003)

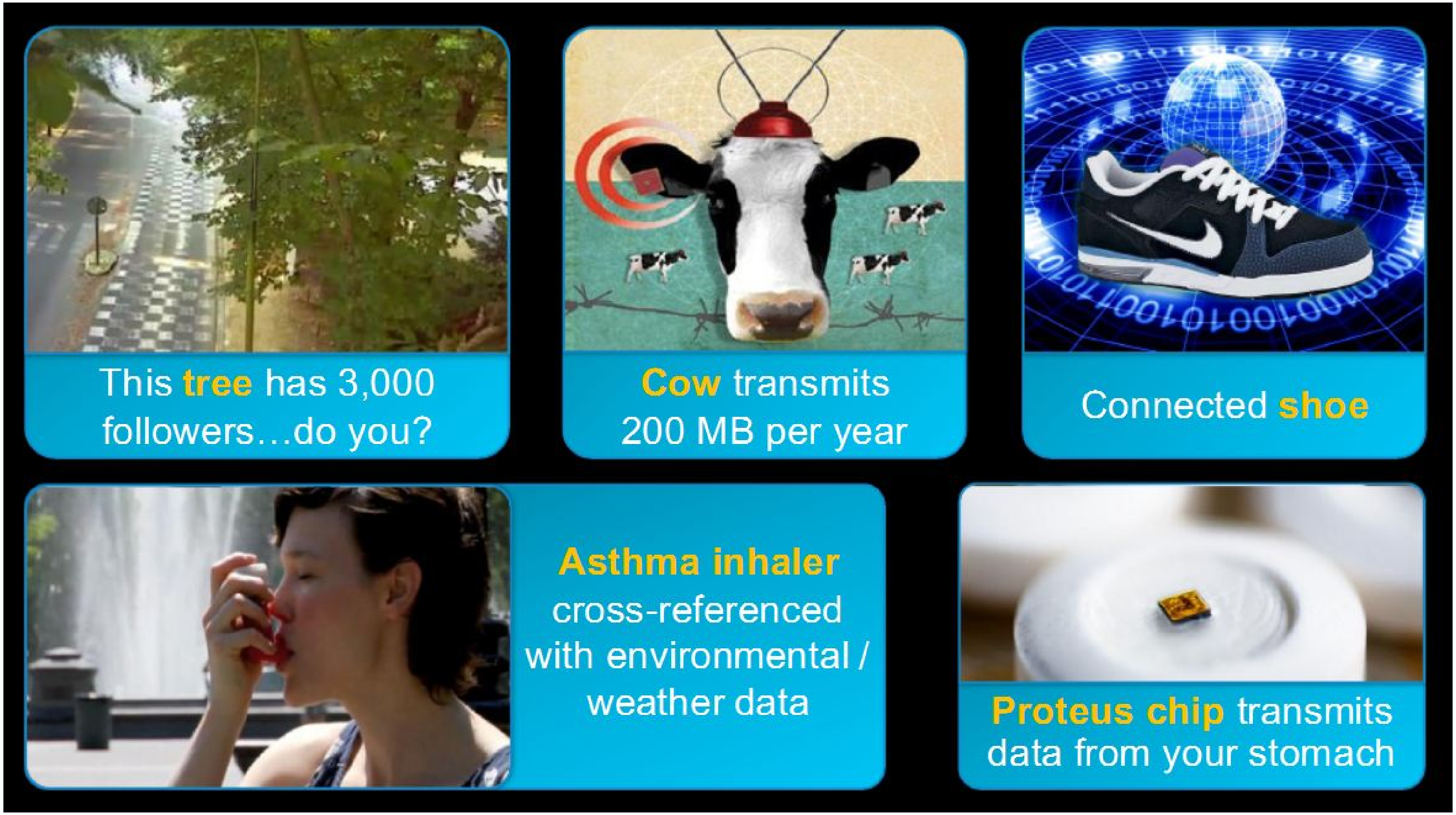
> “By 2011, 20 typical households will generate more traffic than the entire Internet did in 2008.”
>
> Jim Cicconi, VP, AT&T

