Big and Open Data for the Sustainable Development Goals (SDGs)

Alex de Sherbinin, PhD
Associate Director for Science Applications
Center for International Earth Science Information Network (CIESIN)
The Earth Institute at Columbia University

Chair, CODATA Roads Data Development Group

Co-Coordinator, Population-Environment Research Network (PERN)

Kavli Symposium
16 February 2015

Acknowledgments

- Jessica Espey, Program Leader, Monitoring and Accountability for Sustainable Development, Sustainable Development Solutions Network
- Emannuel Letouzé, Director, Datapop Alliance
- Bob Chen and Marc Levy, Director & Deputy Director, CIESIN

Outline

1. A word from our sponsors
2. What are the SDGs and why are they important?
3. Why are data and monitoring important for sustainable development?
4. Can “big data” fill all our data gaps?
5. Bringing it all together

1. A WORD FROM OUR SPONSORS

CIESIN Background

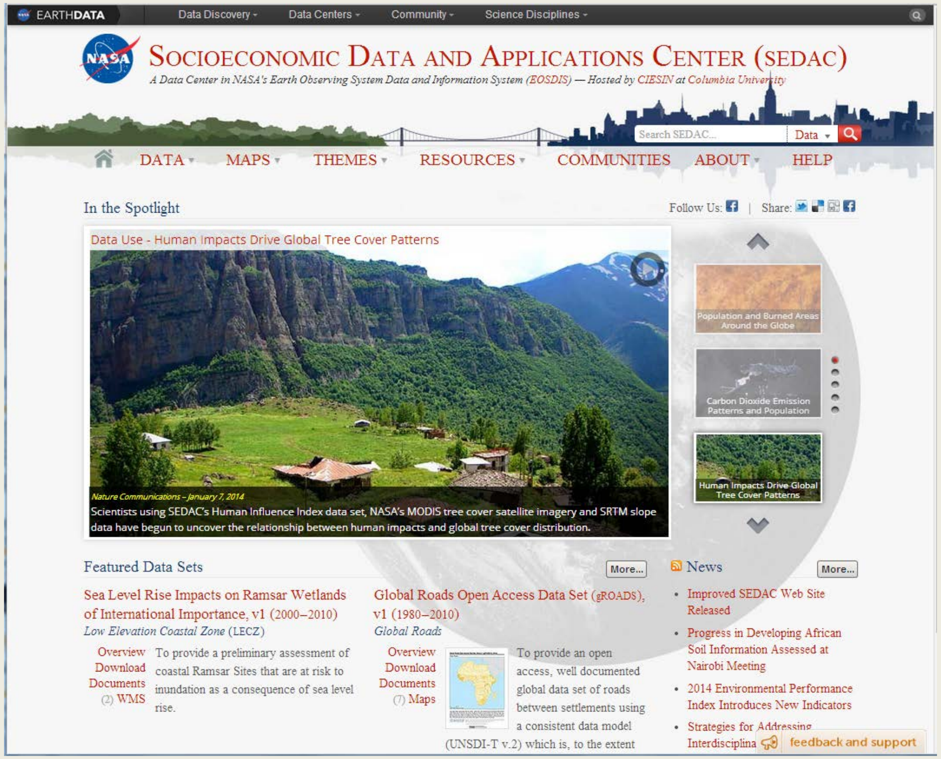
CIESIN
The Earth Institute at Columbia University

- CIESIN founded in 1989 as a non-profit consortium based in Michigan
- In July 1998, CIESIN became a center within the Earth Institute
- CIESIN has more than 40 professional staff from the social and natural sciences, information technology & data management plus many students, postdocs, and visitors
- CIESIN has 4 divisions:
  - Science Applications
  - Information Services
  - Information Technology
  - Geospatial Applications

Center for International Earth Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

NASA Socioeconomic Data and Applications Center (SEDAC)

CIESIN
The Earth Institute at Columbia University



SOCIOECONOMIC DATA AND APPLICATIONS CENTER (SEDAC)

A Data Center in NASA's Earth Observing System Data and Information System (EOSDIS) — Hosted by CIESIN at Columbia University

In the Spotlight

Data Use - Human Impacts Drive Global Tree Cover Patterns

Nature Communications – January 7, 2014

Scientists using SEDAC's Human Influence Index data set, NASA's MODIS tree cover satellite imagery and SRTM slope data have begun to uncover the relationship between human impacts and global tree cover distribution.

