Introduction to “Big Data” (and Friends)

Stephen F. Harper
Global Director, Environment and Energy Policy
Intel Corporation

Environmental Law Institute
October 2013

First, a Few Definitions…

- **"Big Data"** = "…is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database transfer, analysis, and visualization." (Wikipedia)
  - Data are often "unstructured"

- **"Cloud Computing"** = "…describes a variety of different types of computing concepts that involve a large number of computers connected through a real-time communications network such as the Internet." (Wikipedia)

- **The "Internet of Things"** = "…refers to uniquely identifiable objects and their virtual representations in an Internet-like structure…Equipping all objects in the world with miniscule identifying devices or machine-readable identifiers could be transformative of daily life." (Wikipedia)

What is “Big Data”?

- Volume
- Variety
- Velocity

But Big Data Isn't the Real Story

- Data is data, however "big"
- The real story for environmental management and sustainability is ICT – Information and Communications Technology – and its use
- ICT is the mechanism whereby data, big and small, is collected, analyzed and acted upon

Hype v. Reality

- At this point, “big data” more potential than reality
- Intel estimates:
  - 6% of enterprises make decisions with big data analytics
  - 9% of enterprise workloads reside in public cloud
- Focus of this presentation is broader than “big data”
- Focus is on ICT – information and communications technology – and how it impacts corporate environmental management and sustainability generally

Hype Becoming Reality

"From the dawn of civilization until 2003, humankind generated 5 exabytes of data. Now we produce 5 exabytes every two days…and the pace is accelerating."

Eric Schmidt, Google

[1 exabyte = 10 x 18 0s bytes]

Hype Becoming Reality

- Some estimates show Gig Data growing at 40% per year
- From dawn of civilization to 2003, mankind generated 5 exabytes of data
- By 2012…2.7 zetabytes
- Exabyte = 1 x 18 0s bytes
- Zetabyte = 1 x 21 0s bytes
- IDC estimate: by 2015, 15 Billion devices connected to Internet

The “Internet of Things”

- Also known as the “Industrial Internet”
- Exponential growth in sensing and silicon – “Intelligence Everywhere”
- Eventually, the network will be “like the wind” – you will see its impact, but not the thing itself

Aggregate Demand for Computing Accelerating

What I am Going to Cover

- Corporate environmental management applications of big data and ICT
- The “macro story” – the enabling impacts of ICT
- “Smart Grid” and building energy management applications
- “Smart City” applications
- Water and natural resource applications
- BUT, I am not going to address Privacy and Security…

Big Data for Corporate Environmental Management

- Performance benchmarking against competitors
  - E.g., SIA and World Semiconductor Council
- Meeting reporting standards
    - E.g., GRI, CDP, UN Global Compact
- Establish predictive capabilities
- Sound corporate data management suggests that a first step is to inventory corporate data needs/requirements and sources
  - Regulatory
  - Voluntary (e.g., CDP and ISO)
  - Stakeholder needs
- The more and more complex data you have, the more sense the ‘cloud’ makes, whether private or public
- Be wary of correlation v. causation difference

Different Stakeholders, Different Data Needs

- Governments
- Investors
- Employees
- Suppliers
- Community groups

Crowd-Sourced Environmental Data

- Air and water sensor technology is approaching relatively low price points – future is foreseeable whereby crowd- or citizen-sourcing of real-time ambient environmental data is practical
  - The “Arduino” platform is path breaking
- Intel’s “Frugal Innovations” laboratory example
- China’s Danger Maps, enabled by Baidu Maps, pinpoints pollution hotspots

Crowd-Sourced Data – A Two-Edged Sword

- As the air and water monitoring gets better, companies will be able to install very dense sensor networks and provide much greater, and more real-time performance transparency to their communities and other stakeholders

- Being such a transparency leader can create later embarrassment and legal liability if monitors reveal environmental “excursions”

- Crowd-source data may conflict with, and thereby challenge, the validity of existing data and models that air and water quality management programs are based upon

Testing and Registering Chemicals

- Recent Nobel Prize winners for Chemistry – their work mostly done on computers, using lots of data and sophisticated modeling techniques
- In the future, big data and computer modeling will play a growing role in “testing” chemicals and registering them under national regulatory regimes

Technology and Carbon Emissions

Drive Computing
to Be More
Energy Efficient

~2% Opportunity

Use Computing to Improve
Energy Savings Outside
Information and
Communications Technology

98% = The Big Opportunity

"Macro Story" Evidence

- American Council for an Energy-Efficient Economy (ACEEE) studied this issue and concluded:
  - ICT seen as a major factor in improving energy efficiency of US economy during the Internet era
  - "For every extra Kwh of electricity that has been demanded by ICT, the US economy increased its overall energy savings by a factor of about 10…" (2008)

- The Climate Group and the "Global e-Sustainability Initiative" published a report entitled, "*Smart 2020: Enabling the Low Carbon Economy in the Information Age*" (2008), concluding:
  - Smart 2020 concludes that ICT strategies could reduce **up to 15% percent** of global emissions in 2020 against a "business as usual" baseline

- US Addendum to *Smart 2020* report, prepared by Boston Consulting Group indicates that ICT strategies could reduce US carbon emissions by **up to 22 percent** by 2020 vs. business-as-usual

- **TAKE AWAY**: ICT strategies offer huge potential for addressing climate challenge BUT there is a huge gap between actual performance (ACEEE) and potential (Smart 2020)

Macro Story – Increasing the EE of Other Sectors

Automation

Industrial Robots

Logistics for Transportation

LEED Certified Buildings

Smart Motors

Smart Power Delivery

Substitution

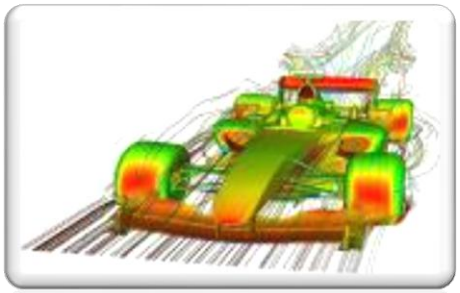
Video Conferencing

Online Entertainment

e-Commerce

Paperless Office

De-Materialization



Converting Atoms to Bits

On-line Banking

Digital Media Content

The “Cloud” as an Energy Efficiency Driver

- The Data Center and the network is at the center of the “macro story”

