The Sensor Web: Unpredictable, Noisy, Loaded with Errors

Alan F. Smeaton
CLARITY: Centre for Sensor Web Technologies
Dublin City University

science foundation ireland
fondúireacht eolaíochta éireann

Presentation Overview

CLARITY
clarity-centre.org

- Re-capping how far we've come in information growth, especially with connected devices, called the sensor web;
- Linking information growth, innovation, unpredicted disruption … forms a cycle;
- Present 5x "canned" examples/applications of the sensor web, each local, but causing unpredicted disruption;
- Outline 3 areas of major challenge - discovery, physical-online, real-time web;
- Takehome message … with much more information we can do things we didn't think of.

Two worlds

- Working premise is that we live in two different worlds … physical, and online;
- We digitise the physical into the online but we have an online, born-digital world also;
- Initially work-based, then entertainment, now its social;
- And we've got here very very fast.

How far have we come?

>20 Years Ago

No Cellphones

No Internet

No Digital Cameras

An Offline World

>20 Years Ago

No Cellphones

No Internet

No Digital Cameras

An Offline World

>10 Years Ago

No Mobile Web

No Broadband

No MP3

The Internet Dawn

CLARITY
clarity-centre.org

>20 Years Ago

No Cellphones

No Internet

No Digital Cameras

An Offline World

>10 Years Ago

No Mobile Web

No Broadband

No MP3

The Internet Dawn

5 Years Ago

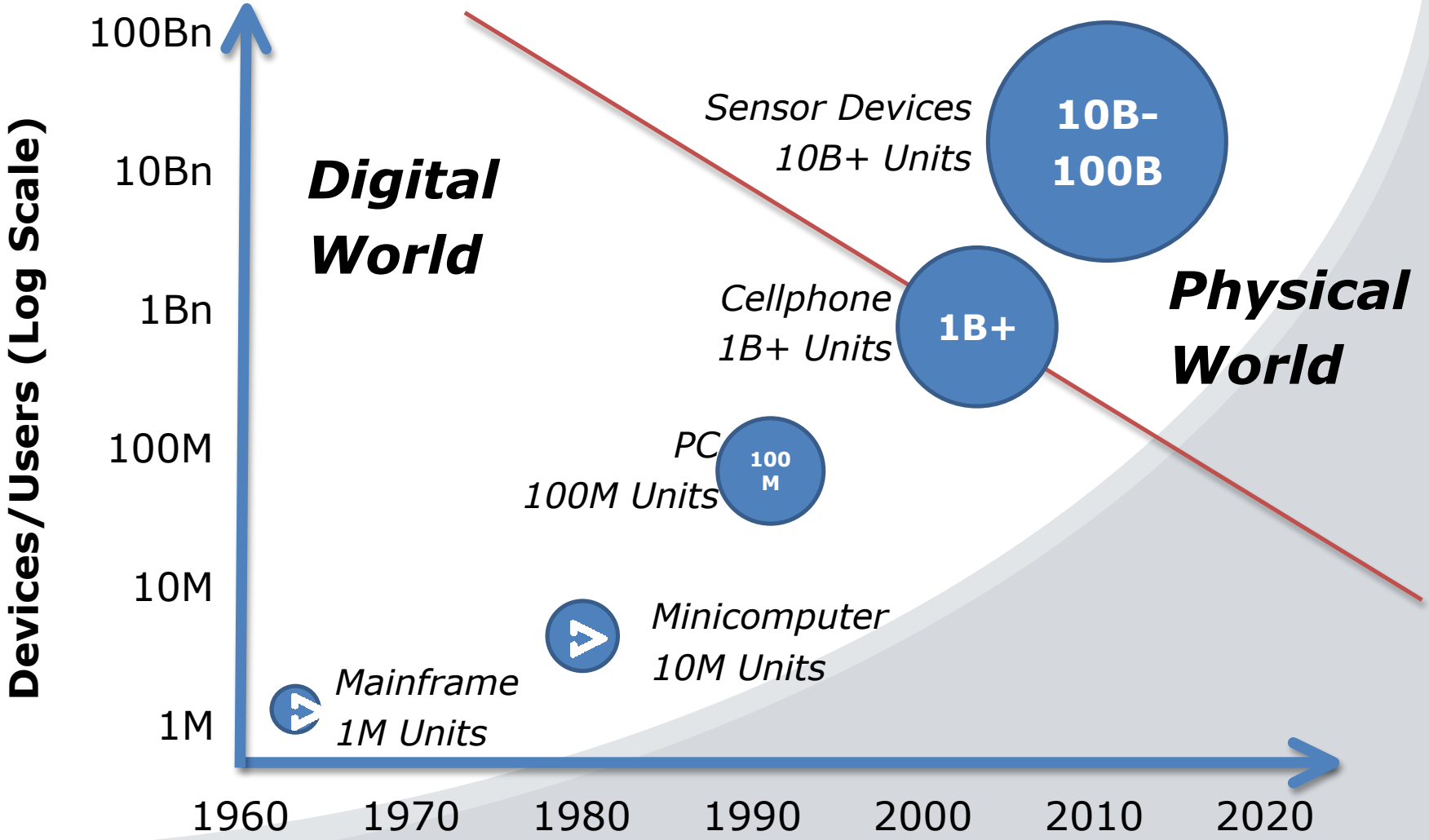
No Facebook

No Twitter

No YouTube

Social Media Age

… and how quickly have we got here !


Devices/Users (Log Scale)
100Bn
10Bn
1Bn
100M
10M
1M
1960
1970
1980
1990
2000
2010
2020
Digital World
Physical World
Sensor Devices
10B+ Units
10B-
100B
Cellphone
1B+ Units
1B+
PC
100M Units
100
M
Minicomputer
10M Units
Mainframe
1M Units

1T Connected Devices ?

- During the last month (Aug 2010), the 5 billionth device was plugged into the Internet (Network World) and by 2020 that will be 22B (IMS);

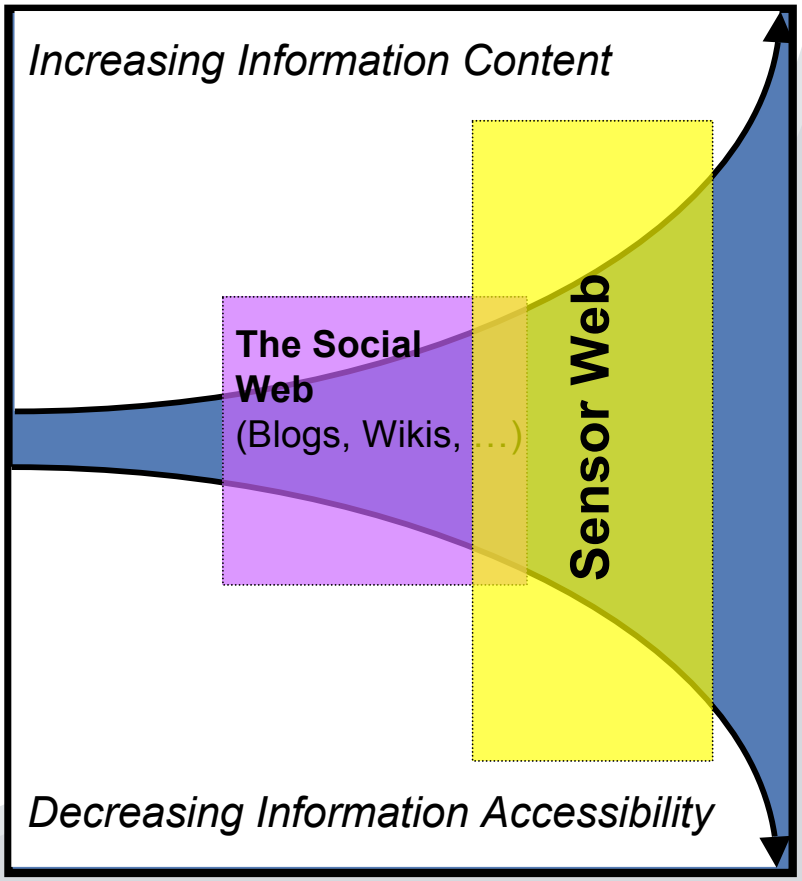
- Cell phones and consumer electronics are the obvious devices but smart grids, surveillance, traffic monitoring … and sensor networks ... the stealth growth in connected devices;
- Emergence of cheap, reliable, flexible sensors - the *Internet of Things* - will help to bridge our physical and online worlds;
- Called the Sensor Web, and it has some real challenges.

The Sensor Web

- With a broad definition of a sensor, there is lots of evidence for emerging sensor web;
- The Economist (Feb 2010) reports
  - Sloan Digital Sky Survey telescope in NM collected more astronomical data in its first few weeks in 2000, than the history of astronomy, now 140 TB but the Large Synoptic Survey Telescope in 2016 will acquire that amount … every 5 days;
  - Wal-Mart handles 1M xactions/hour feeding databases of 2.5 petabytes, x167 Library of Congress books
- We could go on …

The Sensor Web

- The Web
- The Social Web
  - Dynamic user generated content (UGC) - conventional search tools unable to cope
- The Sensor Web
  - Increasing availability of cheap, robust, deployable sensors as information sources
  - Networked, mostly wireless, hence global, and integrated - WSNs
  - Dynamic and reactive but noisy, and unstructured data-streams