Featured Data Sets

Sea Level Rise Impacts on Ramsar Wetlands of International Importance, v1 (2000–2010)
Low Elevation Coastal Zone (LECZ)
Overview
Download
Documents
(2) WMS
To provide a preliminary assessment of coastal Ramsar Sites that are at risk to inundation as a consequence of sea level rise.

Global Roads Open Access Data Set (gROADS), v1 (1980–2010)
Global Roads
Overview
Download
Documents
(7) Maps
To provide an open access, well documented global data set of roads between settlements using a consistent data model (UNSDI-T v.2) which is, to the extent

News

- Improved SEDAC Web Site Released
- Progress in Developing African Soil Information Assessed at Nairobi Meeting
- 2014 Environmental Performance Index Introduces New Indicators
- Strategies for Addressing Interdisciplina

- Close to 200 global spatial and tabular data sets
- Focus on human dimensions of environmental change and the integration of social and Earth science data, especially with remote sensing
- Provide direct support to scientists, applied and operational users, decision makers, and the media!

SEDAC

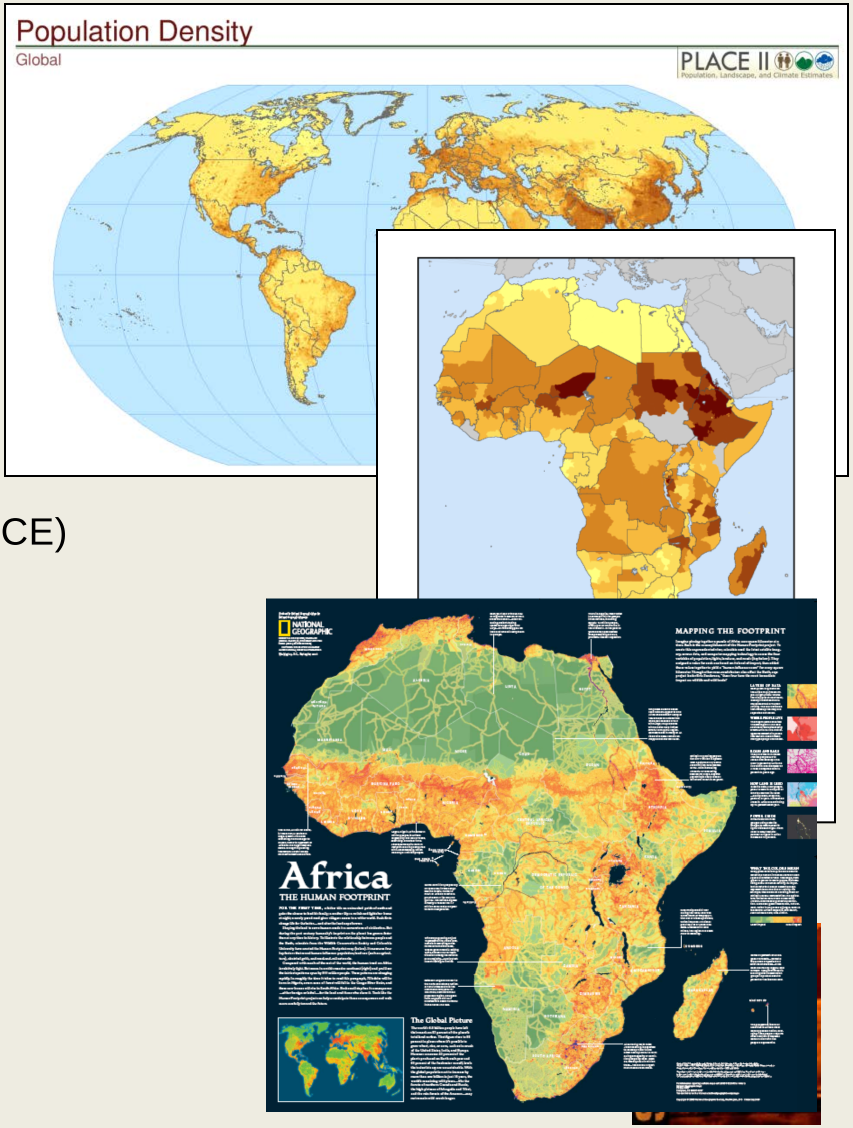
Key SEDAC Global Environmental & Socioeconomic Datasets