- The Carbon Disclosure Project (CDP) commissioned Verdantix to examine the impact of a broad US roll-out of Cloud Computing, based on extrapolation from existing case studies:
  - Huge CO2 emissions reductions
  - Huge financial savings
  - Strong positive financial ROI
  - Indirect benefits from increased business process efficiencies and organizational flexibility

SOURCE: The Carbon Disclosure Project/Verdantix: Cloud Computing: The IT Solution for the 21st Century

Big Data and the Smart Grid

Smarter Generation

Smarter Transmission and Distribution

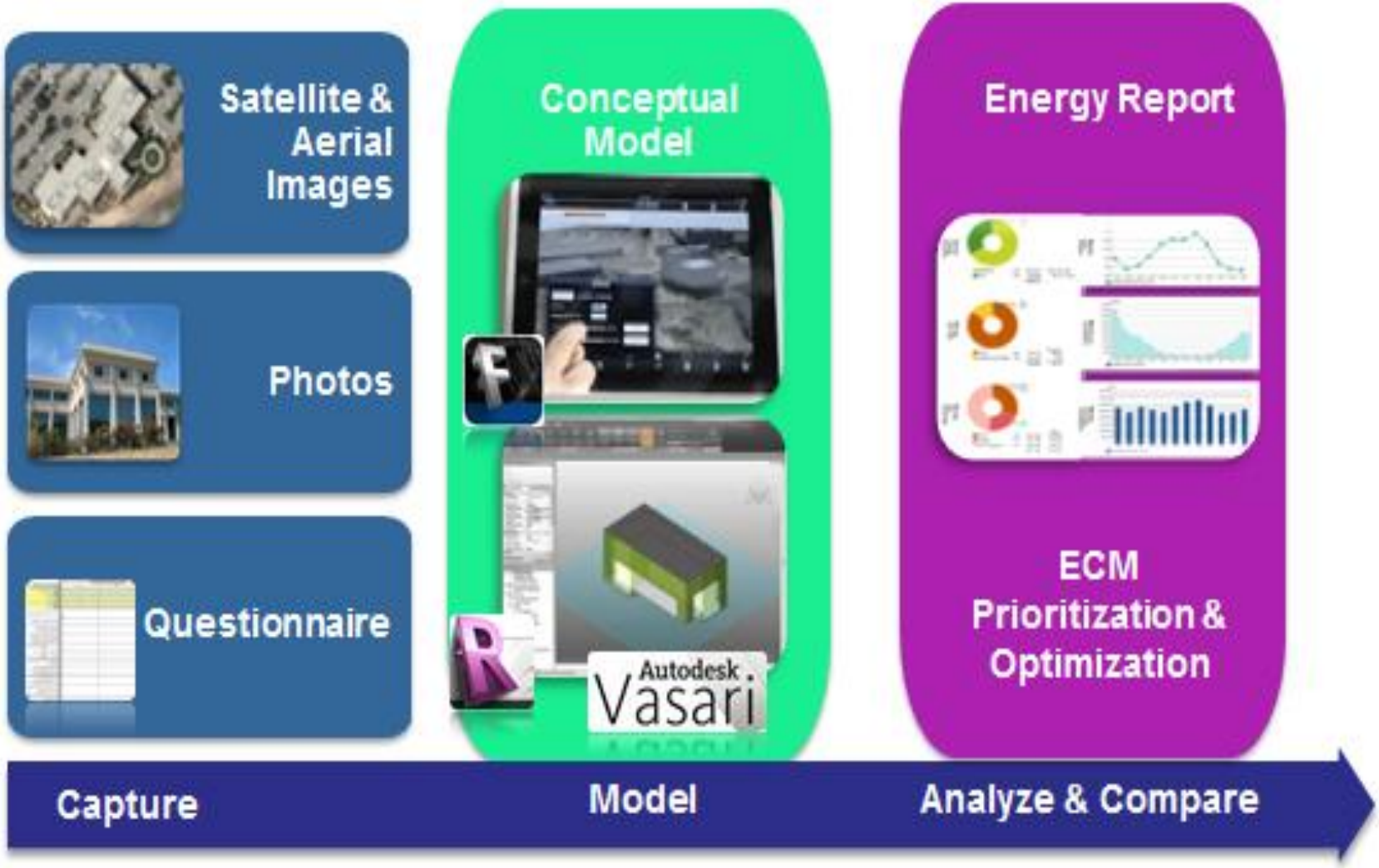
Smarter Demand-Side Management

Grid Security, Management & Communication

Whole-Grid Modeling, Simulation, Monitoring and Control

Rapid Energy Modeling Process

Source: Autodesk



Empire State Building - a private sector success story

Integrated whole-building retrofit

Reduce Loads

- Windows
- Insulation
- Daylighting
- Occupancy sensors

Upgrade Systems

- Chiller plant retrofit,
- VSDs
- VAV air handlers

Monitoring & Controls

- Direct digital controls
- Demand-controlled ventilation
- Tenant energy mgmt system