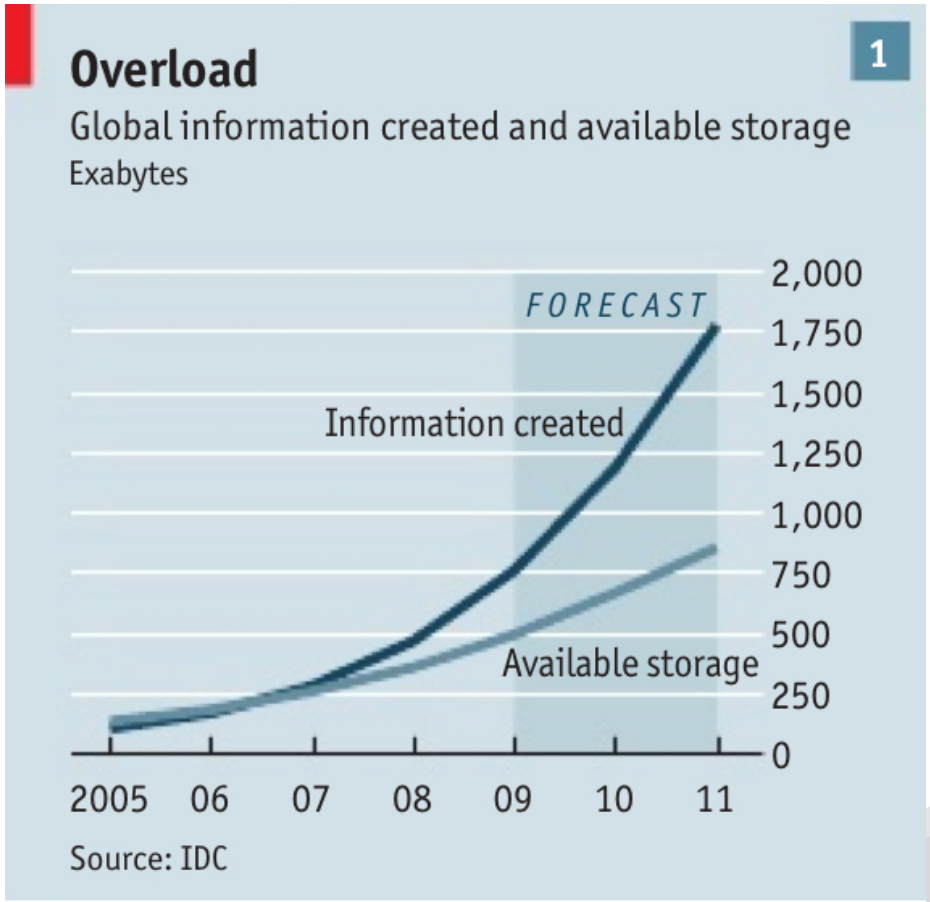
Some technologies prepared

Manufacturing, Networking;

Imagine the ability to stream every movie ever made in less than 4 minutes?

Cisco CRS-3, March 2010

Some unprepared



There is an unimaginably vast amount of data in the world, rapidly getting vaster, and we can't even store it all.

Social Networks

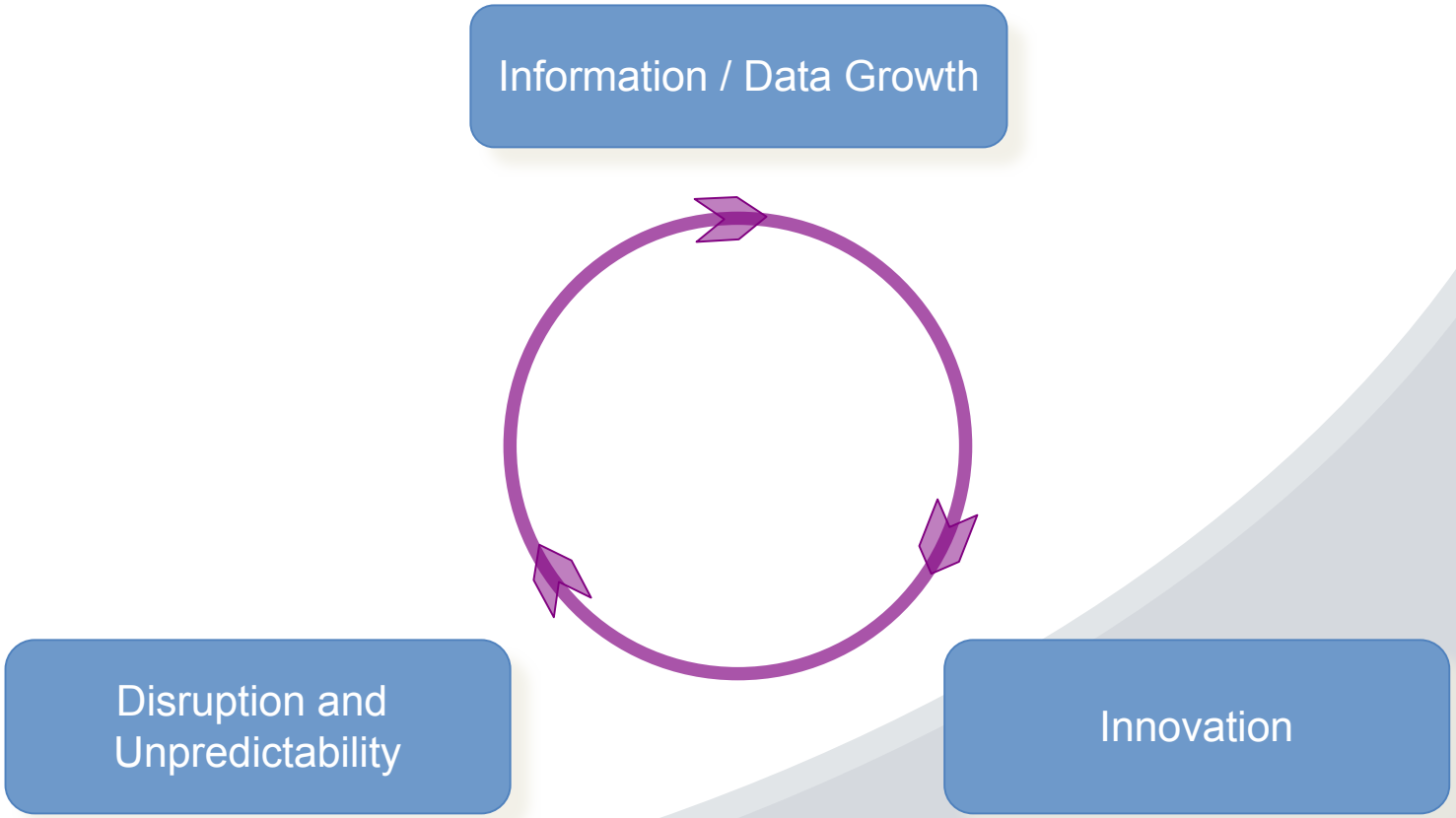
CLARITY
clarity-centre.org

- Much of our online content now comes from social networks - 500M Facebook users, 40B Facebook photos …
- Map of the social network ‘space’ 2010;
- Area of ‘land’ ≈ number of users;
- This is where all the UGC comes from … and this is part of the sensor network, people are sensors.

Info. Growth Consequences

- We observe this growth, but realise that connectivity and information growth brings more then just technical challenges like managing, and searching (see later);
- Information growth is disruptive, unpredictable & dynamic, goes hand-in-hand with innovation;
- Innovation is constant, but its not a constant as the rate of innovation increase is exponential – Kurzweil's Law of Accelerating Returns;
  - Microsoft's Encarta disrupted the encyclopedia market then along comes Wikipedia to disrupt Microsoft;
  - A simple web startup like Craigslist has effectively wiped about $40bn of value from the US newspaper industry;
  - SMS lead to microblogging.

Innovation and Info. Growth



Sensor Web: Applications

- There are lots of statistics, anecdotes, but rather than flood with examples, we illustrate several of our *canned* applications of the sensor web …
  - Visual lifelogging wearable camera (SenseCam)
  - Monitoring domestic energy usage
  - Sensing in sports - cycling
  - Environmental monitoring
  - Online sensing of real world events
- Each is a sensor web *in a can* but even *"in a can",* they are all unpredictable and disruptive.

1. Visual Lifelogging

- Lifelogging is digitally recording, or sensing, daily life
  - Sometimes for a reason
    Work e.g. security personnel, medical staff, etc.
  - Personal e.g. diaries, etc.
- Sometimes for posterity
  - Recording vacations, family gatherings, social occasions
- Sometimes because we can
  - and we're not yet sure what we'll do with it e.g. MyLifeBits

SenseCam Images

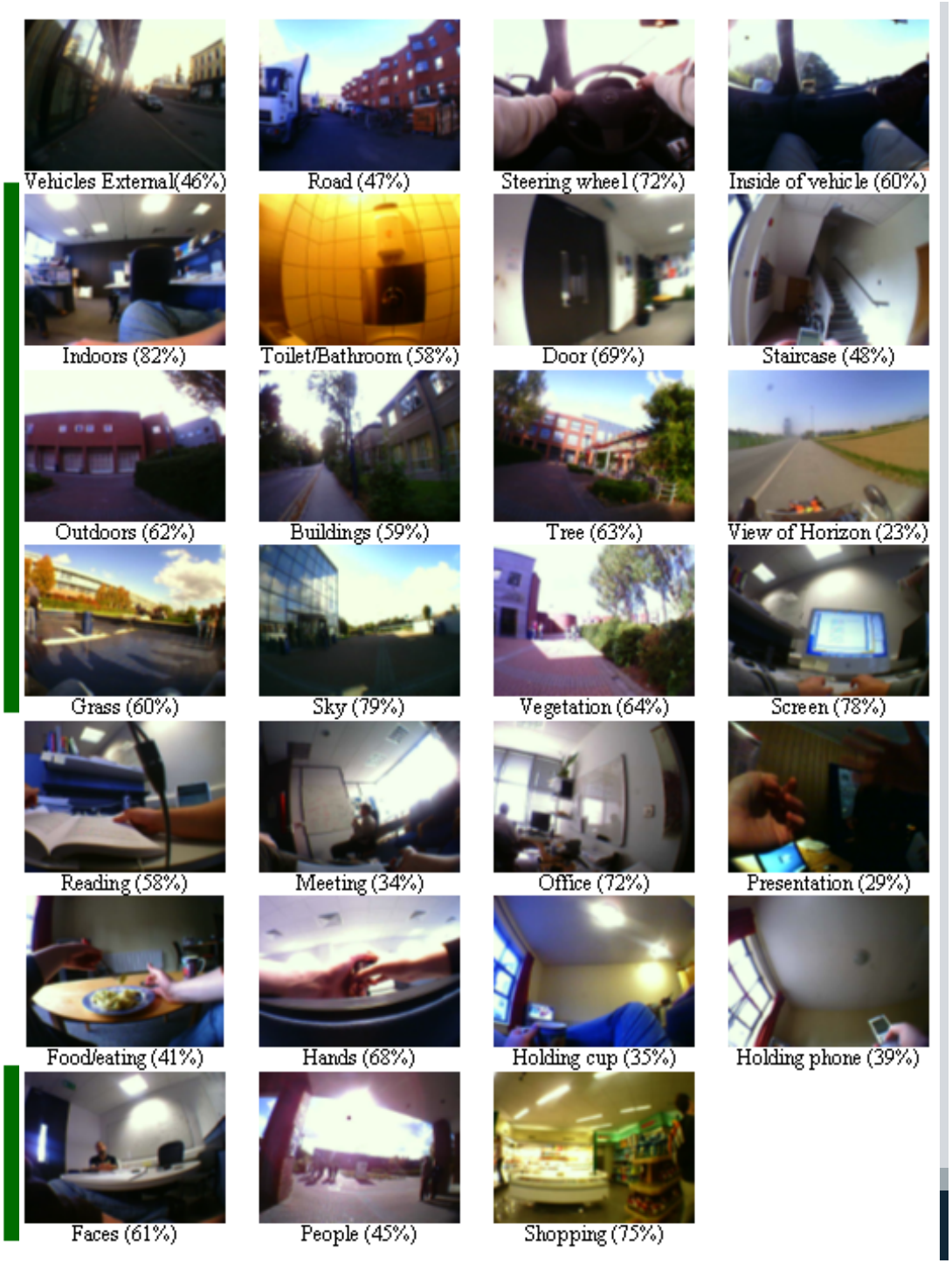
- Wearable camera plus sensors in-built, logging wearer's activities and experiences on-board;
- Images not published so privacy not an issue;
- One user wears SC for over 4 years, all day
  - Each with GPS position
- Experiences:
  - Most people don't notice camera
  - Those that do always remember !
  - Most people don't mind it
  - Have been spotted/greeted by people who have heard about the 'guy with the camera'