Expected to double over the next 18 months

2012- 91% expected to be video

Source: VentureBeat, IDC, C|Net, TheGuardian, UK

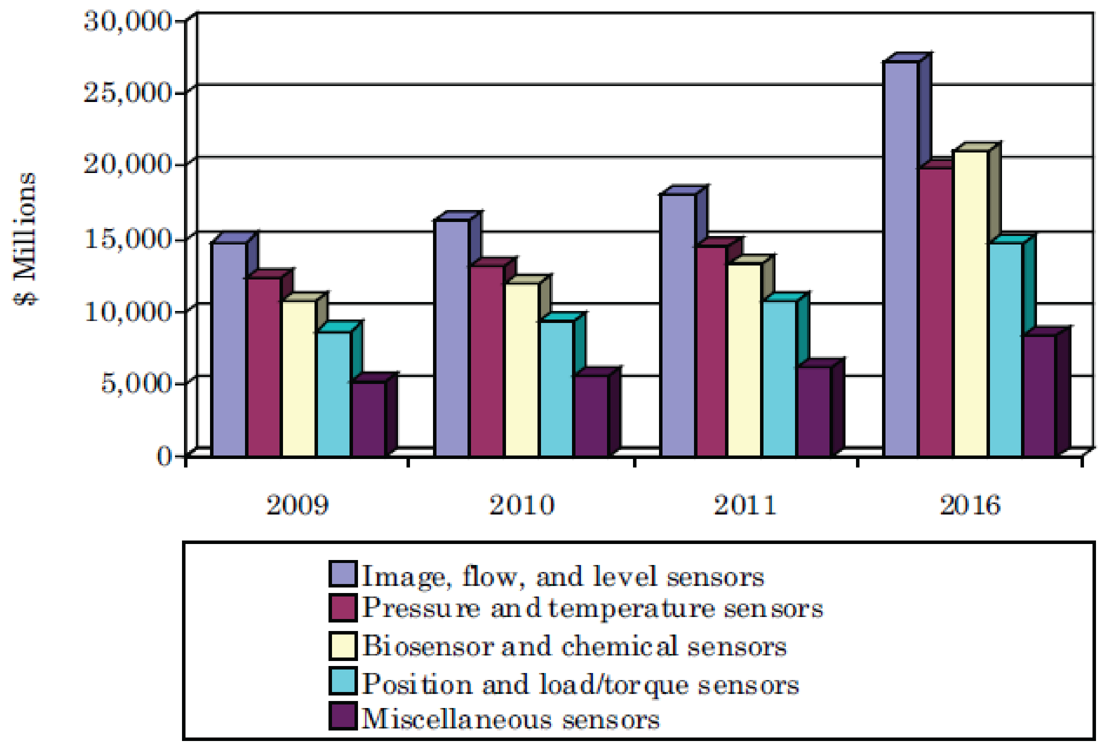
New Internet Inhabitants



Sensors: Technologies and Global Markets

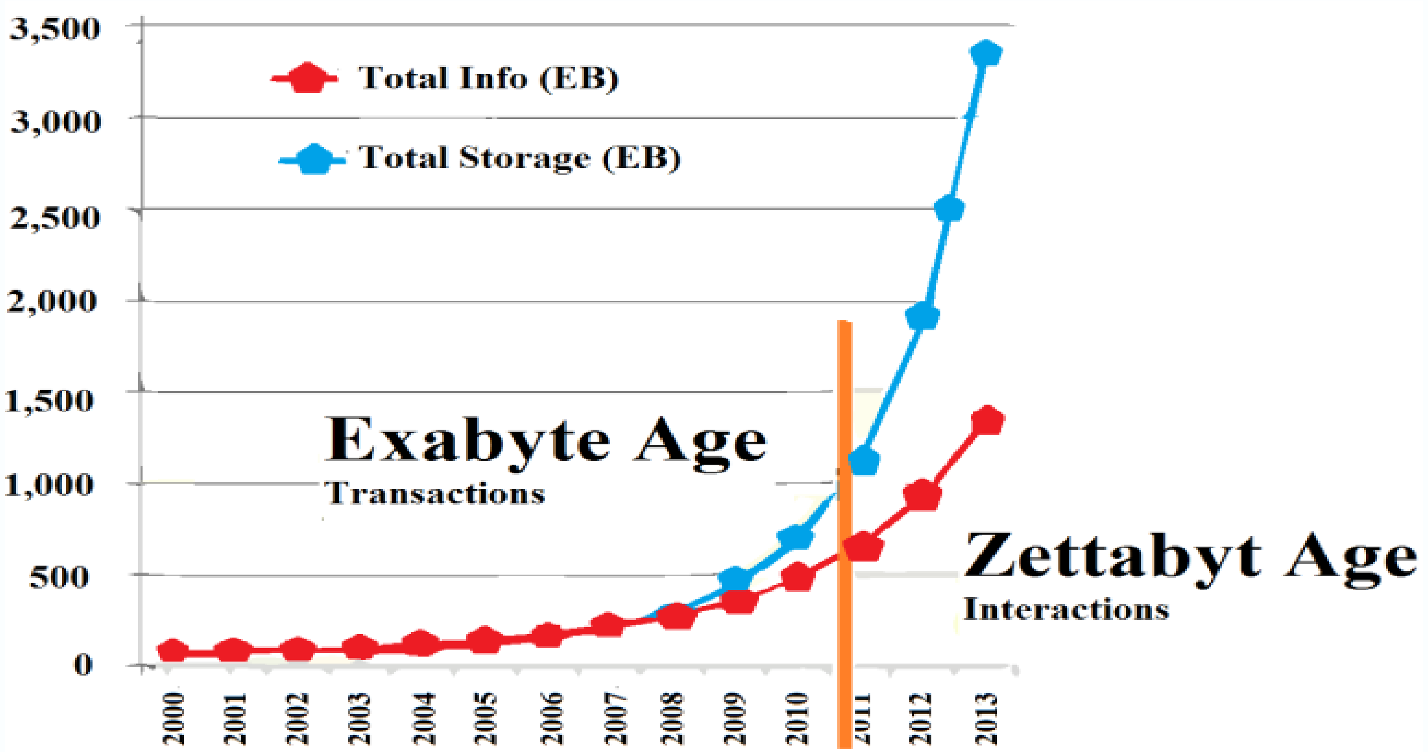
The market for biosensors and chemical sensors is expected to experience the highest growth, at a compound annual growth rate (CAGR) of 9.6% during the 5-year period from 2011 to 2016. This sector is expected to be worth $13 billion in 2011 and $21 billion in 2016.

SUMMARY FIGURE
GLOBAL MARKET FOR SENSORS, 2009-2016
($ MILLIONS)



Source: BCC Research

Big Data



The total amount of data generated on earth exceeded one Zettabyte (ZB) in 2010. It is predicted that data volume will grow exponentially as depicted (www.teradata.com)