RESULTS

38% energy savings

3 year payback on incremental cost

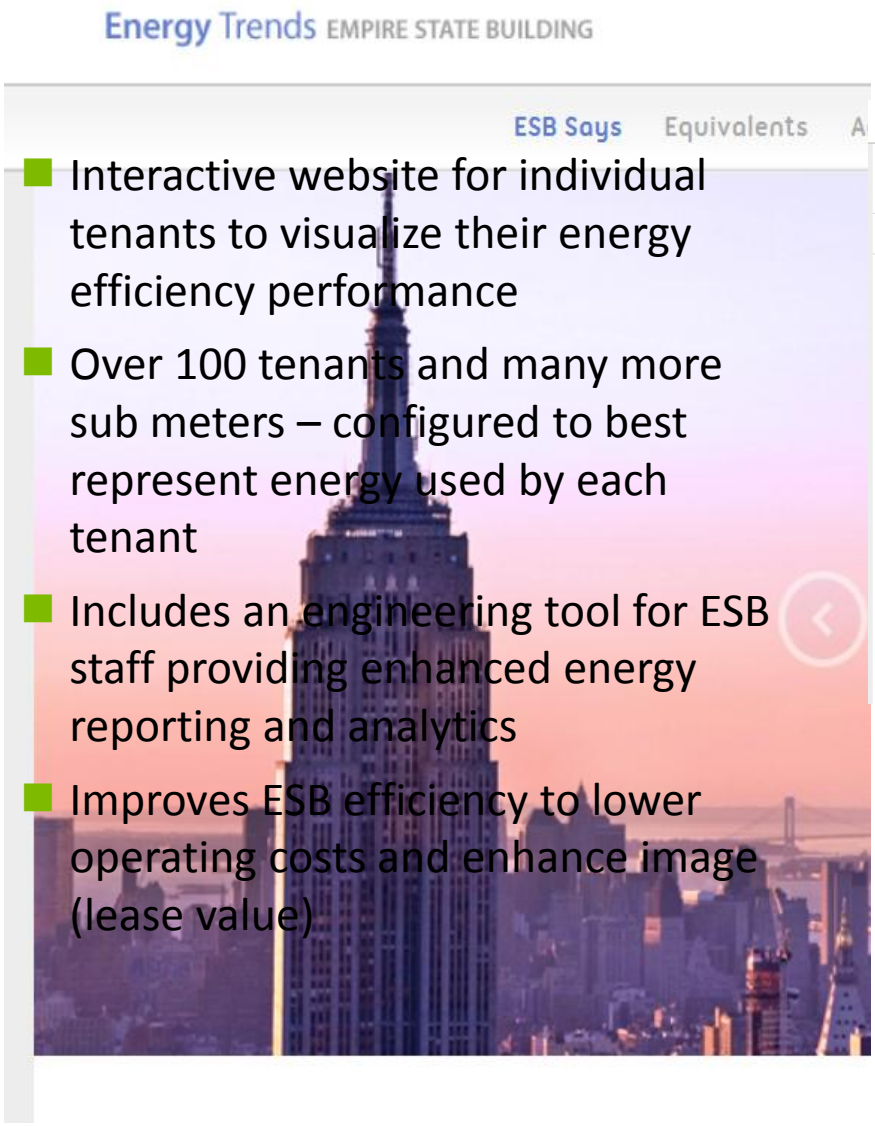
Anticipate **LEED Gold** certification

Energy Star **Top 10%** of U.S. office buildings

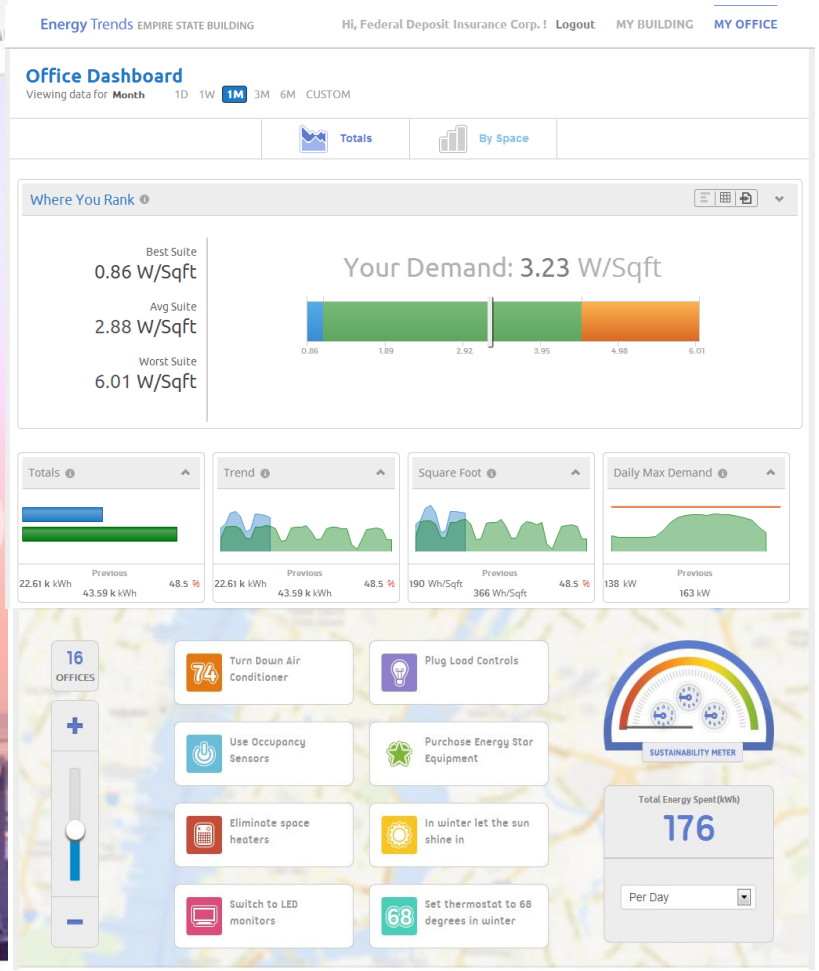
105,000 tons of CO_2e eliminated

Savings of **$4.4 million** annually

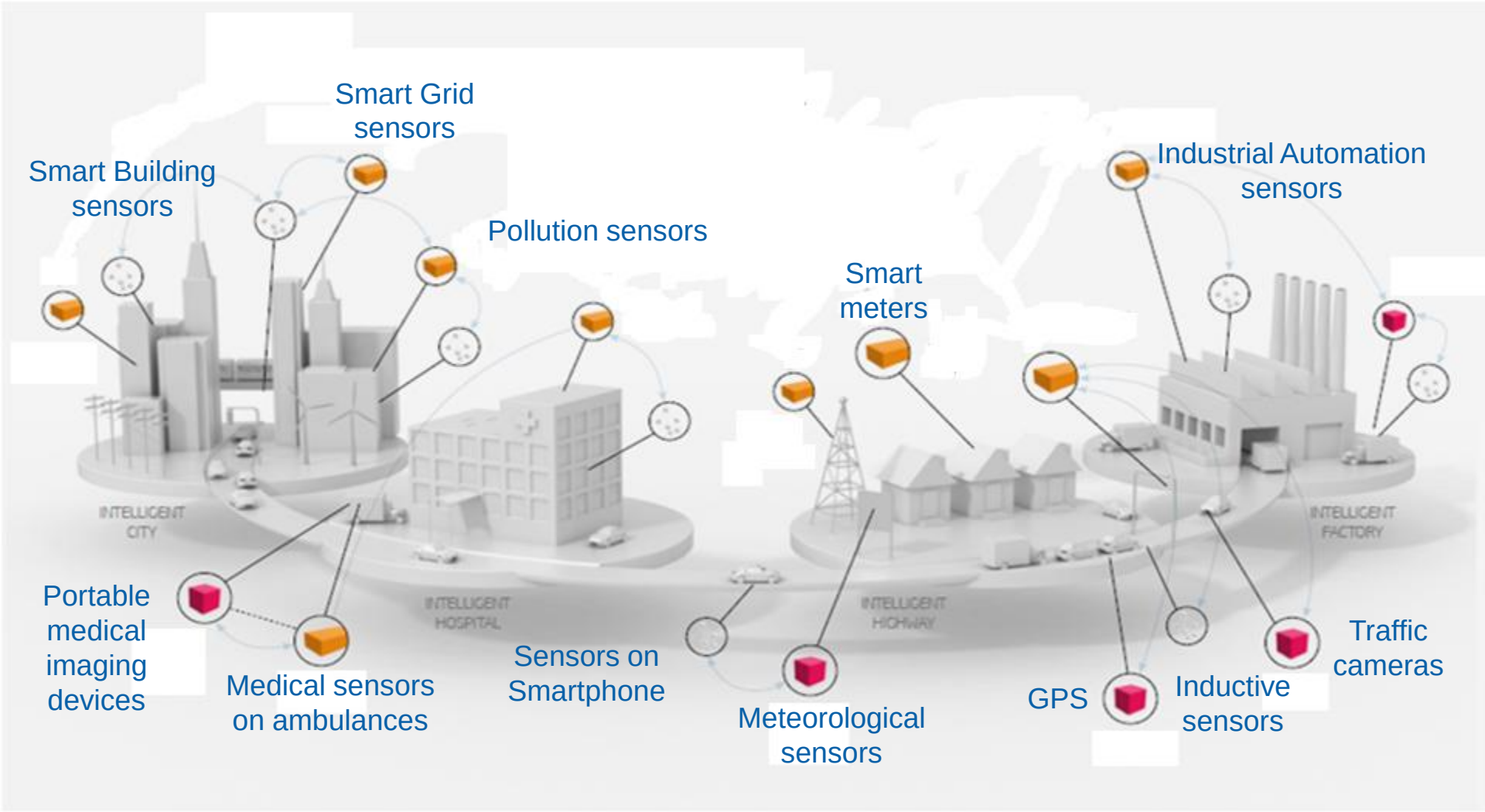
Empire State Building – Tenant Energy Portal



- Interactive website for individual tenants to visualize their energy efficiency performance
- Over 100 tenants and many more sub meters – configured to best represent energy used by each tenant
- Includes an engineering tool for ESB staff providing enhanced energy reporting and analytics
- Improves ESB efficiency to lower operating costs and enhance image (lease value)



“Smart Cities” --