Activity Recognition

27 “concepts” defined

Automatic detection

Evaluation
obtained an average
F-score 0.65



The application …

- Memory recall, a memory prosthesis for memory-impaired early-stage Alzheimer's, etc.
- Memory science says … we need to:
  - Group images into distinct semantic "events"
  - Suggest more interesting/distinctive semantic events
  - Associate related semantic events
  - Provide additional retrieval cues from other sources
- Called cued recall, trigger our own memories, not a memory replacement but triggers our own cascade of Proustian moments.

My Visual Diary with SenseCam

hlee: 76,430 photos (25 days)

MY ACCOUNT | SIGN OUT | ABOUT

CALENDAR

◀ MAY ▶ 2006

S	M	T	W	T	F	S
30	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31	1	2	3
4	5	6	7	8	9	10

DURATION ▶

CAPTION SEARCH

WEEKLY SUMMARY

Selected day is shown below in the context of whole week. Move mouse curosr over to see other similar Events in the week

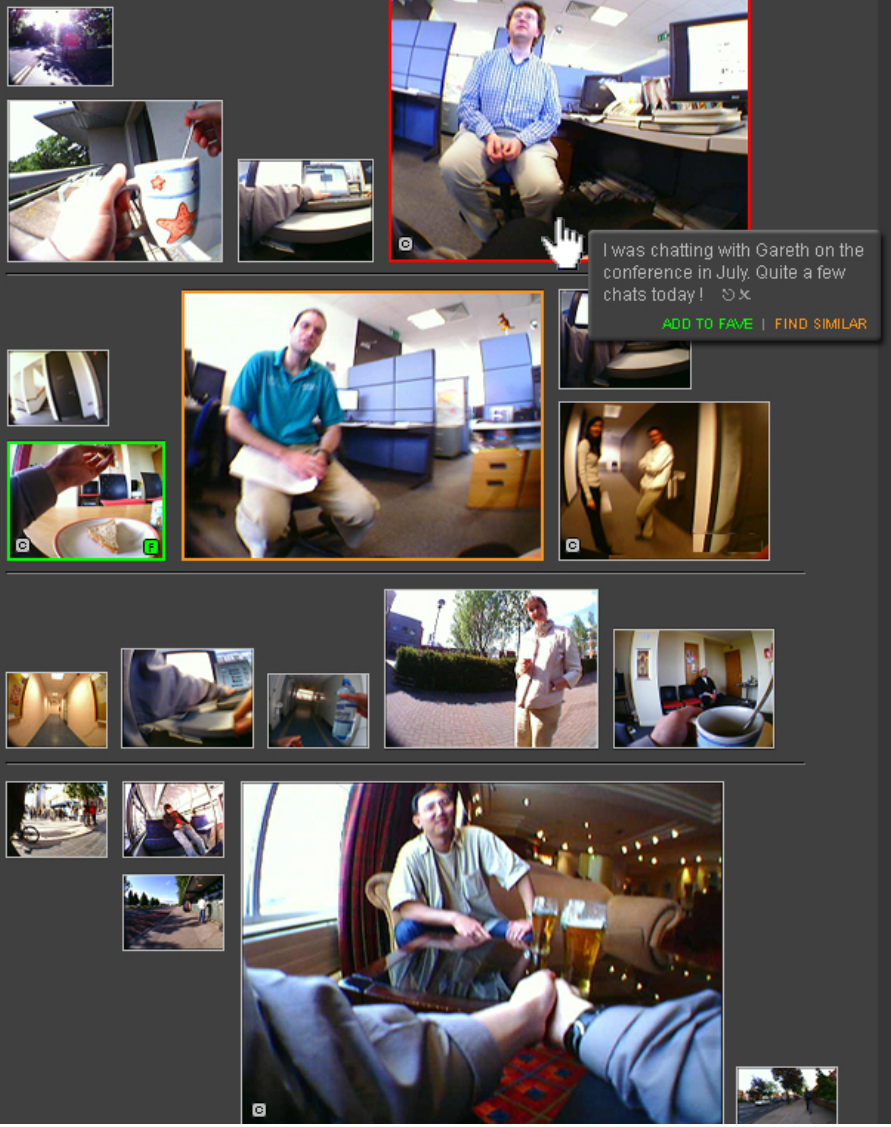
S
M
T
W
T
F
S

29 May 2006

19 EVENTS

Drag the slider bar to adjust the number of Important Events

0 am | 6 am | 12pm | 6pm | 0am



I was chatting with Gareth on the conference in July. Quite a few chats today!