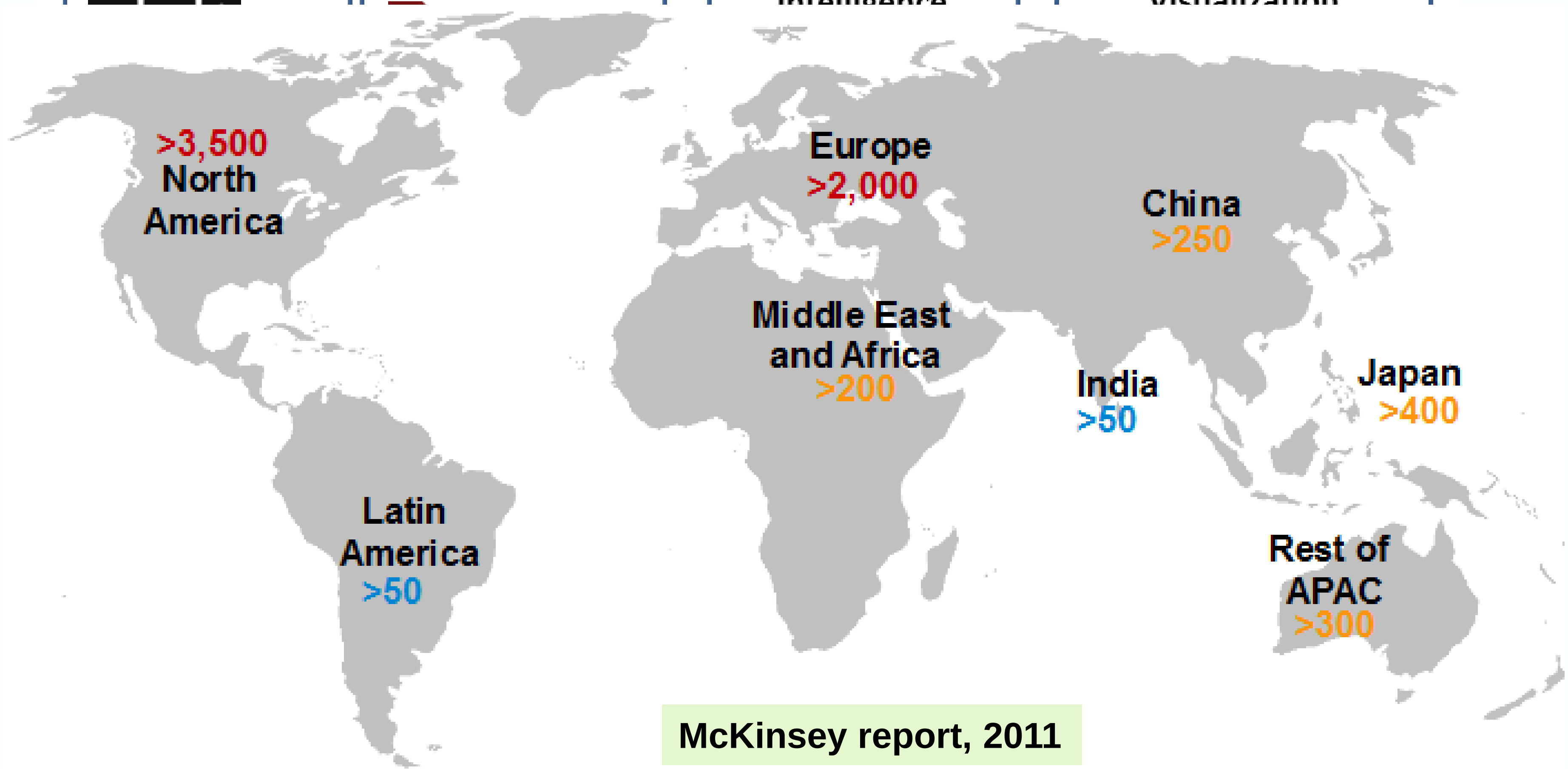
Big Data Landscape

Log Data Apps

Vertical Apps

Business Intelligence

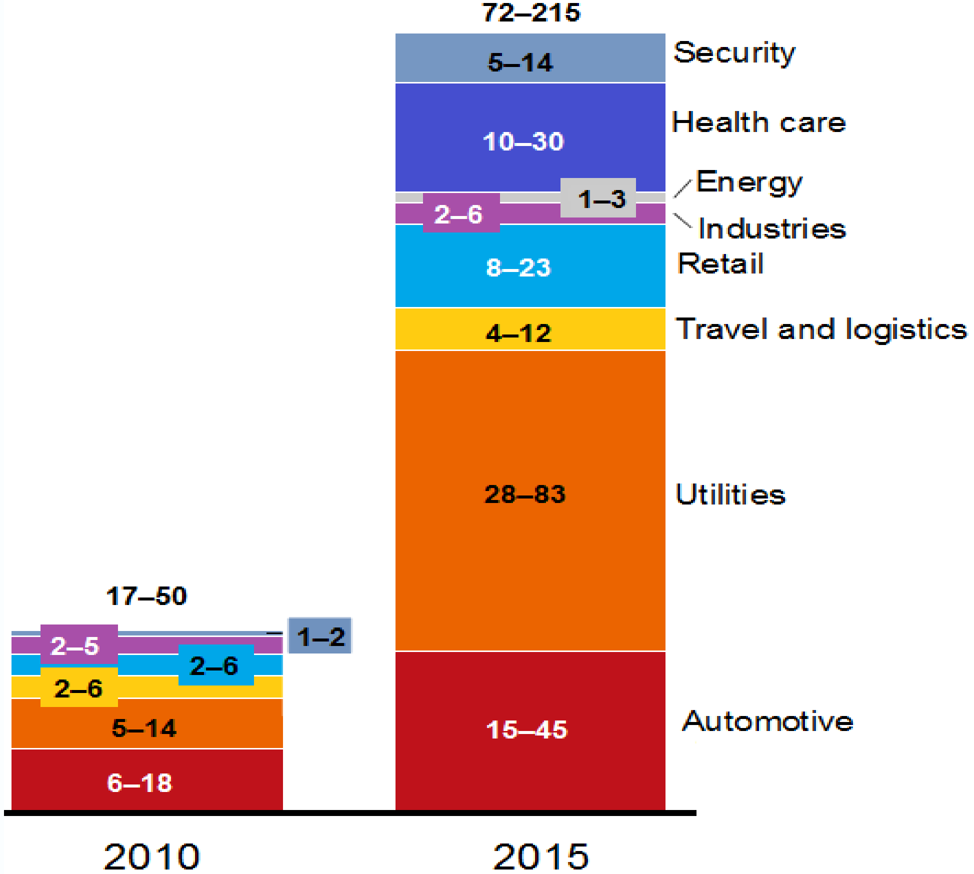
Analytics and Visualization

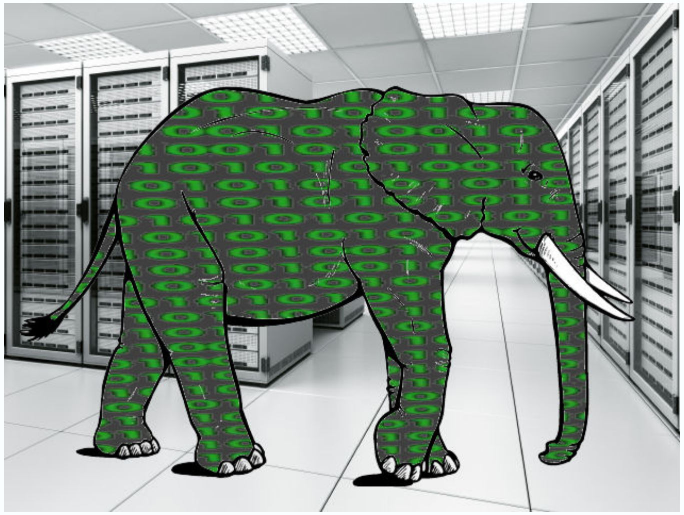


dave@vcdave.com

http://blogs.forbes.com/davefeinleib/

IoT & Big Data





Data generated from the Internet of Things will grow exponentially as the number of connected nodes increases. Estimated numbers of connected nodes based on different sectors are presented in Millions

CHALLENGES IN BIG DATA MANAGEMENT

- ❑ High volume of processing using low power consumed digital processing architecture.
- ❑ Discovery of data-adaptive Machine learning techniques that can analyse data in real-time.
- ❑ Design scalable data storages that provide efficient data mining.

EU FP7 OpenIoT Objectives

The Motivation:

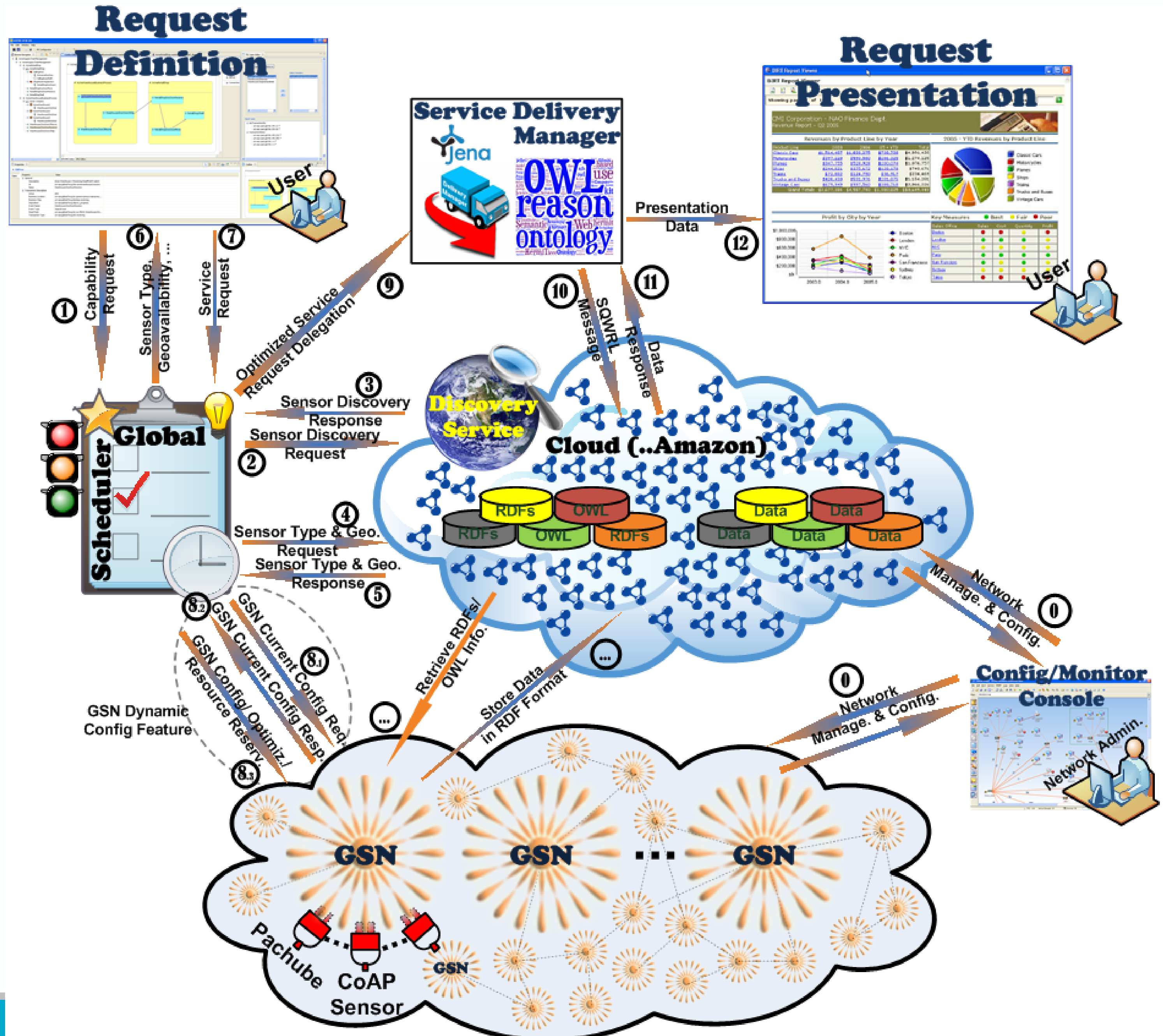
- Despite the proliferation of pervasive grids, participatory sensing and on-demand sensing services (e.g., «Location-as-a-Service») **there is no easy (and generic) way to offer utility based IoT services**
- IoT end-users and providers still need to deploy their own sensors and devices

The Open IoT Goal:

- Research and provide **an open source middleware framework** enabling the dynamic, self-organizing formulation of self-managing cloud environments for IoT applications
- Sensing-as-a-Service
- Converge IoT and IoS - Cloud Computing

www.openiot.eu

OpenIoT High Level Architecture
(Simple Example with dynamic GSN config)