A World of Big Sensed Data



Imperial College London

intel

UCL

Quadruple Helix In Action:
Sustainable Connected Cities
Dublin – CityWatch

Citizen-Enabled Sensor Networks

Cities Informing Its Citizens

Gamification and Incentivization

Intel Labs Europe

Dublin City Council
Comhairle Cathrach Bhaile Átha Cliath

TRINITY COLLEGE DUBLIN
COLÁISTE NA TRÍONÓIDE

Citizens

CityWatch

Citizen Powered – Better Living

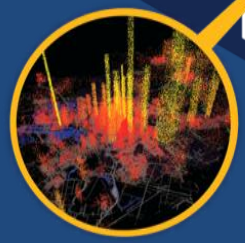
Data Fusion
Integration of heterogeneous urban data sets

Open Government Data

Gamification

Participatory Sensing
Citizen Mobile devices

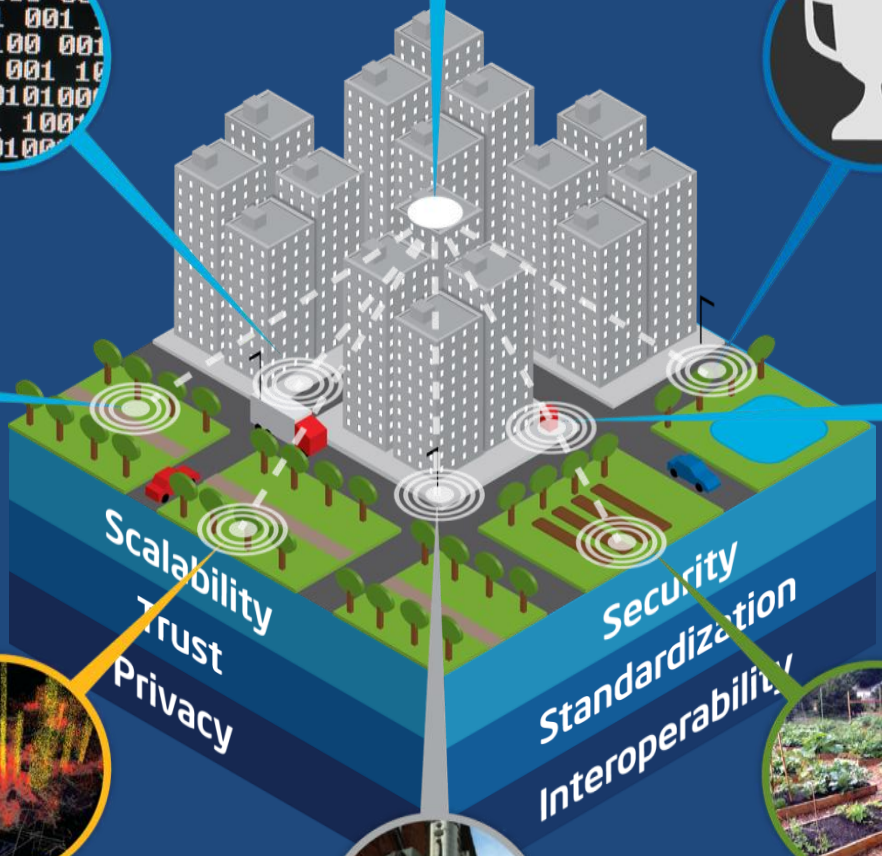
Citizen Centric Services



Opportunistic Sensing
System dynamically selects a sensor to collect data (e.g., temperature sensor on a smart vehicle)