ADD TO FAVE | FIND SIMILAR

MY FAVOURITE EVENTS

25 Favourite Events are shown below. Click on the photo to replay all photos within the Event.

| 1 | 2 | 3 |

Sort by: TIME | SIMILARITY | #PEOPLE

16:20 (Duration: 08m 43s)
14 APR 2006 ▶

13:45 (Duration: 14m 05s)
14 APR 2006 ▶

10:02 (Duration: 23m 56s)
13 APR 2006 ▶

14:39 (Duration: 15m 30s)
12 APR 2006 ▶

11:25 (Duration: 08m 21s)
12 APR 2006 ▶

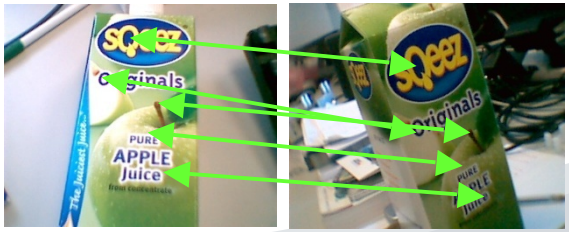
09:52 (Duration: 01m 03s)
12 APR 2006 ▶

Unpredicted: market analysis

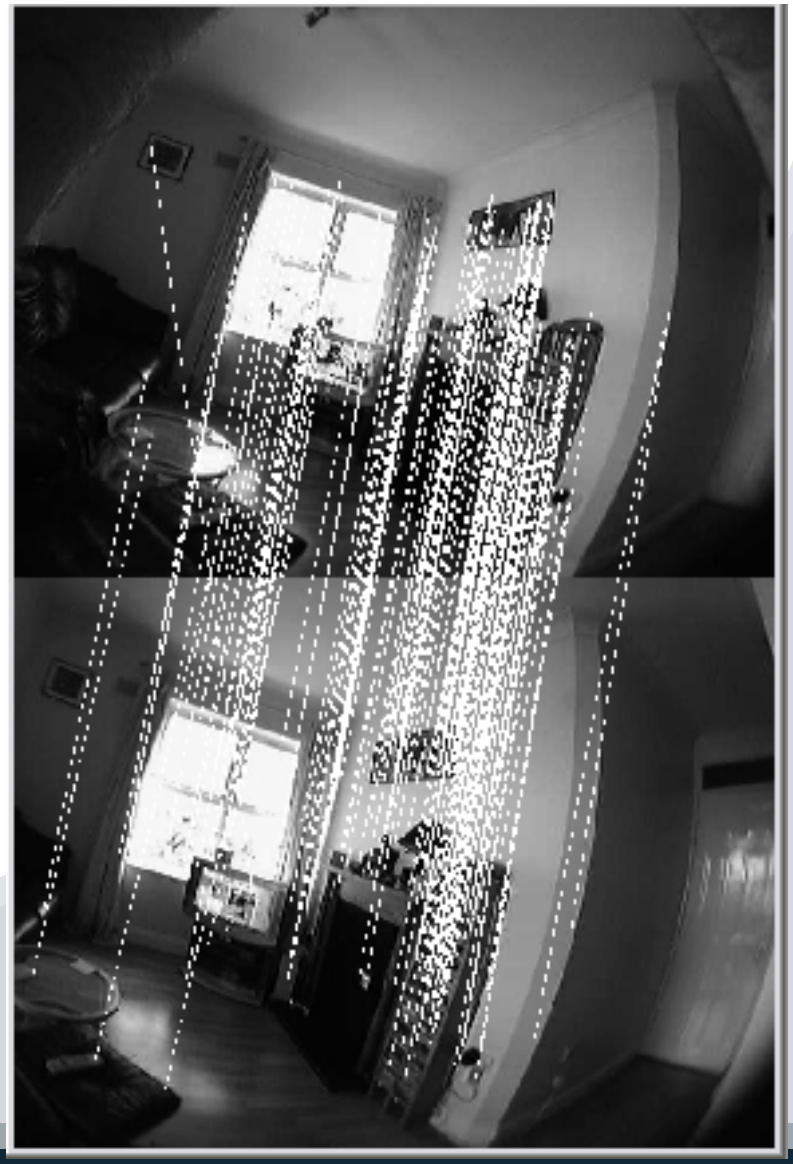
SURF features are extracted

Each feature point casts a weighted vote against multiple database images

Votes are accumulated & the best match is found

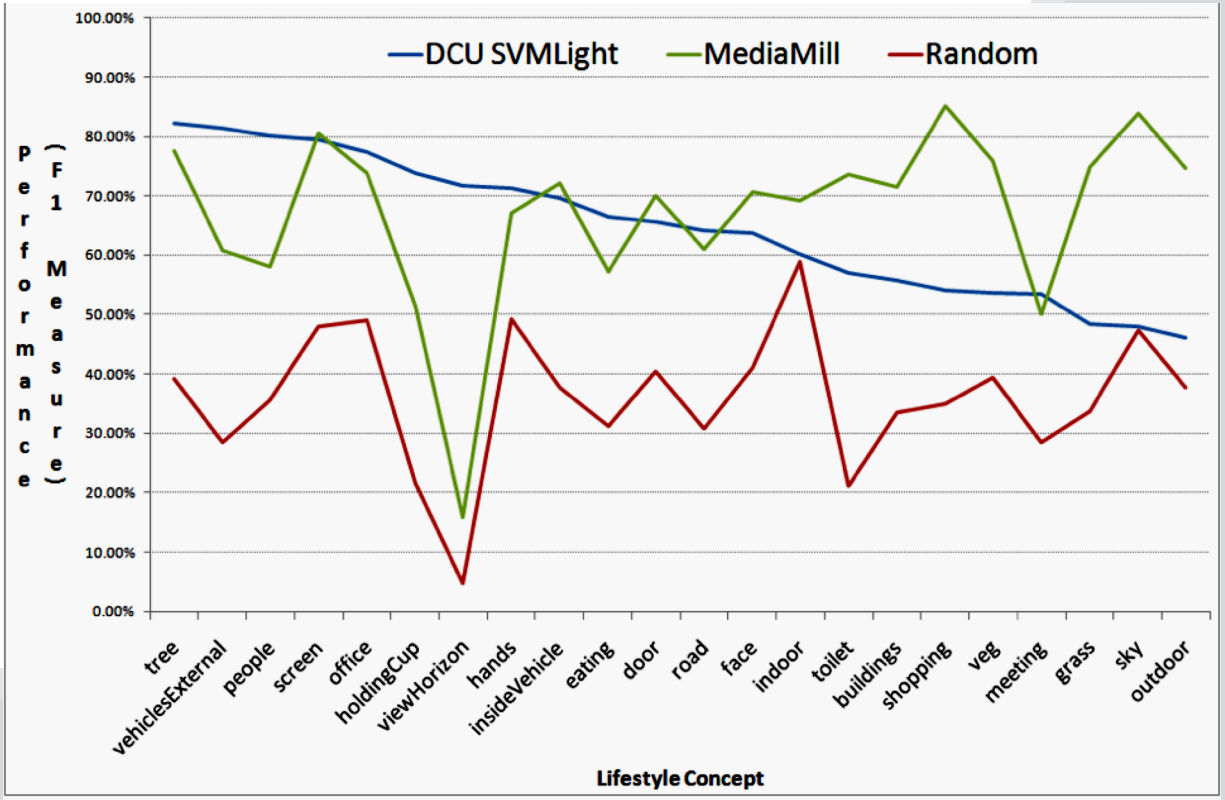


Unpredicted: setting detection



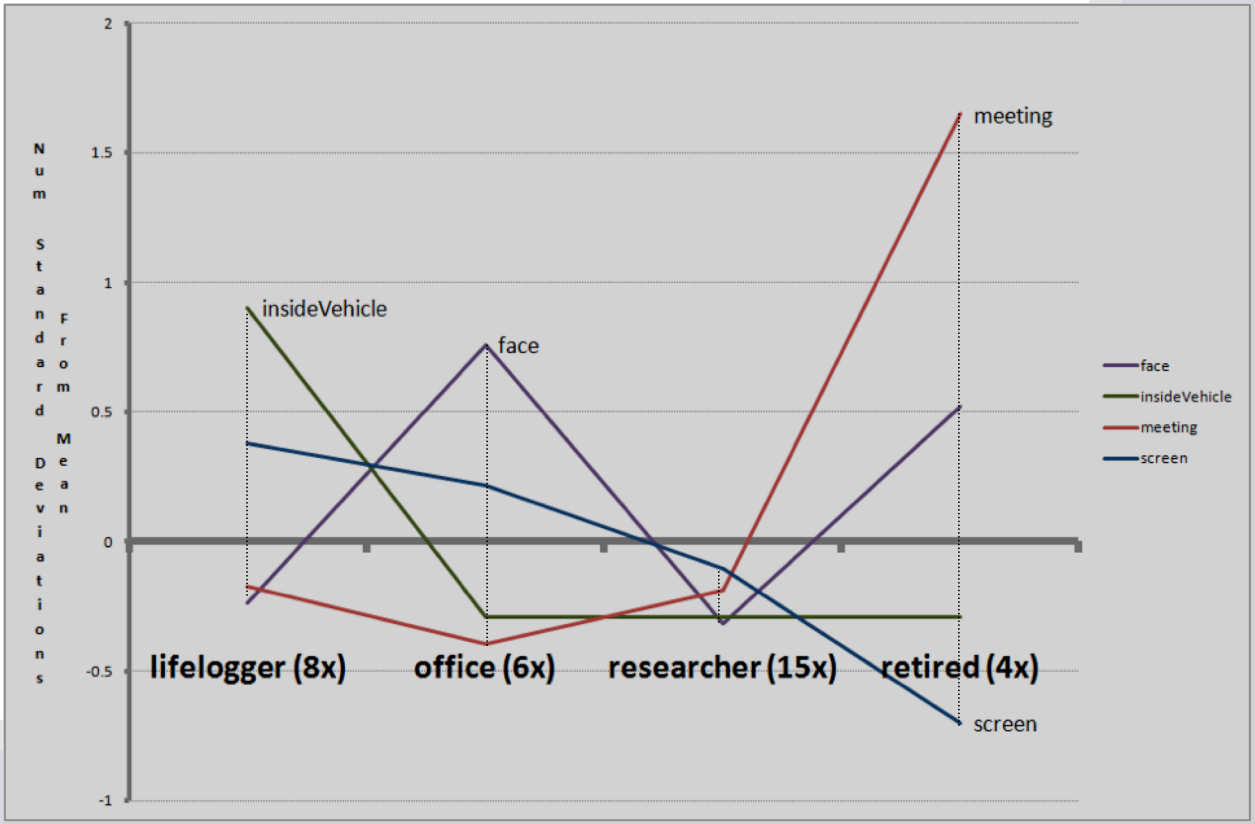
Unpredicted: lifestyle analysis

- Tested and evaluated 22 of 27 concept detectors on 3.5 M SenseCam images, representing 43,072 events or +1 M minutes of 33 participants drawn from 4 social groups



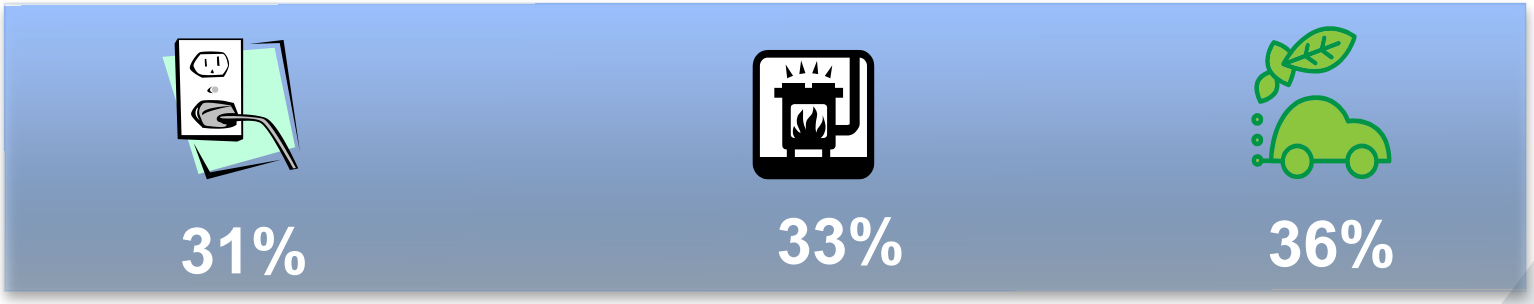
Unpredicted: lifestyle analysis

- Can compare across groups, individual vs. peers, individual vs. groups, etc.
- Illustrative example, take …
  - *meeting*
  - *face*
  - *inside-vehicle*
  - *screen*
- Now working with public health and ethnographers on lifestyle (larger samples)



2. Domestic Energy

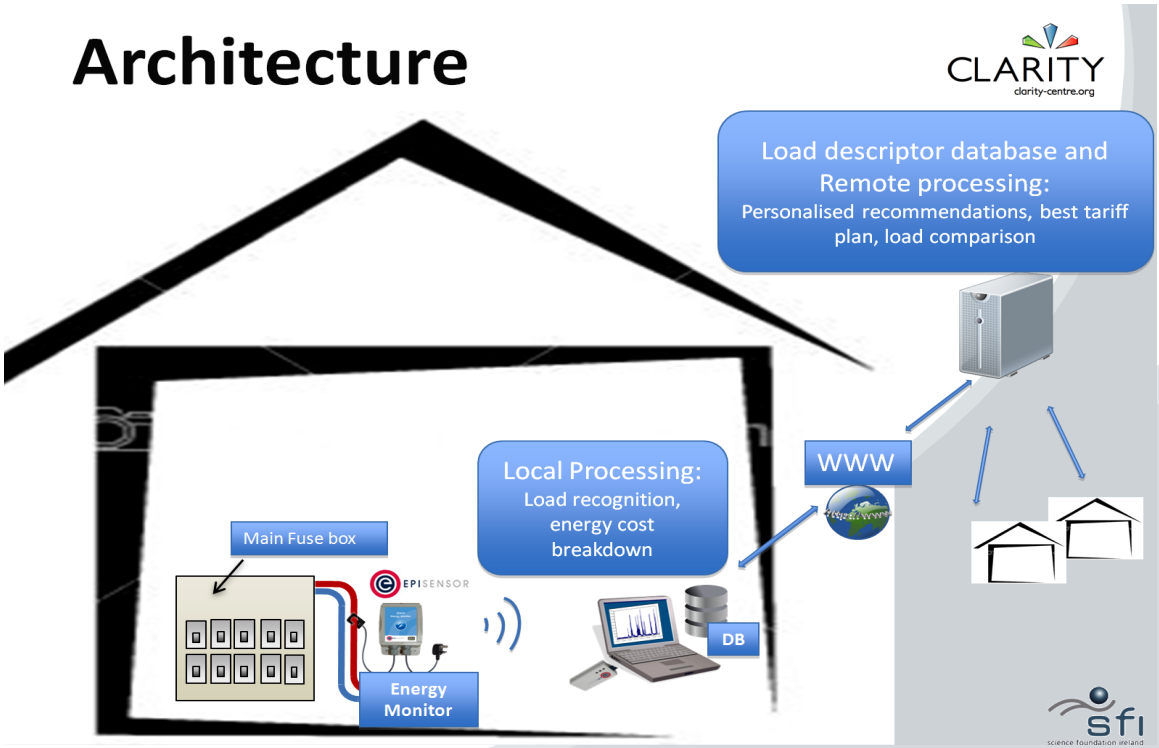
IRELAND's ENERGY-RELATED CO_2 EMISSIONS (SOURCE: SEAI)



- Smart energy metering projects springing up;
- Attractions of data collection, load-based tariffs, real time user feedback and management of peak demands;
- Also supports micro-generation, and energy trading using storage from EVs;
- Smart meter trials generate sensor readings.