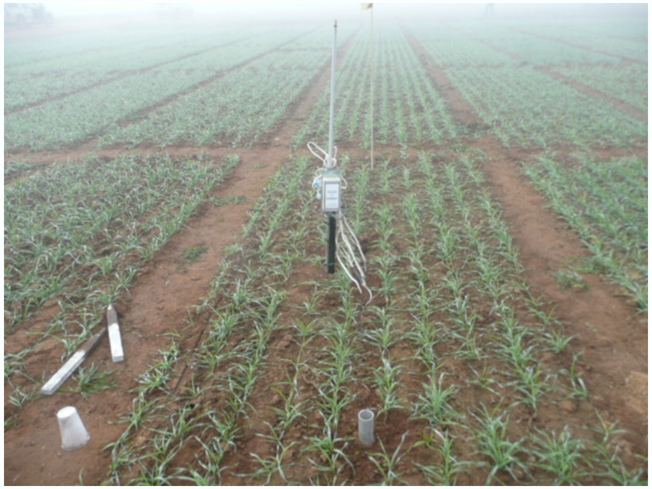


CSIRO

OpenIoT experimental test-bed

For the High Resolution Plant Phenomics Centre's Phenonet project

- Measure environmental and plant physiology parameters in the field
- Improve the quality and scale of data available to plant breeders from grain trial plantings



IE Lab contribution

- Design and programming of sensor network
- Testbed for declarative programming of sensor networks
- Fast browser-based data display & analysis using reusable components
- See http://phenonet.com

CSIRO

What is GSN ?

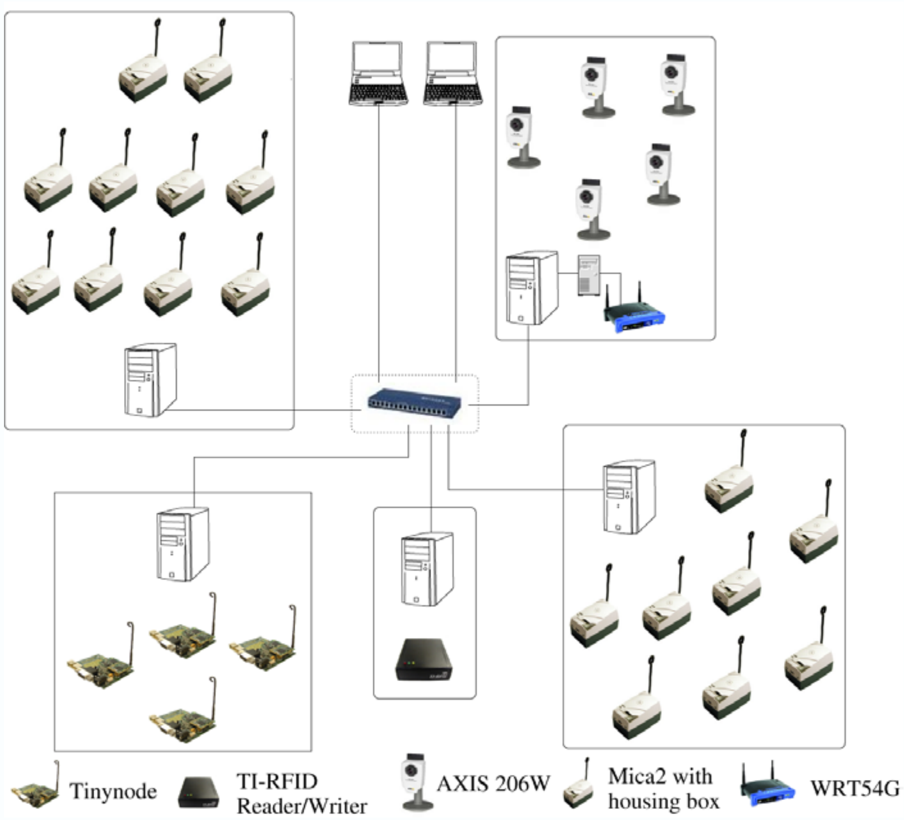
Sensor information management solution

Runs on base computers (Java)

Integrates data streams with R/Matlab,…

Publish/Subscribe with continuous data processing

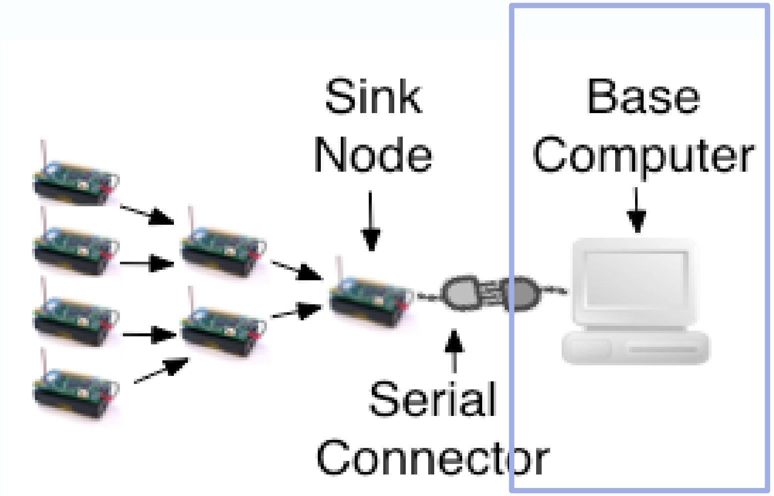
Many other advanced features…



OpenIoT is extending and transforming GSN to a Cloud Service

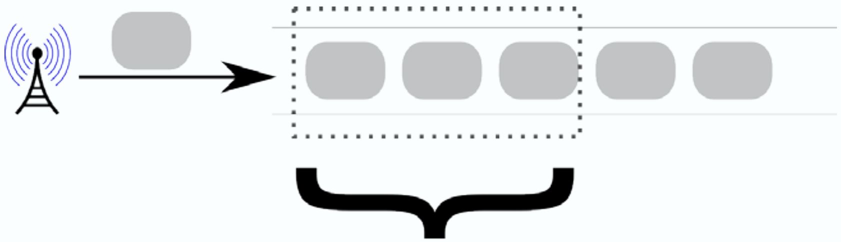
GSN Focus

Acquire data from sensors, RFIDs, motes, cameras, and other information sources (anything that can produce data streams)