Green City Map
Highlight initiatives, resources and events

Fixed Sensing
Sensors are fixed at a place in the environment (e.g., meteorological sensors, CCTV

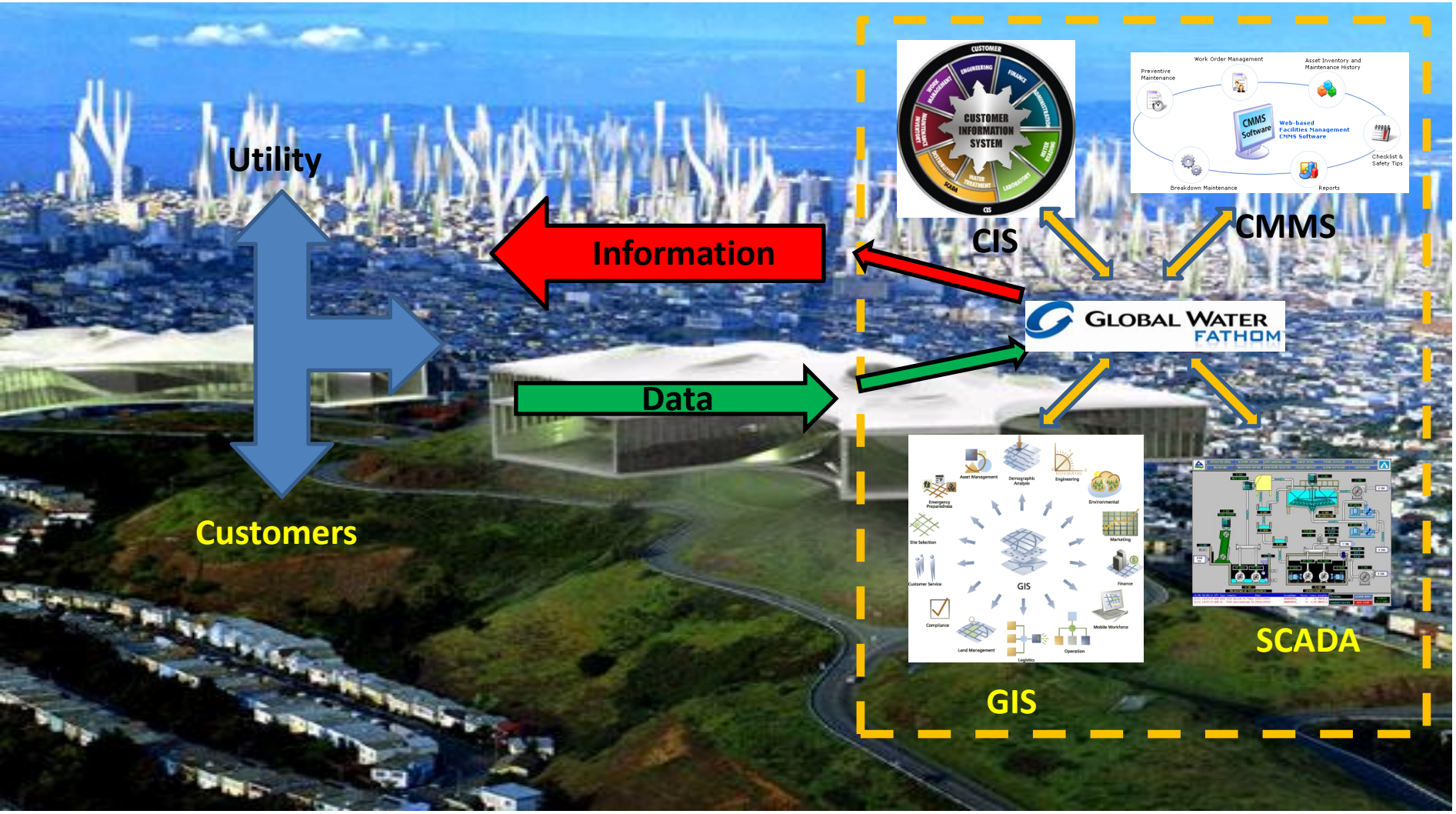


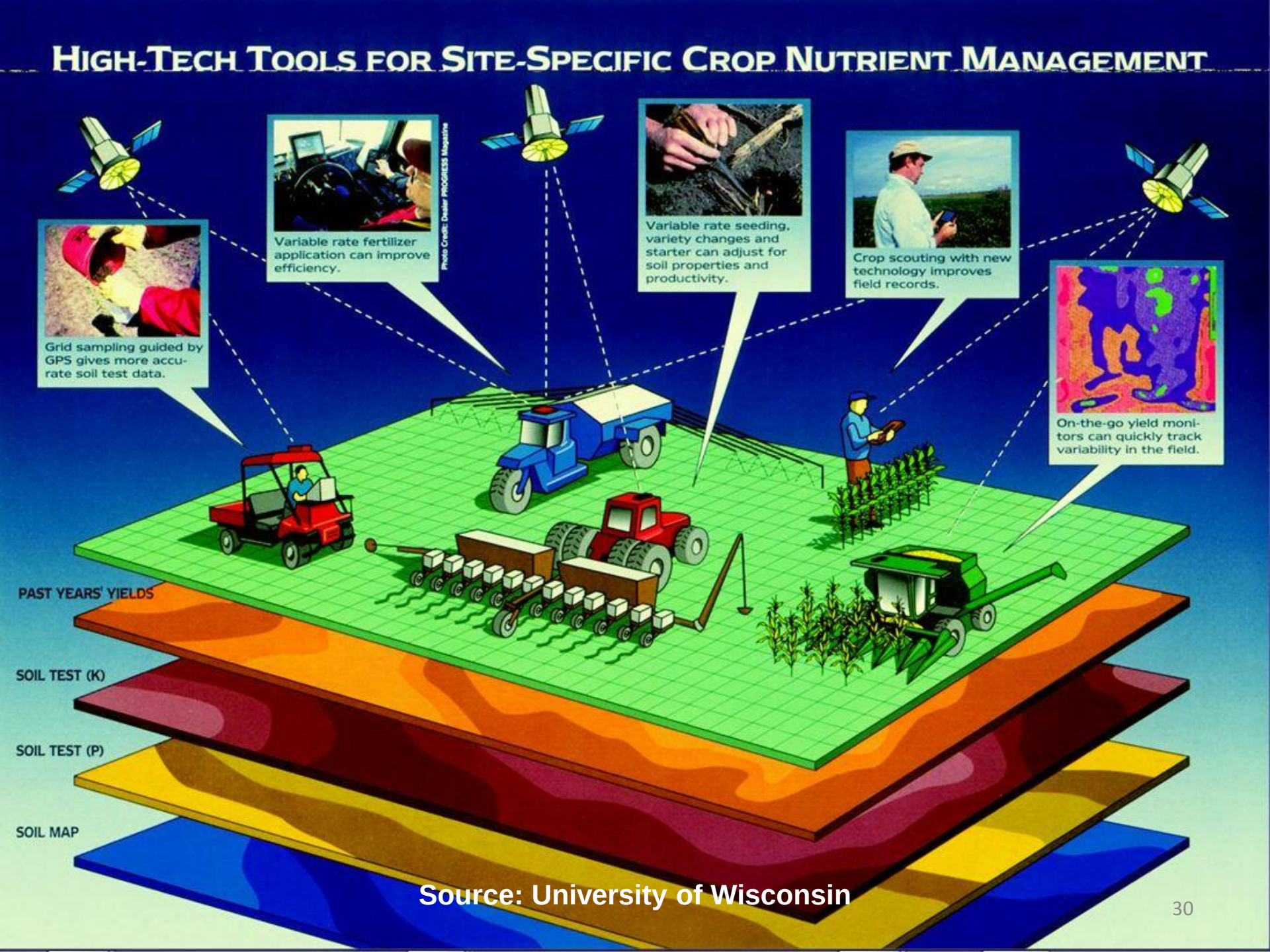
Big systems optimization: Smart parking

Motes use <400μW on Average. For LA, with 40,000 parking spots, that implies total mote power of about 15W. Mote technology is from Dust Networks

Source: Mark Noworolski, Streetline Networks, via Jonathan Koomey

Smart Grid for Water





Source: University of Wisconsin

Participatory Simulation Games at UVa

The project began in the Fall 2008 with a group of faculty members representing seven of the University's eleven schools.

The goal was to find a means to introduce students to the issue of sustainability in a complex socio-environmental-economic system.

the uva bay game™

Katie Jones Current Year: **2002**
Land Developer, James River Watershed, Virginia

LOGOUT / EXIT

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020

My Activity

VIEW MAP LAYERS

My Net Worth

Bay Health

My Financial Position

Cash:	$0
Land:	10 acres
Land & Equipment Value:	$254,876
Total Assets:	**$254,876**
Long-term Debt:	$133,436
Net Worth:	$121,440
Total Liabilities and Equity:	**$254,876**
Maximum Allowable Purchase:	$600,000
Revenue:	$3,600
+ Gain / Loss:	$0
- Operating Expenses:	$1,800
- Interest on Debt:	$0
- Taxes Paid:	$360