2. Domestic Energy

- So we built the following



2. Domestic Energy

CLARITY
clarity-centre.org

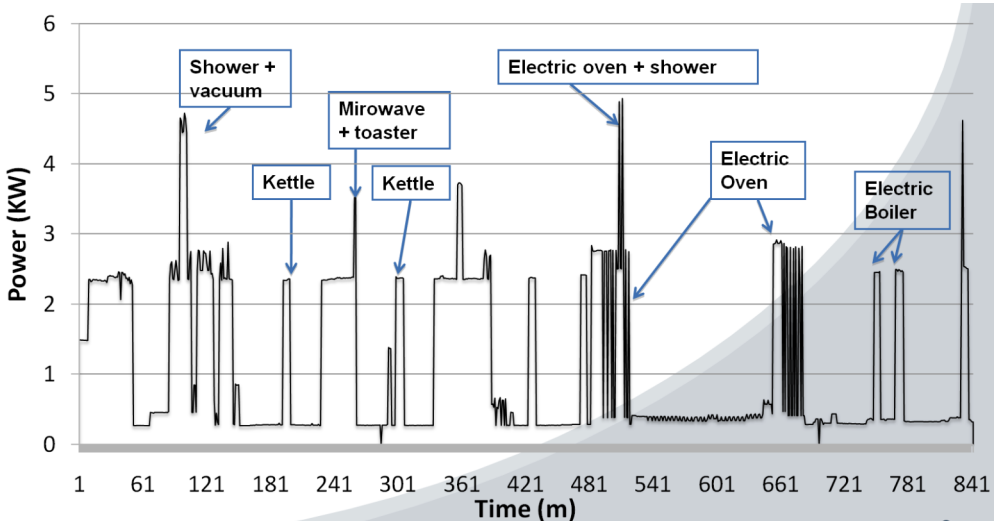
ZEM-30 Energy Monitor

- Monitoring cuff placed on/around live line of mains supply
- Reports multiple power/current/voltage values every minute
- 15,840 sensor readings per house per day
- Normal 5-7pm weekday peak in electricity consumption

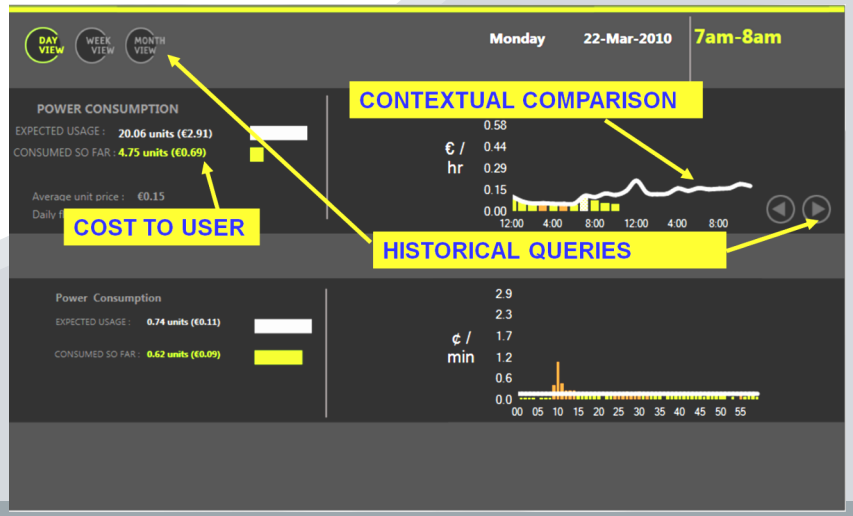
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Mon	23.30%	10.27%	2.29%	0.41%	1.70%	9.66%	17.55%	29.96%	38.69%	38.50%	39.92%	34.18%	36.60%	29.62%	32.49%	43.57%	51.69%	71.31%	91.22%	83.01%	71.93%	61.94%	54.76%	44.15%
Tue	29.29%	6.49%	0.50%	0.30%	4.10%	6.27%	10.33%	32.53%	36.70%	45.49%	42.51%	36.06%	33.77%	35.37%	41.86%	42.16%	52.22%	75.31%	100.0...	77.10%	71.93%	73.14%	60.81%	44.28%
Wed	20.00%	7.48%	0.01%	0.00%	3.66%	8.78%	15.70%	29.18%	43.00%	39.66%	37.49%	34.90%	30.08%	27.72%	34.92%	34.47%	50.77%	68.57%	99.50%	91.12%	76.78%	60.16%	53.13%	40.88%
Thu	21.24%	5.30%	1.93%	1.41%	3.92%	7.47%	16.15%	43.85%	44.76%	45.73%	43.50%	41.94%	47.68%	35.01%	50.02%	53.48%	69.46%	86.10%	98.74%	95.09%	70.76%	55.37%	49.95%	39.99%
Fri	21.27%	9.18%	4.50%	1.90%	2.00%	7.33%	14.02%	29.99%	45.46%	40.45%	35.79%	28.99%	27.58%	37.63%	43.37%	38.67%	47.61%	58.54%	76.15%	75.50%	74.42%	65.59%	52.69%	41.93%
Sat	28.16%	18.36%	6.12%	3.75%	3.07%	9.19%	6.31%	7.59%	22.78%	41.62%	48.43%	45.68%	49.70%	53.88%	60.18%	47.98%	55.20%	75.86%	84.03%	70.21%	61.38%	51.15%	45.15%	42.90%
Sun	27.93%	18.69%	9.97%	10.04%	4.21%	12.08%	7.45%	7.73%	20.35%	32.55%	53.73%	63.35%	57.50%	49.85%	49.29%	59.99%	68.39%	76.22%	93.03%	82.27%	80.68%	67.26%	61.48%	37.75%

2. Domestic Energy

- Typical usage profile



- Touchscreen for in-home display;
- Passive visualisation thru smart textile furnishings;
- Can recognise major appliances.



2. Domestic Energy

USE CASE: SHOWER USAGE

- 1 home, Sep-09 to Jul-10
- 3,391,443 sensor readings ...520 showers
- Average duration = 8 minutes (min = 3 min, max = 27 min)
- 8am Tuesday morning is peak time

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Mon	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	9.15%	26.14%	53.59%	49.02%	75.16%	66.67%	54.90%	33.99%	25.49%	0.00%	10.46%	1.96%	4.58%	7.19%	11.76%	0.00%	0.00%	2.61%
Tue	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	7.19%	50.33%	100.0...	64.71%	37.91%	57.52%	0.00%	37.25%	5.88%	6.54%	6.54%	1.96%	5.88%	0.00%	0.00%	3.27%	11.11%	0.00%
Wed	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	9.15%	61.44%	79.08%	52.29%	49.67%	38.56%	31.37%	18.30%	6.54%	11.76%	2.61%	9.15%	11.11%	7.19%	0.00%	0.00%	3.92%	0.00%
Thu	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.2...	77.12%	60.78%	96.73%	56.21%	58.17%	19.61%	33.33%	22.22%	9.15%	14.38%	0.00%	0.00%	13.07%	7.19%	8.50%	3.27%	0.00%
Fri	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	...76%	46.41%	64.05%	64.05%	13.07%	30.72%	16.34%	31.37%	19.61%	3.27%	33.99%	25.49%	0.00%	0.00%	8.50%	1.96%	0.00%	0.00%
Sat	0.00%	0.00%	0.00%	0.00%	0.00%	0.00...	0.00%	13.73%	29.41%	33.33%	13.07%	33.99%	85.62%	79.08%	50.33%	29.41%	33.33%	14.38%	5.88%	5.88%	4.58%	0.00%	7.19%	3.27%
Sun	0.00%	0.00%	0.00%	0.00%	0.00%	0...0%	0.00%	18.30%	18.95%	13.73%	43.14%	32.03%	50.33%	25.49%	30.07%	18.30%	28.76%	24.18%	4.58%	0.00%	7.84%	7.84%	4.58%	0.00%

Morning Showers

2. Domestic Energy

- What people wanted, based on trials in homes ?
  - (near) realtime feedback and warnings;
- Unpredictable disruption comes in the lifestyle patterns and analysis;
- Unpredictably, now working with health monitoring agencies, monitoring older people life patterns in their homes.

3. Sports - Cycling

- Race Around Ireland cycle race
  - Non-stop, 1,350 miles, solo or team of 4, some hills, mostly flat;
- Strategy and pacing are key to completion
  - When to rest, how to pace, who to race, recovery;
- Human performance scientists interests
  - How the human body recovers, how to maximise performance over 4/5-day endurance, what recovery techniques work best.