Stream Processing Engine

- Window-based processing
- Example:
  Sliding Value 1, Window Size 3

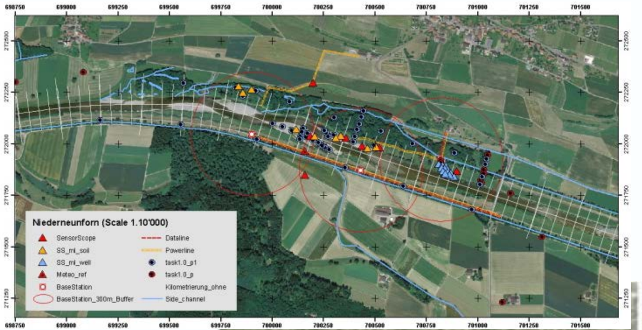


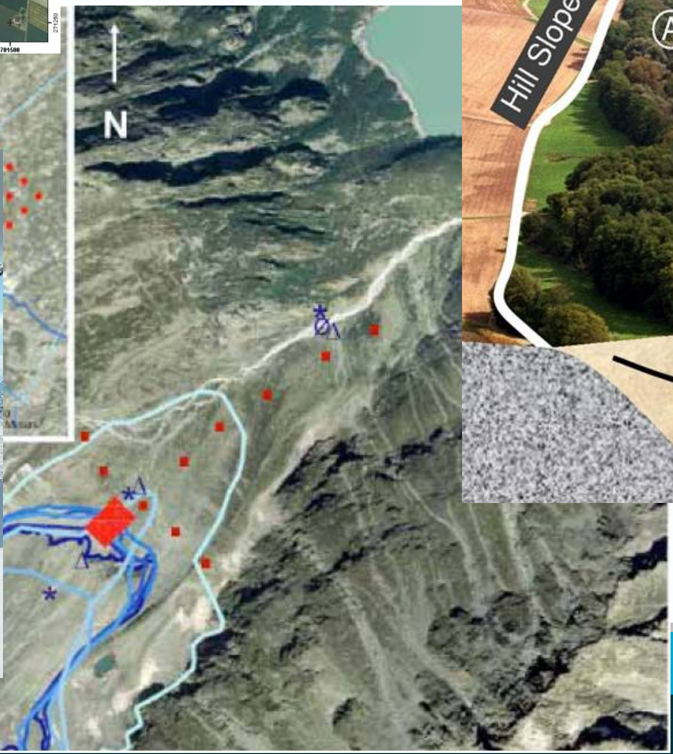
Publish/Subscribe services

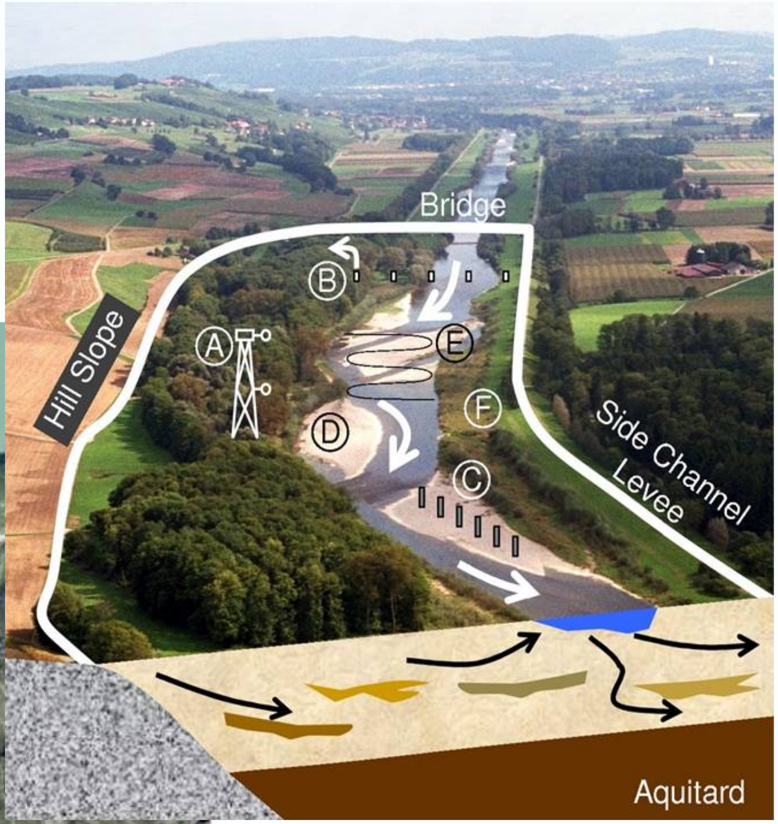
Archiving

Deployments of GSN

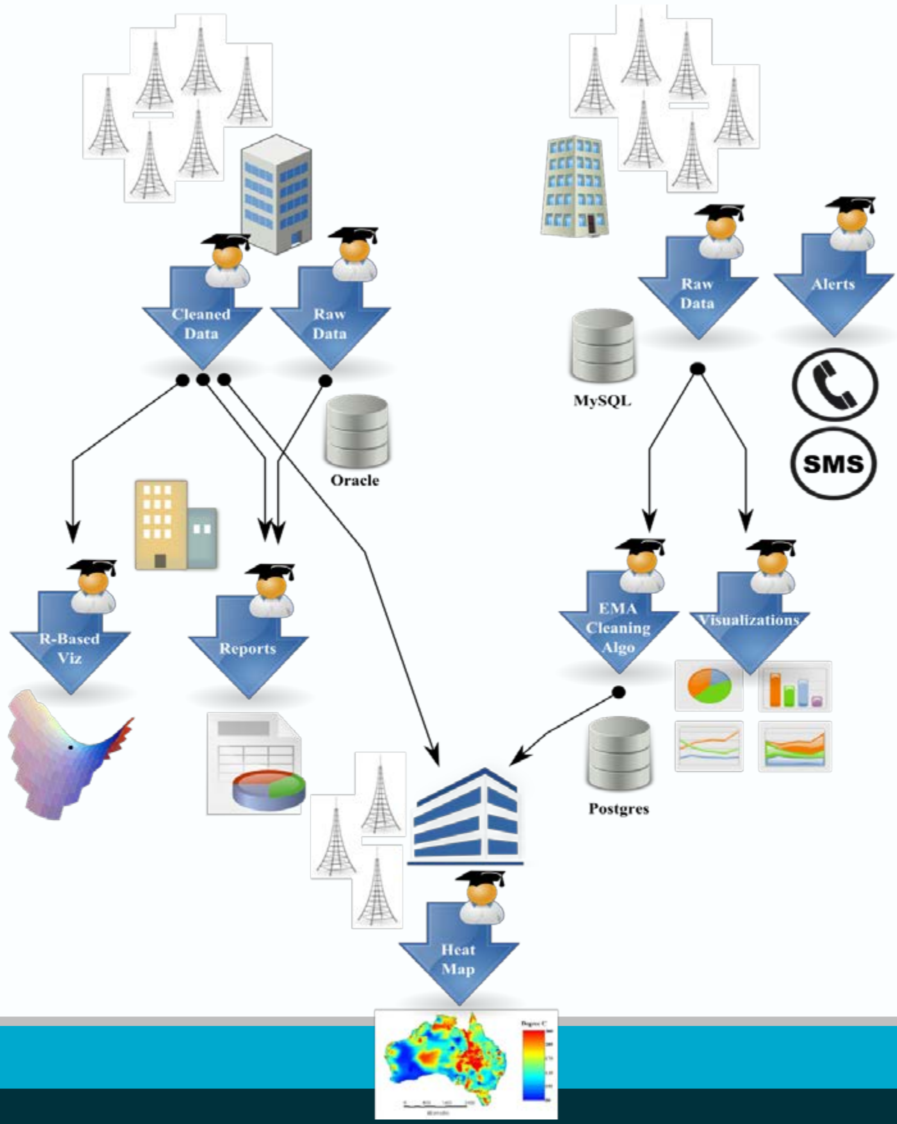
Some of the Existing Deployments





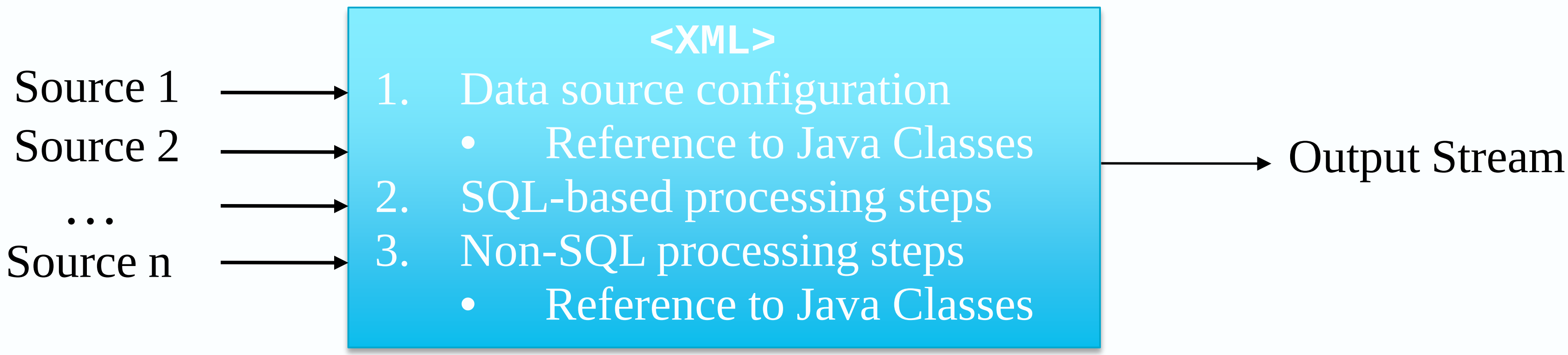


Building GSN Applications



- **GSN Applications are collections of interconnected Virtual Sensors**
- **Example:**
  - 9 Virtual Sensors
  - Each virtual sensor has to configure its data source
  - Publish/Subscribe Model
  - Continuous Queries