Net Income:	**$1,440**

View Financial Reports

Land Developer

Help

My Land Holdings:

	Acres	Value	Sell Acres
Infill Sustainable	10	$271,648	

Total Sale: $0

Purchase and Develop Land:

	Acres
Greenfield Conventional	
Greenfield Sustainable	
Infill Conventional	
Infill Sustainable	

Total Purchase: $0

Submit Decisions

Assets N P

Land Development: Conventional Greenfield

Cost to Buy and Develop: *per acre*	$15,422
Development Tax: *per acre*	$0
Value of Developed Land: *per acre*	$16,539

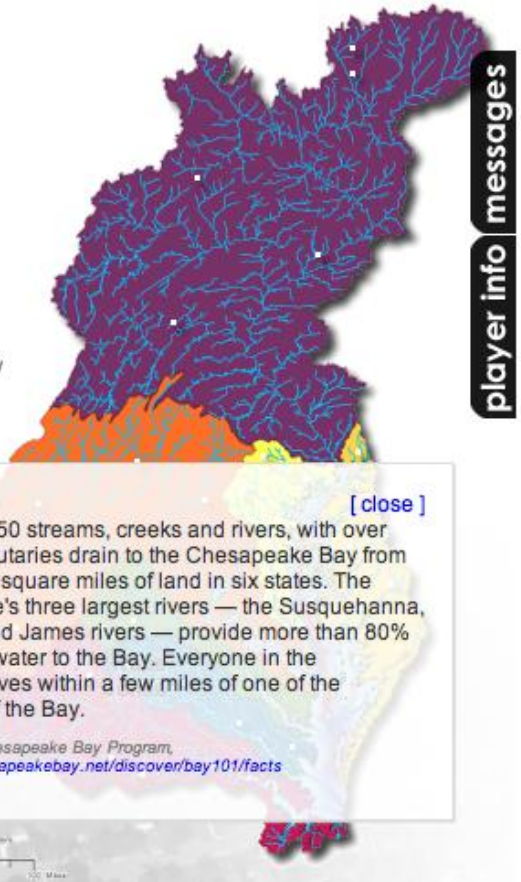
Revenue Comparison

Greenfield acreage is land in the rural areas which is currently forested, used as farmland, or otherwise vegitative. This land will be converted to developed land.

Conventional development practices yield the quickest and most cost effective turnaround, but often without regard to land disturbance, land planning, and it's long term environmental effects.

Learn More

Susquehanna
Potomac
Eastern Shore
Patuxent
Rappahannock
York
James
Chesapeake Bay



player info messages

[close]

More than 150 streams, creeks and rivers, with over 100,000 tributaries drain to the Chesapeake Bay from over 64,000 square miles of land in six states. The Chesapeake's three largest rivers — the Susquehanna, Potomac, and James rivers — provide more than 80% of the fresh water to the Bay. Everyone in the watershed lives within a few miles of one of the tributaries of the Bay.