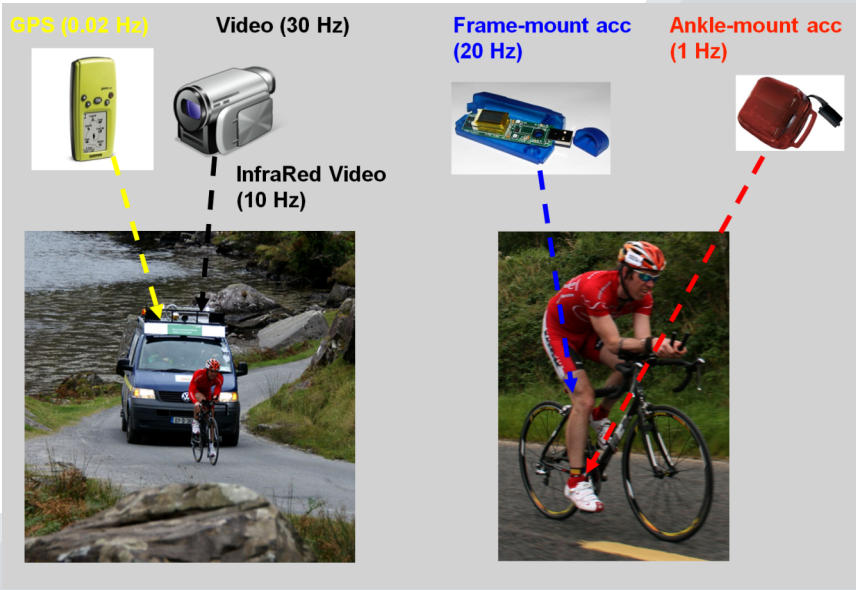
3. Sports - Cycling

So we instrumented a team with sensors

- On Body …
  - Heart rate, GSR, respiration, fluid intake, food intake, sweat analysis, blood analysis, massage,
  - Ankle-mounted WIMU gives cadence, cycle/rest times

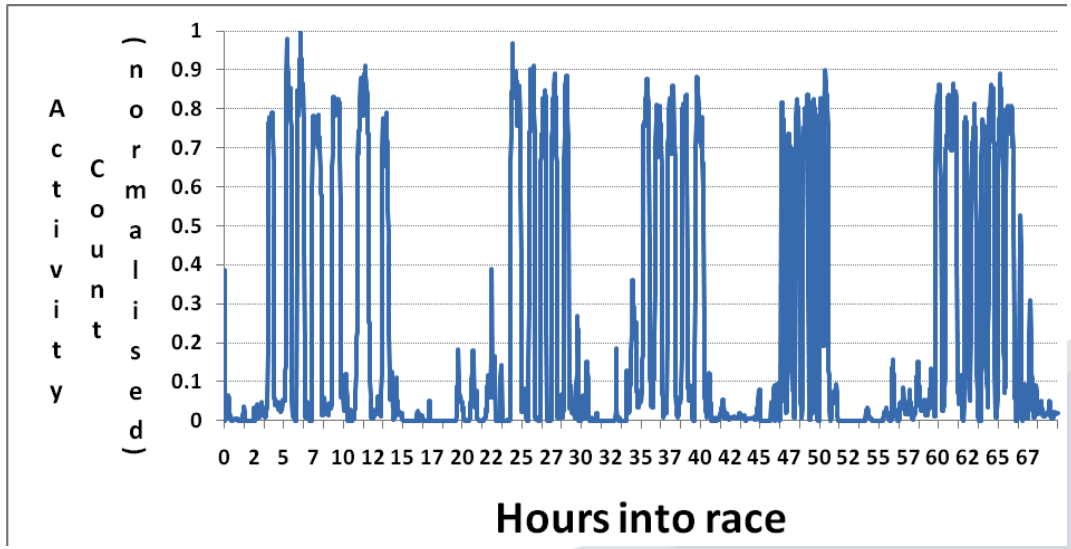
- On Bike …
  - WIMU gives yaw and roll, in/out saddle
  - Power gauge gives power output

- Environment …
  - Weather, wind, temperature, road surface, wet/dry, day/night video tracking using geospatial measurement jeep



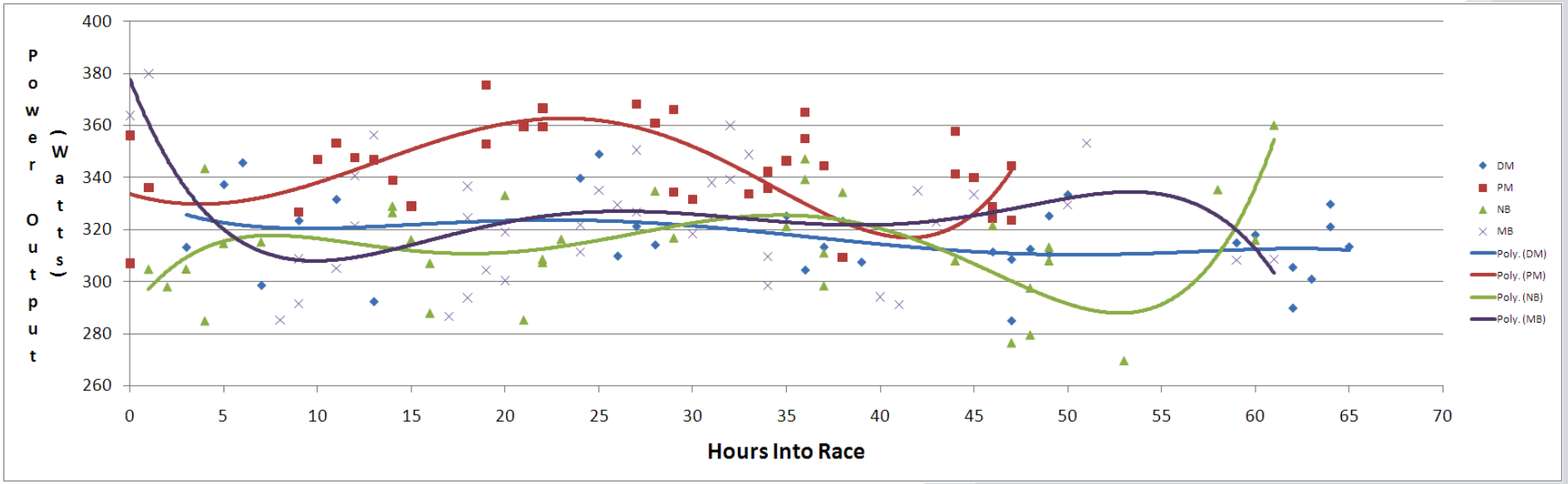
3. Sports - Cycling

- What do people want ?
  - Study of performance (degradation) over time, impact of different recovery techniques
  - Insights into endurance performance



3. Sports - Cycling

- For sports scientists, the following is gold … aggregated from multiple sensors;

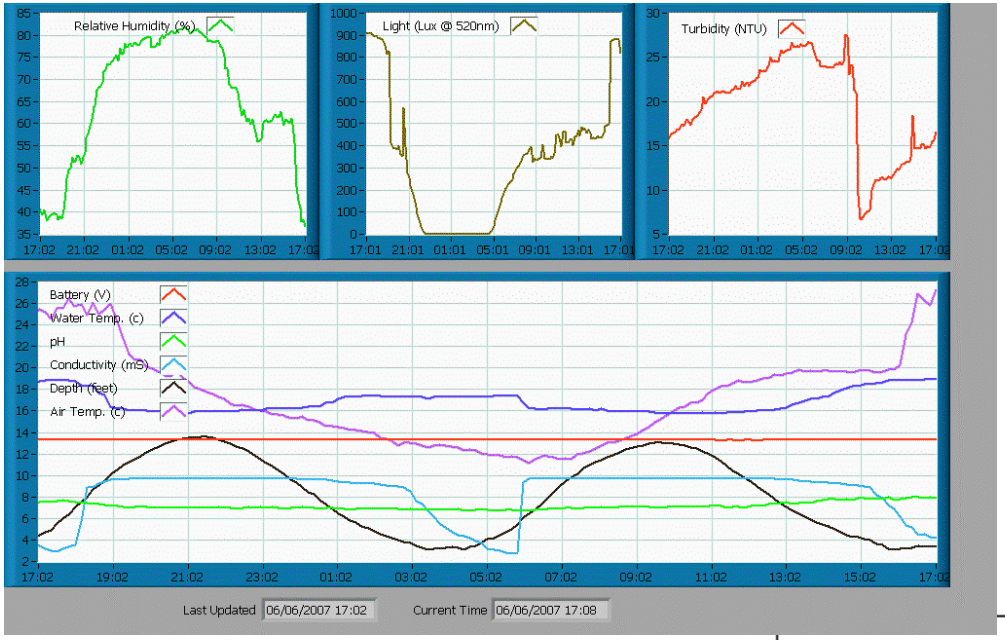


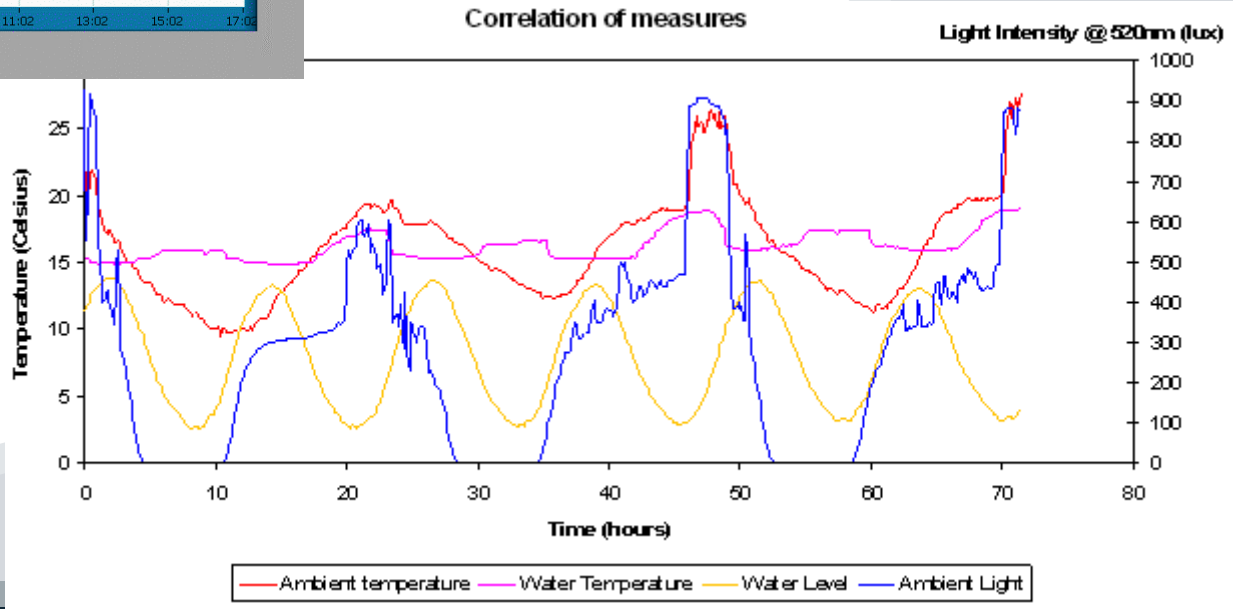
- Disruption comes from doing this in real time, team management decision-making.