GSN Virtual Sensors



Virtual sensor are specified by XML scripts

Include stream data processing steps specified via SQL-like queries

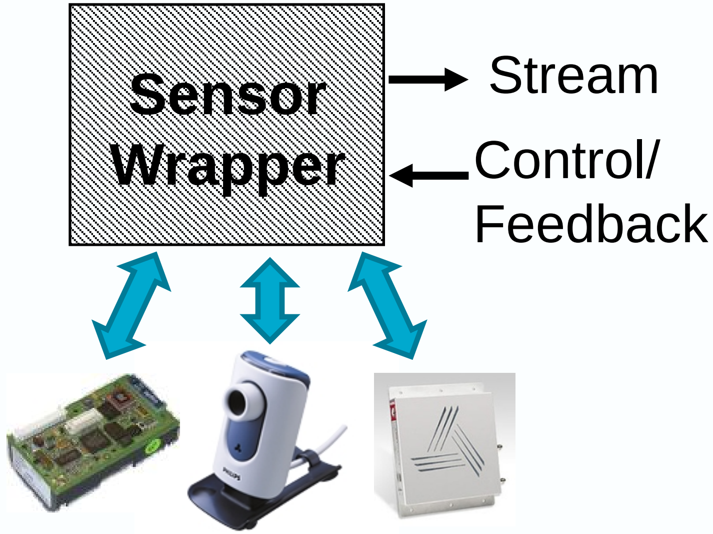
Include References to Java classes of additional operators, including

- Configuration of data steam sources via provided wrappers
- Non-relational stream data processing operators, e.g., for spatiotemporal processing, alerting, for data cleaning, visualization

Flexible Data Acquisition for Virtual Sensors

GSN Sensor integration support wrappers

- GSN provides wrappers and APIs for adding more
- Wrappers produce streams
- Interact with sensors
  - Wrappers consume control cmds



20+ Wrappers already available

- TinyOS, CSV, HTTP-Get, Serial Port, UDP, RSS Feeds, …

An introduction on how to write a new wrapper

- sourceforge.net/apps/trac/gsn/wiki/writing-wrapper
- sourceforge.net/apps/trac/gsn/wiki/template-wrapper

CSIRO Applications

Sensing the crop environment and performance

Wireless sensor network:

- Nodes with sensors, microcomputer and radio
- Self configuring mesh network

Two types of Nodes

- Plot nodes and weather nodes
- Sensors: rain, hail air temperature, pressure, relative humidity, wind speed and direction, etc.

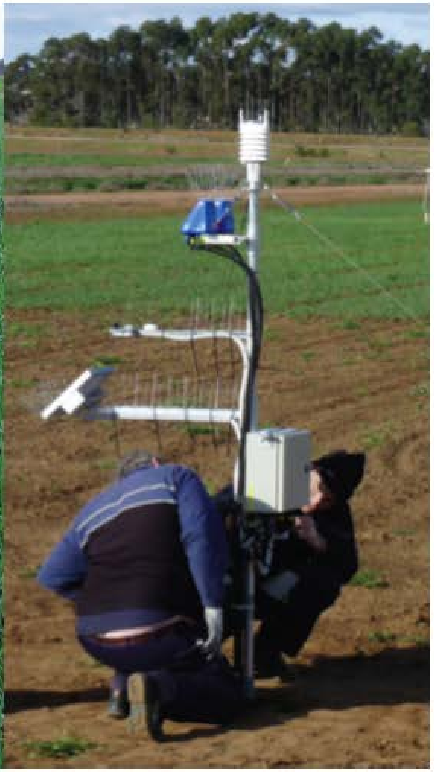
Time resolution in the order of minutes

Spatial resolution – per plot where needed

- "Golden Data Points" are captured
- Site visits are transformed to virtual (online) site visits

The Phenonet Network

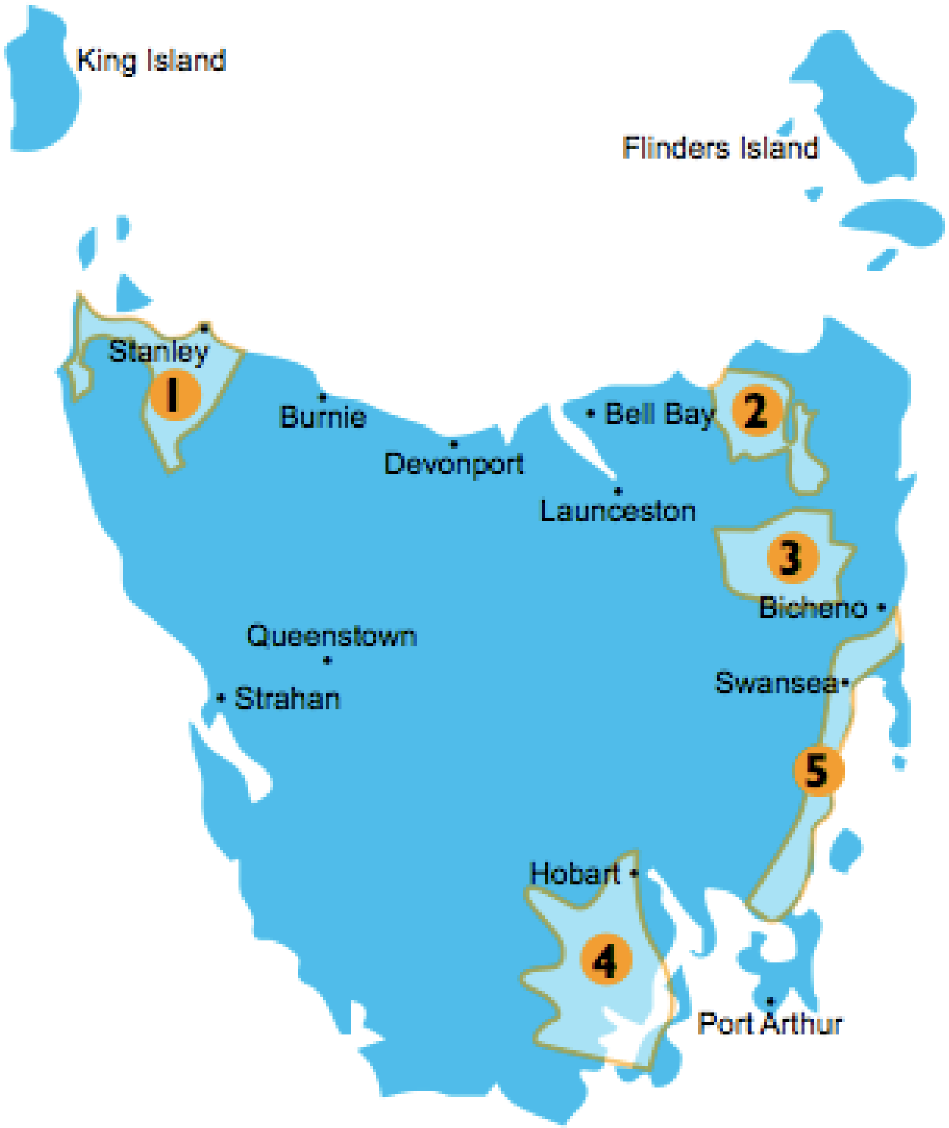
Plot Nodes



Weather Nodes

Priority areas

1. Circular Head – NW Tasmania
   Agriculture (dairy), aquaculture, carbon markets
2. Scottsdale – NE Tasmania
   Agriculture (dairy, fruit and viticulture), food logistics, carbon markets
3. South Esk – Northern Midlands, NE Tasmania
   Catchment management, flood prediction, carbon markets
4. Triabunna – East Coast
   Agriculture (pasture, fruit and viticulture), aquaculture, food logistics, carbon markets
5. Huon Valley – SE Tasmania
   Aquaculture, food logistics, carbon markets, catchment management



Aquaculture monitoring

- Strategic relevance
  - Improve productivity of the salmon and oyster aquaculture industry through provision of real-time awareness of critical environmental variables.
  - Build supply chain resilience to natural events (e.g. harmful algal blooms) that may force temporary closure of some production facilities.