Source: The Chesapeake Bay Program
http://www.chesapeakebay.net/discover/bay101/facts

Metrics Standings

Bay Health: 80.0 % Anoxic Region: .84 N: 191 M. lbs/yr P: 7 M. lbs/yr

[close]

Simulation Standings

	Rank	Watershed	Net Income (relative to year 2000)	Nutrient Runoff / Flow (relative to year 2000)	Nutrient Runoff / Population (relative to year 2000)	Overall Score
↑	1	Patuxent	1.04	0.93 N 0.95 P	1.08 N 1.12 P	98 / 100
→	2	James	1.04	1.10 N 1.14 P	1.34 N 1.4 P	84 / 100
↓	3	Eastern Shore	1.04	1.05 N 1.31 P	1.26 N 1.67 P	84 / 100
↓	4	Rappahannock	1.04	1.11 N 1.16 P	1.35 N 1.42 P	84 / 100
↓	5	York	1.04	1.12 N 1.17 P	1.37 N 1.44 P	80 / 100
↓	6	Potomac	1.03	1.10 N 1.24 P	1.35 N 1.56 P	79 / 100
↓	7	Susquehanna	1	1.07 N 1.29 P	1.32 N 1.67 P	62 / 100

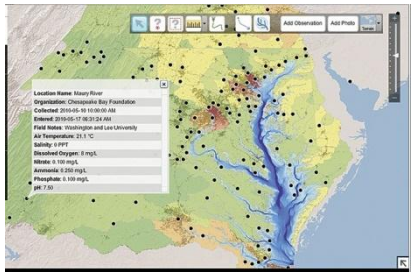
Click here for expanded standings

Big Data and Watershed Management

- Large, complex ecosystems like the Chesapeake Bay are rich sources of many types of data
- US EPA/State program for improving Bay water quality is all based on a very complex model built on millions of data points
- NGOs also use big data and ICT

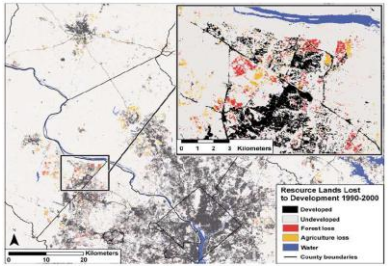
Combining Innovative Technologies with Large Landscape Conservation

Biological Field Assessments



Crowd Sourced Data

In-Situ River Water Quality Monitoring



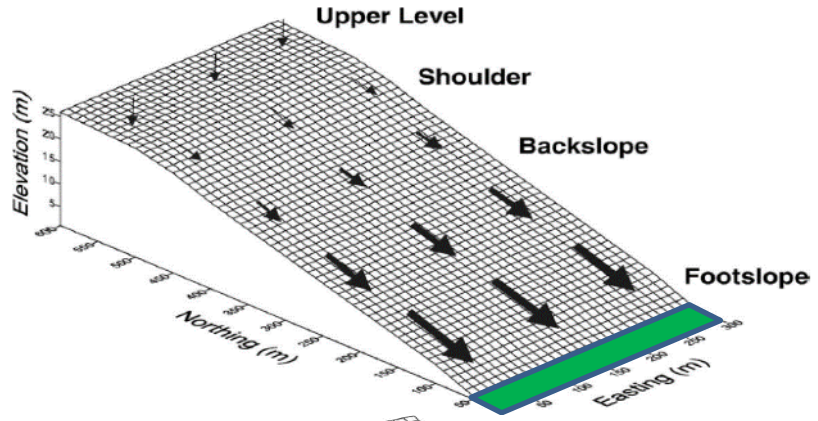
Remote Sensing Assessments

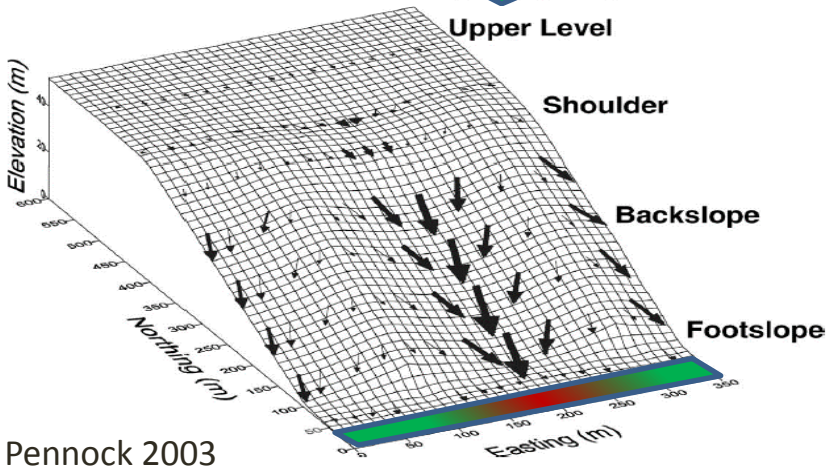
Better Understanding of Land Impacts on Water Quality

University Participation

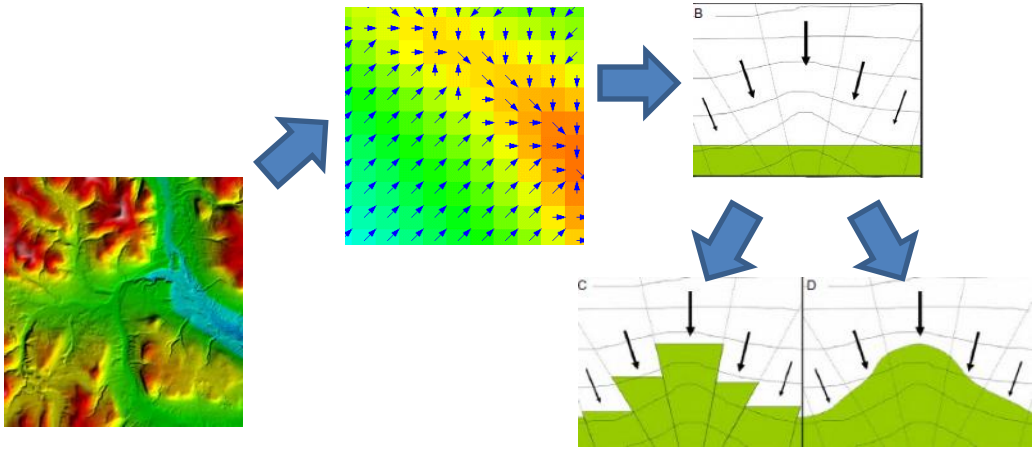
Concentrated Flow Path Mapping

- Most models assume consistent flow off the land
- In reality, some areas get more runoff than others, which can overwhelm buffers' filtering capacity
- Can help determine optimal buffer placement and size