4. Environmental Monitoring

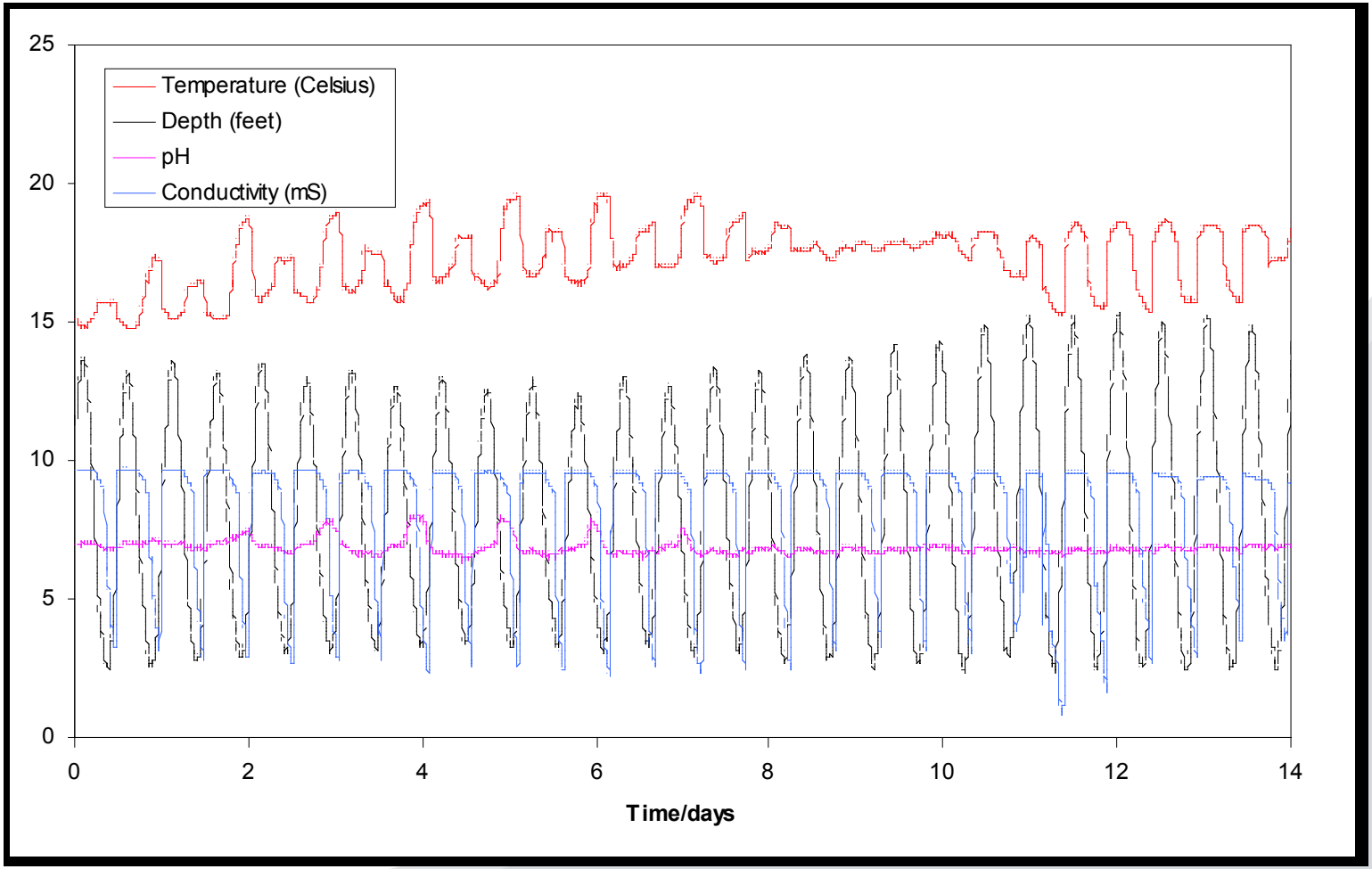
- River Lee, Cork, Ireland, a city centre tidal river;
- We can sense …
  - Water turbidity (NTU), water temperature (degrees Celsius), pH, water conductivity (milli-Siemens), water depth (feet), dissolved oxygen, nitrates and sulphates, air temperature, relative humidity and light (lux @ 520nm)
- Phosphates are nutrients and are bad because encourage plant growth, consume O_2;
- Sensor platform is microfluidic-based, 1-year field lifetime, shoebox size, solar powered.

5. Environmental Monitoring





5. Environmental Monitoring

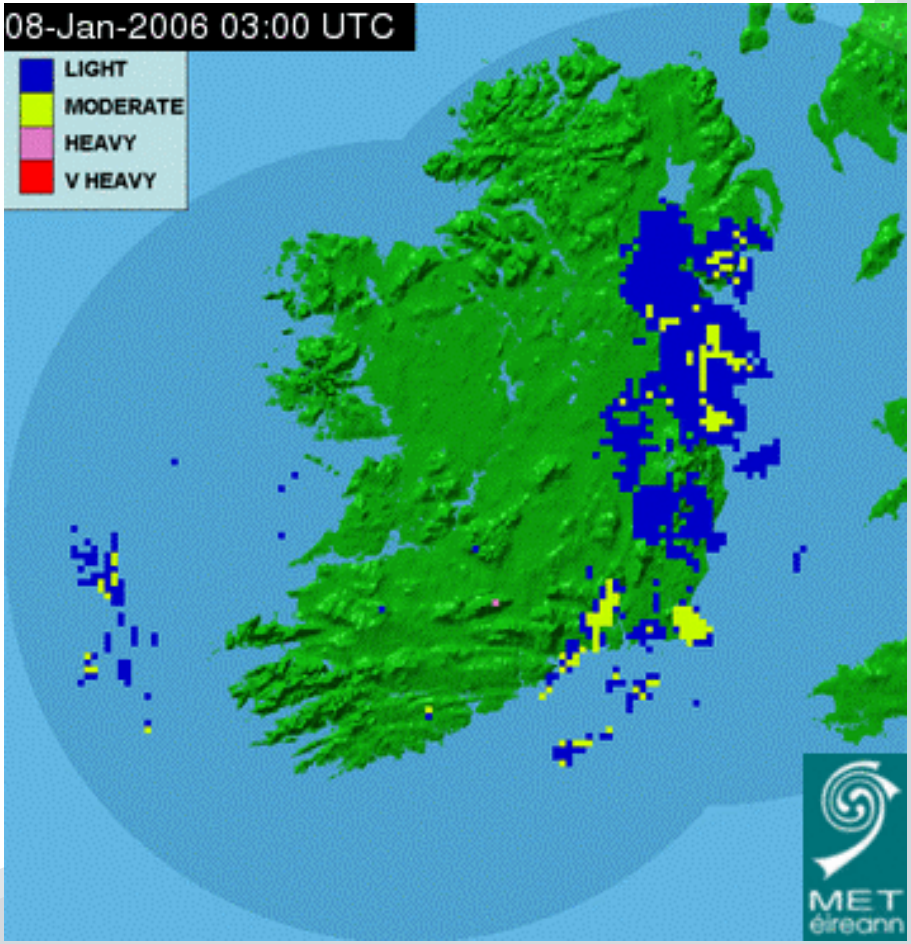


5. Environmental Monitoring

- Several deployments throughout the country, all measuring phosphates, nitrates, DO, as well as “off-the-shelf” sensors, WSN-based;
- Problem with these sensors is longevity … reagents need to be replaced;
- Environmental scientists want readings at times of change, not times of stability … changing times are predictable based on local rainfall and within catchment;
- So we broadened the context and used other sensors.

Met Éireann Met.ie

- We harvest and process to track rainfall in near-term for, the catchment areas for our river sensors;



4. Environmental Monitoring

- What do people want … environmental sensing
  - Long lived sensors;
  - Real time readings;
  - Integration across sensors and sensor sources;
  - Detection and alerting of events, based on combination of factors derived from combinations of sensor readings;
- Unpredictable was broadening sensors … historical rainfall radar used to influence sampling rate and prolong longevity of sensors.

5. World Cup 2010

- June 2010, South Africa;
- Ireland not present because of Thierry Henry's handball;
- 700M (or 1B, whichever) watched the final (Spain won);
- 64 games;
- From the video, we automatically detect goals, penalties, cards, other highlights;
- We have time-stamped in-game betting odds from a major online betting agency.

5. World Cup 2010

- Official FIFA match reports (the groundtruth);

- Newspaper reports, blogs, linked images;
- 48M Twitter posts, time-stamped;
- Twitter-based event detection, a kind of socially-backed highlight detection;
- Twitter-based sentiment analysis, a kind of socially-based opinions on teams, players.

5. World Cup 2010

- Applications …
  - Multimedia/multimodal, search and summarisation;
- The unpredictable disruption comes with
  - “Second screen” viewing, like MetaMirror from Notion;
  - Realtime highlighting and summarisation;

Information Discovery

- We've illustrated several applications of the sensor web that we're involved in …
- Common characteristic is real users, with requirements and needs
  - Visual lifelogging, SenseCam wearable camera
  - Monitoring domestic energy usage
  - Sensing in sports - cycling
  - Environmental monitoring
  - Online sensing of a real world event

- A second characteristic is that each one has ended up with unpredicted disruption in some way;

Challenge Areas

- Sensor Web challenge / opportunity areas:
  - There are very many associated with the sensors (new sensor types, calibration drift, power management, long life), networking (bandwidth, power drain, coverage), WSN architecture (ad hoc networks, data aggregation points, trust and provenance), etc.