- Impact
  - Tasmanian aquaculture industry groups (Oysters Tasmania and Tasmanian Salmon Growers Association)
  - Tasmania Department of Human and Health Services

CSIRO

Aquaculture monitoring

- Science challenges
  - Federation of existing aquaculture data including sensor observations, historical data, and model outputs.
  - Develop incremental learning model to analyse cause-effect relationships. Decision support tools capable of predicting events that may potentially impact productivity.

Smart rural infrastructure

- Strategic relevance
  - Real-time sensing and modelling of environmental and irrigation crop parameters to:
    - Increase water availability
    - Improve water use efficiency
    - Boost agricultural productivity without compromising critical ecosystem services.
- Impact
  - Tasmanian Irrigation (Lower South Esk Irrigation Scheme).

Smart rural infrastructure

- Science challenges
  - Federation/querying of of diverse sensor data streams.
  - Model-sensor integration, framework for managing computer simulation models used to predict weather, stream flow, crop production, and irrigation demand.
  - Provenance-based model-sensitivity analysis.
  - Decision support tools that will maximise irrigation water use efficiency and automate irrigation.

Context-based disaster alerts

- Strategic relevance
  - Delivers on key objectives in the COAG National Disaster Resilience Statement.
  - Improves government and societal capacity to deal with disasters.
- Impact
  - Tasmania State Emergency Services (SES).
  - Local communities.
  - Utilities.

Context-based disaster alerts

- Science challenges
  - Alerting system is dependent on information from many different sources including live sensor feeds, spatial feature services, and output from predictive models.
  - Representing context information in a unified manner
  - Developing a methodology for analysing different contexts
  - Predicting causes and effects
  - Providing context-aware alerting services.

Mobile Crowd-sensing Applications

- Environmental – monitor the phenomena related to the natural environment
  - pollution levels (Common Sense[1]), water levels in creeks (CreekWatch[2]), wildlife habitats
- Infrastructure - measurement of large-scale phenomena related to public infrastructure
  - traffic congestion and road conditions (CarTel [3], Nericell [4]),
  - parking availability (ParkNet [5]),
  - outages of public works (e.g., malfunctioning fire hydrants, broken traffic lights),
  - real-time transit tracking

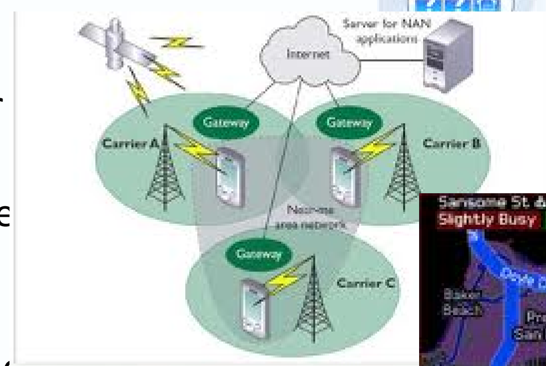
Mobile Crowd-sensing Applications

❑Health and well-being – monitor health and well-being of groups of people

- individuals measure location and bike route quality (e.g., CO_2 content on route, bumpiness of ride), and aggregate the data to obtain the “most” bikeable routes (BikeNet [6])
- individuals take pictures of what they eat and share it within a community to compare their eating habits (DietSense[7]) – community of diabetics

❑Social networking

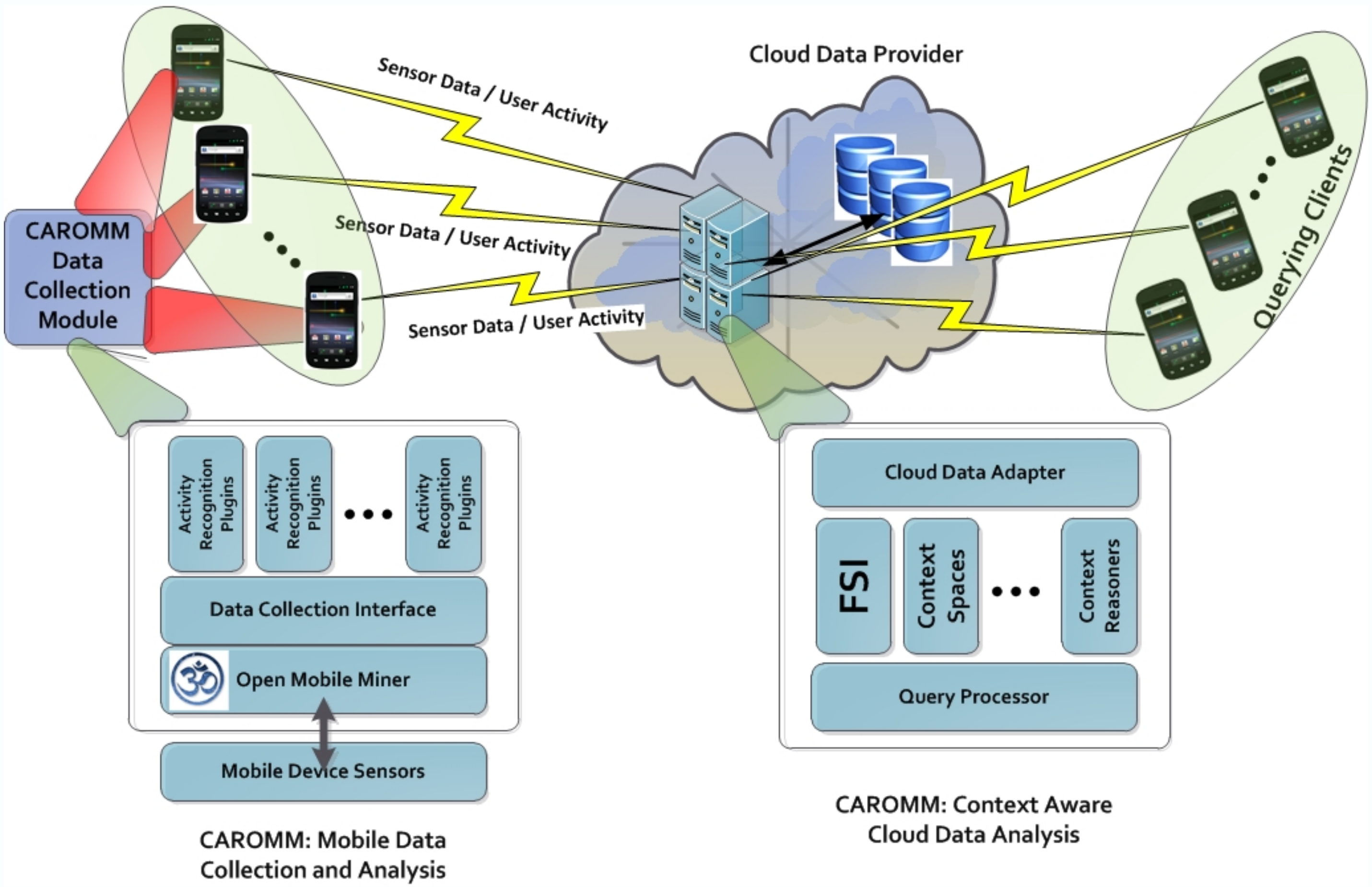
- user activity recognition and sharing with their social networks (Cenceme[8])
- social introduction /matchmaking based on use profile information (Serendipity [9], WhozThat [10])
- group-based, context-aware recommendation c music/videos (SocialFusion [11])
- visualization of mobile user concentration in an area using users’ GPS locations (Citysense [12])