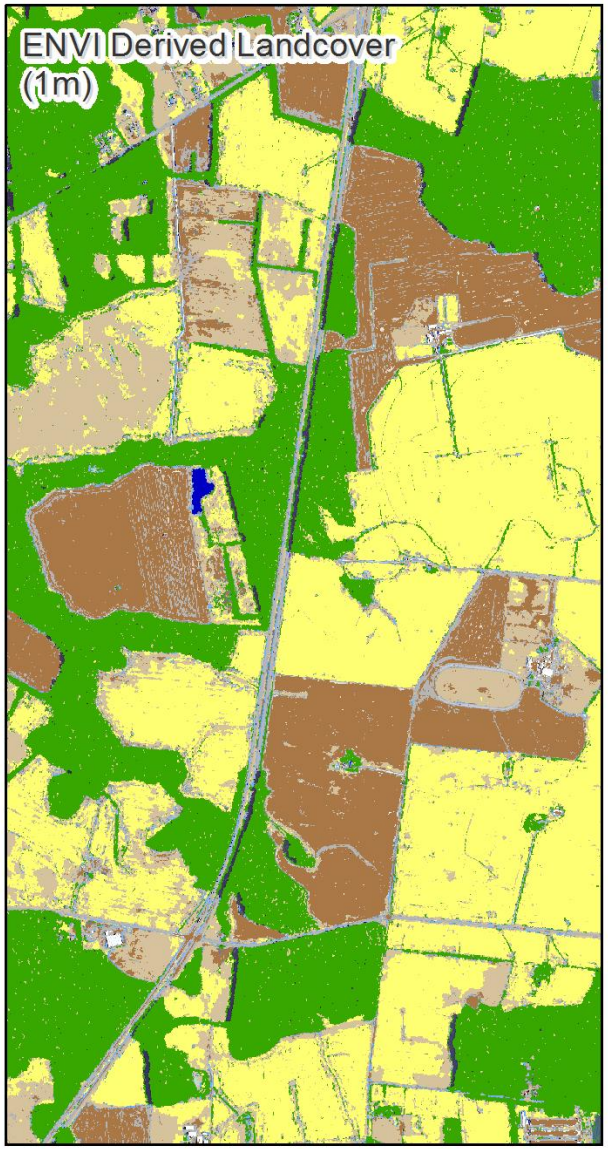


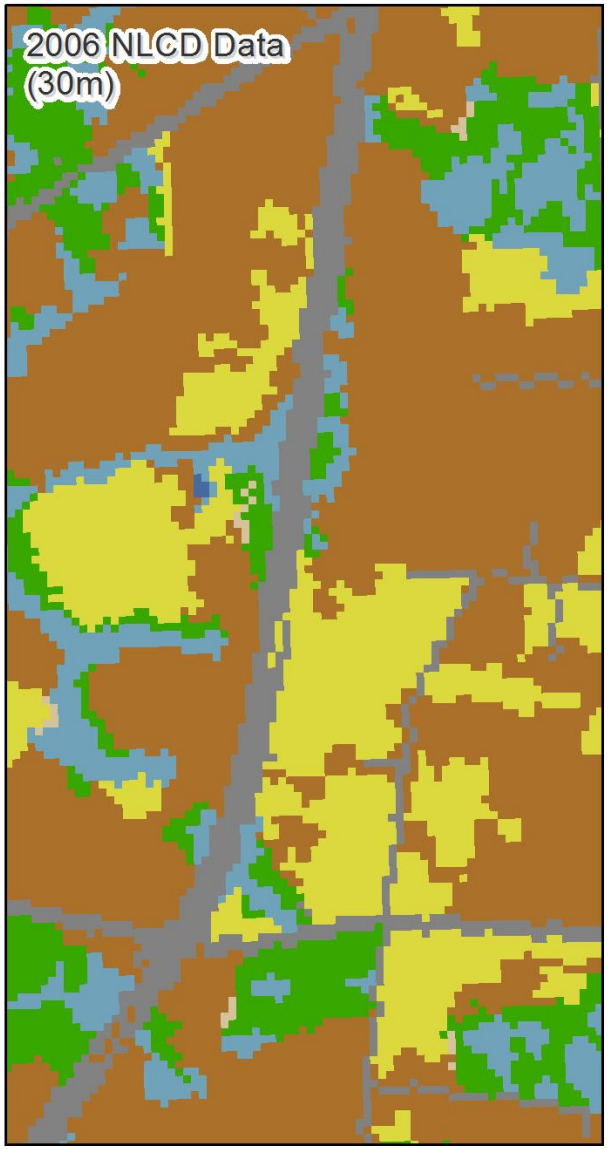


Pennock 2003



Chesapeake Conservancy





Crops
Forest
Water
Shrub/Scrub
Wetlands
Tilled Field
Impervious

Comparison of 1m and 30m Landcover Data in Chester River Watershed

Barren Land
Cultivated Crops
Forest
Developed
Wetlands
Hay/Pasture
Open Water
Shrub/Scrub

Big Data for Development – the UN's Global Pulse

- Launched by UN in 2009, aims to turn big data flows into actionable intelligence

- Uses computational techniques to unveil trends and patterns within and between very large socioeconomic datasets, including:
  - Mobile phone billings
  - Health hotlines
  - Social media

- Data sources must be: digitally generated, passively produced, automatically collected, geographically or temporally trackable, and amenable to analysis

- Develop early warnings of economic distress, epidemics and other disruptions, enabling earlier intervention

Wonderbag

- Designed by South African entrepreneur, the Wonderbag saves lives, frees women from the home, and can reduce carbon emissions*
- Can be outfitted with a heat sensor and radio chip to keep track of carbon savings

*1/2 ton per year of C saved if used 3x/week