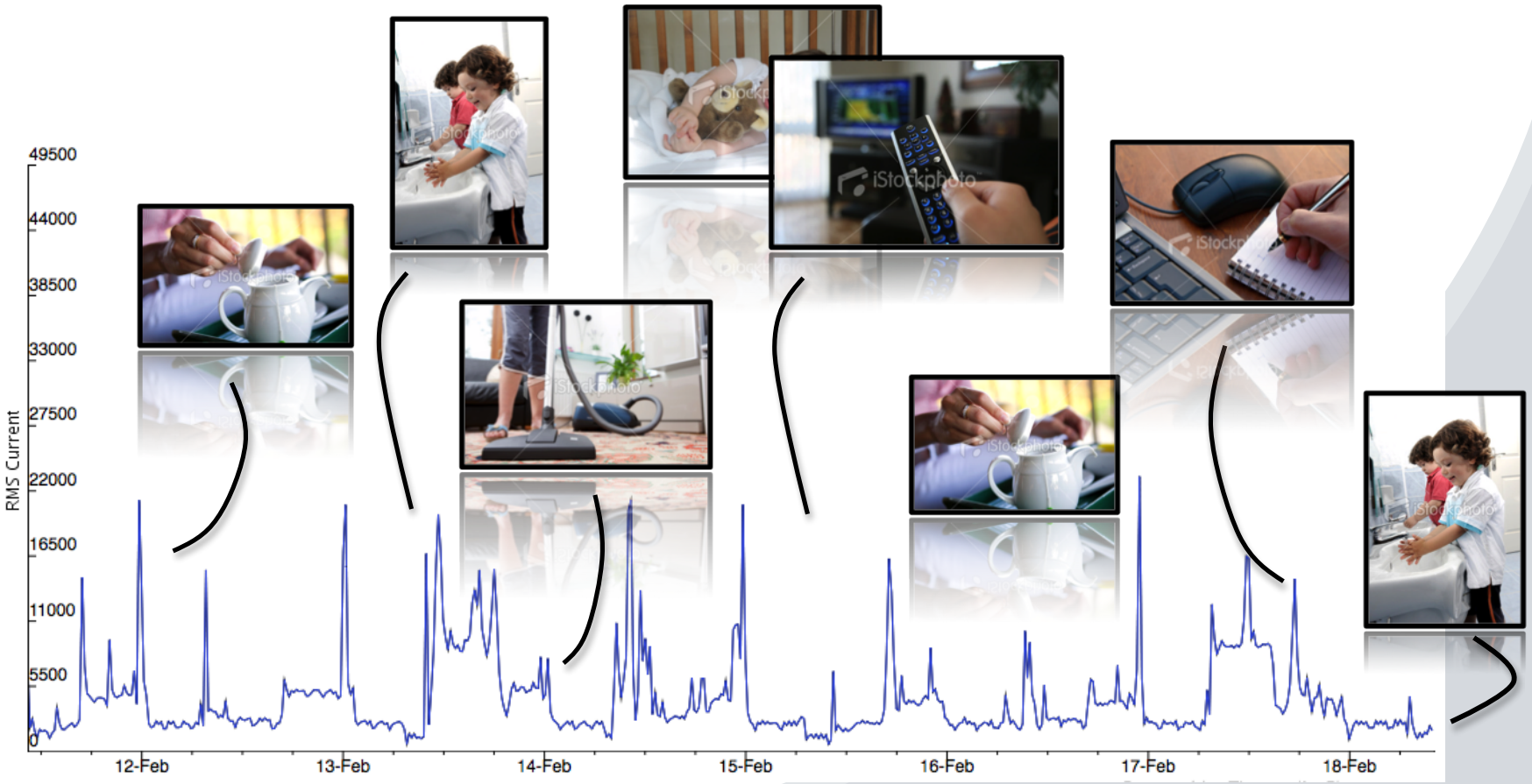
- For us, we see three areas, layered,

A. Information Discovery
B. Bridging the Physical - Digital divide
C. Real time web

A. Information Discovery

- Making Sense of Sensor Data
  - Raw sensor data ⇒ recognising/representing/classifying meaningful events/activities;
- Mining ‘Interesting’ Events
  - Identifying novel and/or significant events ⇒ prioritising relevant events ⇒ event aggregation & information diffusion;
- Supporting Information Discovery in the Sensor Web
  - Also includes Authority and Reputation
- Our examples all involve activity recognition.

Making Sense of Sensor Data



Making Sense of Sensor Data

- Because we have users, we know what events to look for;
- Requires machine learning but the ‘ontology’ so far is flat;
- For each of these we move:
  - Sensor data
    ⇒ Segmented events
    ⇒ Interesting events
    ⇒ Event augmentation/enrichment
    ⇒ Event notification
- That’s where the challenge is, limited domain semantics.

B. Bridging Physical-Online

- Broad definition of sensors includes UGC - not not digitiation of the physical, but born digital;
- Currently microblogging but also SMS, tags;
- Data is availabile on web scale;
- Source of sensor data on physical events
    - dynamic, noisy, unstructured
- We can extract sentiment from blogs and UGC, do polarity classification (+ - /), which is a sensor data value.

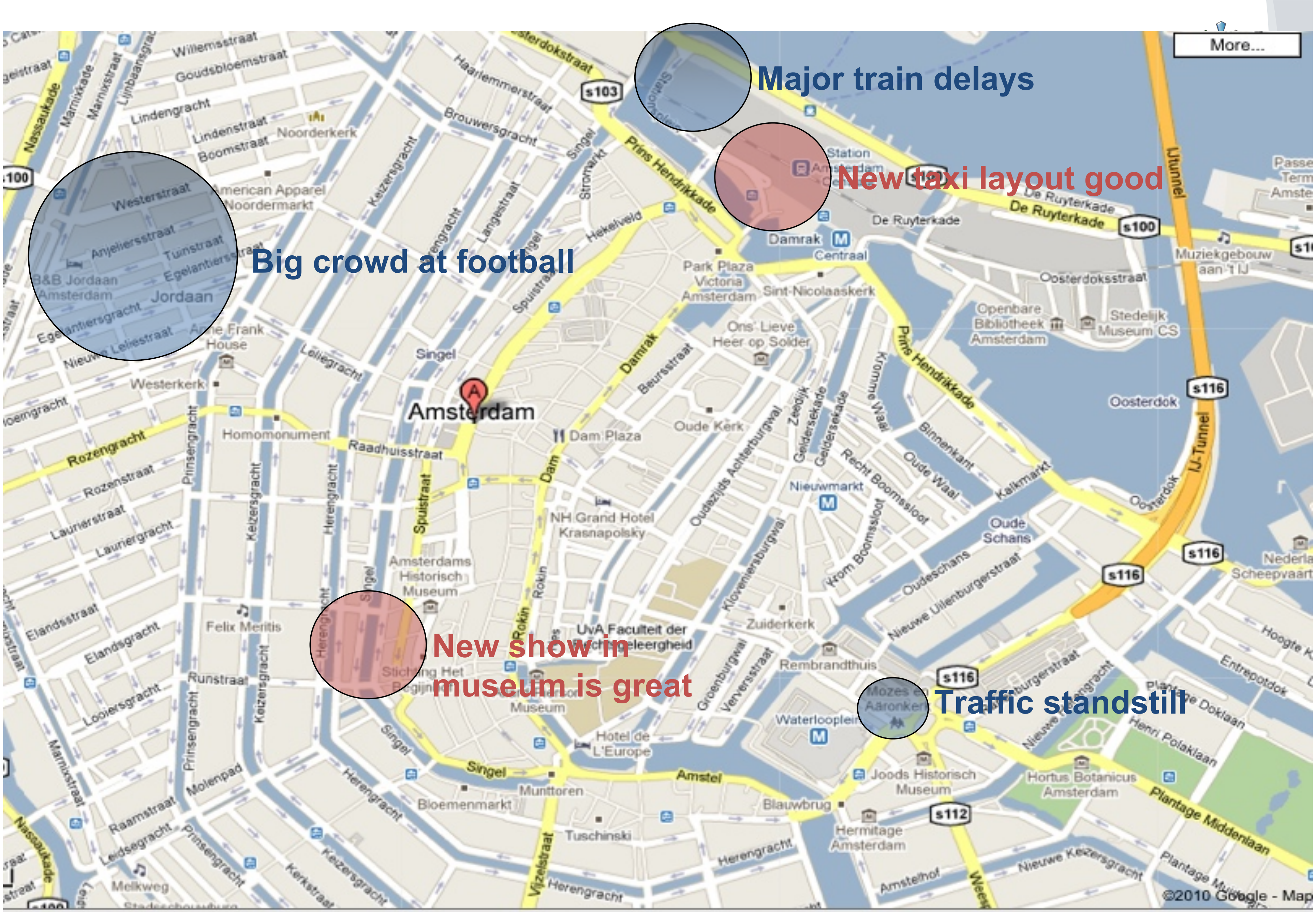
B. Bridging Physical-Online

- Allows us to sense real-time events - scheduled events like concerts, shows, sports, cultural;
- Real-time feedback, alternative views, alternative replays;
- Automatic summarisation of events based on sentiment highs, and lows … mining interesting moments;
- Physical world
  ⇒ born digital
  ⇒ physical.

C. Real Time Web

- Real time online services like Twitter, FB, Foursquare, etc. are simple, and popular;

- Real time physical world of sensor data combined with RT online services makes possible sensing of **unscheduled** real-time events;
- Not just combine but **integrate** physical and online, and in real time;
- RTW is changing how we use information - we're mobile and we're social;
- So many examples of how RTW helps in situations:
  - Mumbai attacks 2008, grad student taking photos in Egypt, various earthquakes, floods, other disasters, neighbour's cat;
- Requires *active* information discovery and open-domain semantics … here's one visualisation:


Major train delays
New taxi layout good
Big crowd at football
New show in museum is great
Traffic standstill
Amsterdam

C. Real time Web

- A fusion of networked physical sensors (people movement from CCTV, Foursquare, traffic flow, environment, weather, noise, bike usage, etc.) and born-digital online (geo-tagged tweets, blogs, event guides, online news, reviews), in real time;

- Colours to indicate sentiment (+/-),
- Transparency to indicate certainty/trust,
- Size to indicate geographical range,
- Playback persistence to indicate temporal span,
- … and then it can be personalised
- … and on mobile
- … and …

Searching Real Time Web

- So there's the picture.

- Where's the multimedia ?
  - Its everywhere, MM has a broad definition
- Where's the semantics ?
  - Its starting to creep in, everywhere, but lightly
- Where's the search ?
  - What's the Google of the sensor web ?

Why is it interesting ?

- Because we are sensing so much of the (physical and online) worlds we now know so much more about our world, and it follows naturally that we can put that new knowledge to good use.
- That's what makes the sensor web compelling.

Re-cap …

- Overview of digital/online and physical worlds … introduced sensor web … discussed information growth … described 5x of our canned/local sensor web applications, emphasising user importance but also unpredicted disruption and outcomes … outlined 3 challenges (information discovery, physical-digital and real time web);

- Thanks to Science Foundation Ireland (funding) and the many people in Dublin who do all this work