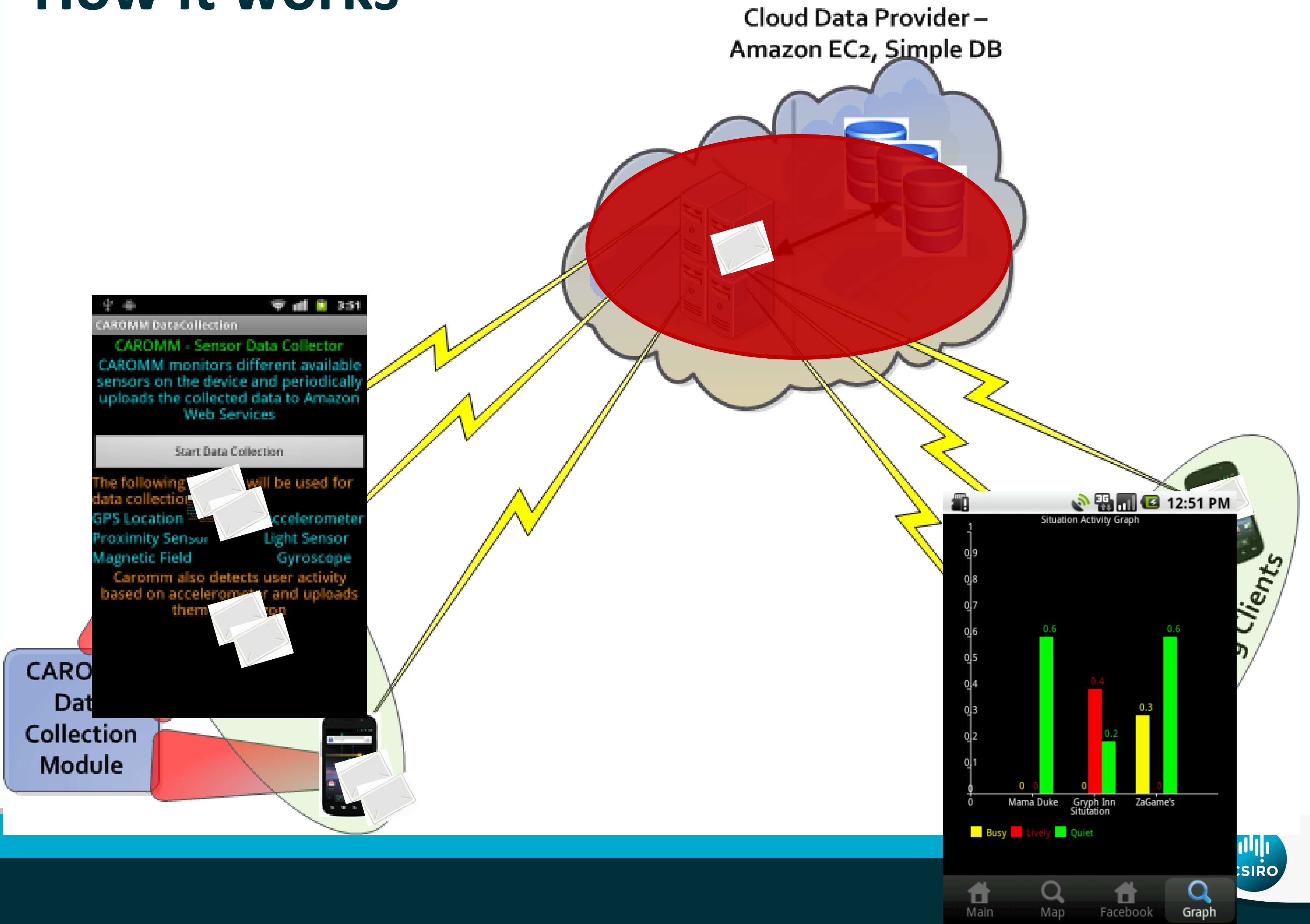
MCS Applications- Issues

- Require mobile devices to continuously sense, process and upload sensed data to the cloud/remote servers
- Data collection should be *cost-efficient* for both the *devices* and the *networks*
    - Local analytics
- Require *infrastructure* to *receive, manage and analyse* large volumes of *real-time data streams* using cloud computing platforms
- Aggregate analytics
- Require *participation* from the user and *willingness* to allow *collection of sensor data*
    - user preferences
    - privacy
    - incentive

Here-n-Now: A CAROMM application

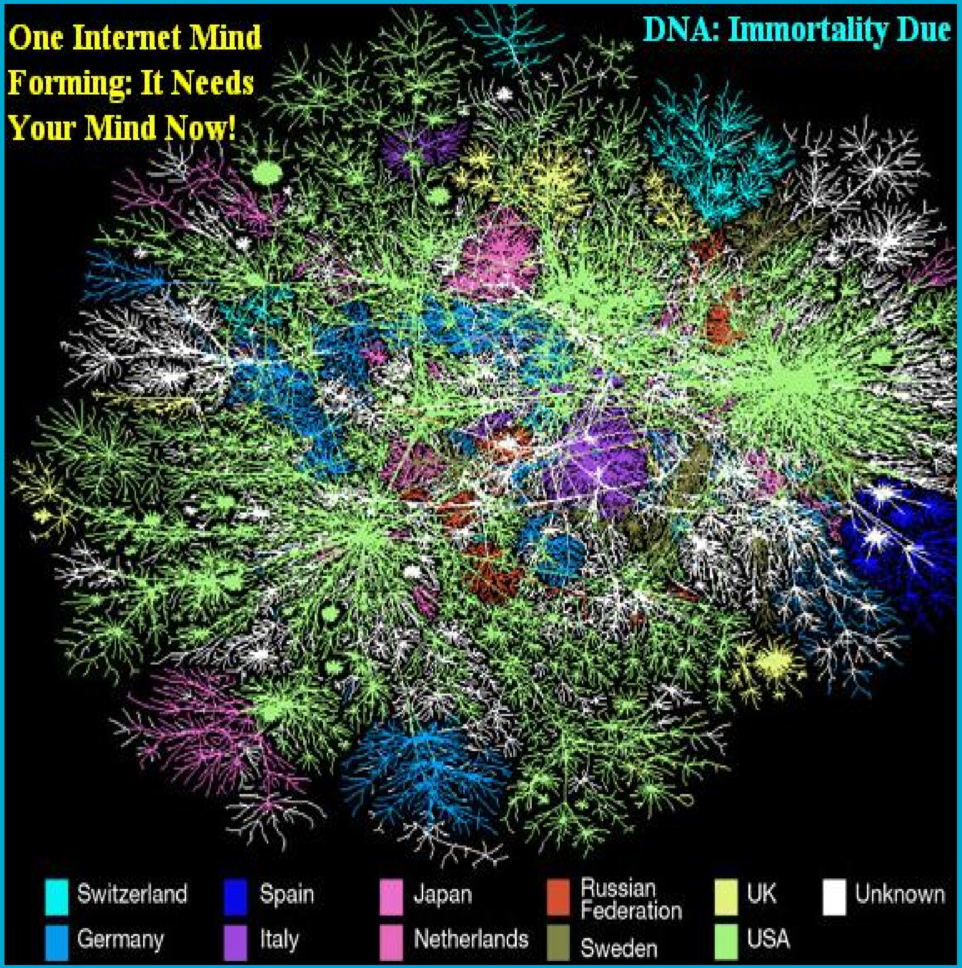


How it works



Cloud Data Provider – Amazon EC2, Simple DB
CAROMM DataCollection
CAROMM - Sensor Data Collector
CAROMM monitors different available sensors on the device and periodically uploads the collected data to Amazon Web Services
Start Data Collection
GPS Location
Proximity Sensor
Light Sensor
Magnetic Field
Gyroscope
Data Collection Module
12:51 PM
Situation Activity Graph
Mama Duke
Gryph Inn Situation
ZaGame's
Busy
Lively
Quiet
Main
Map
Facebook
Graph
Clients

Thank you !



Dr Arkady Zaslavsky, Professor

Science Leader in Semantic
Data Management, CSIRO, Australia

Phone: 02 6216 7132

Email: arkady.zaslavsky@csiro.au

AND

Research Professor

Luleå University of Technology, Sweden