http://sedac.ciesin.columbia.edu

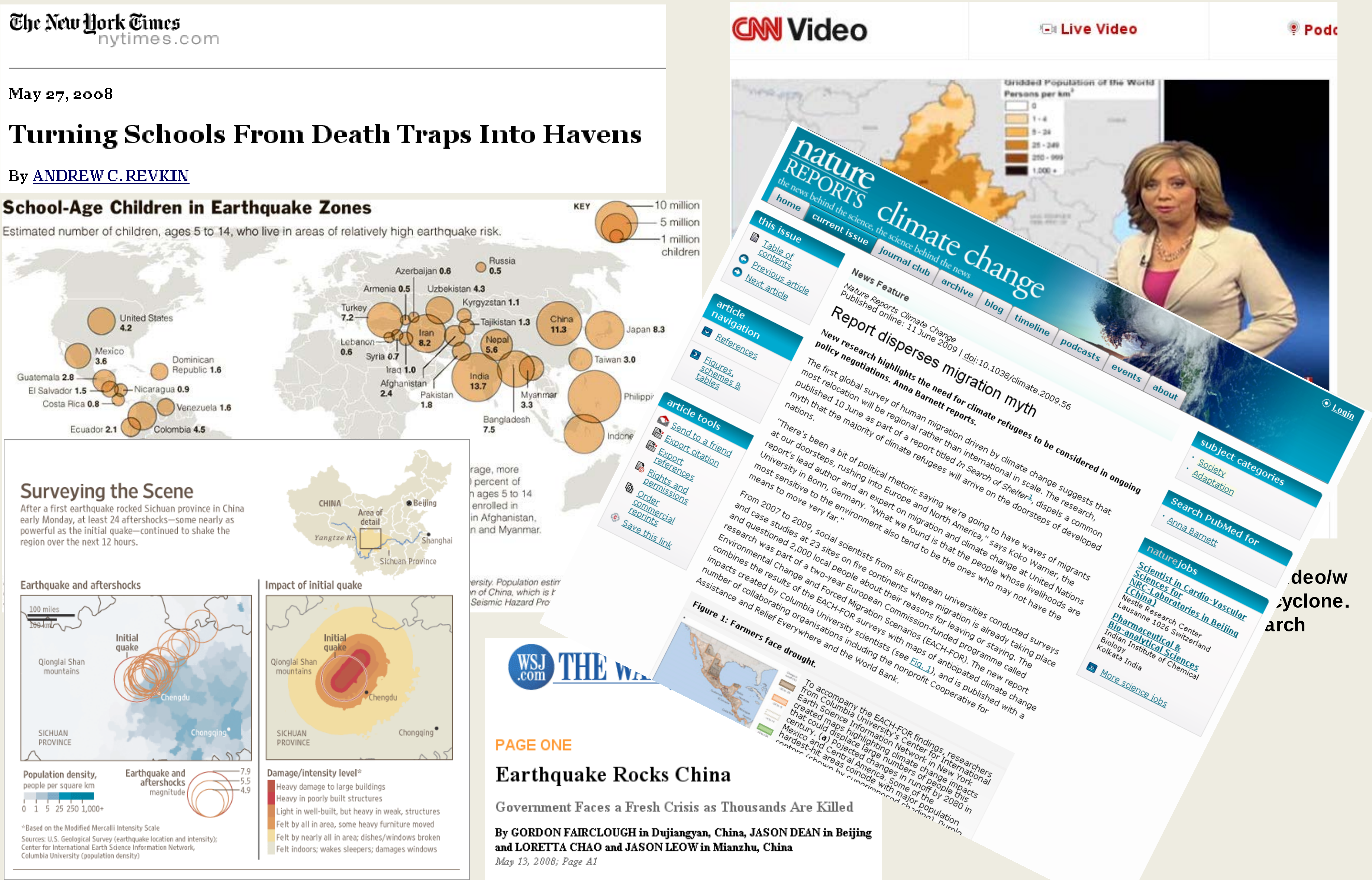
CIESIN

The Earth Institute at Columbia University

- Population Grids
  - Gridded Population of the World, Version 3 (GPW)
  - Global Rural-Urban Mapping Project (GRUMP)
  - US Census Grids
- Poverty Maps
  - Global Distribution of Poverty
  - Global Child Malnutrition
- Climate Related
  - Low Elevation Coastal Zone
  - Ramsar Wetlands at Risk to Sea Level Rise
  - Population, Landscape, And Climate Estimates (PLACE)
  - Global Natural Disaster Hotspots
- Environmental Indicators
  - Environmental Sustainability Index (ESI)
  - Environmental Performance Index (EPI)
  - Human Appropriation of Net Primary Productivity
  - Human Footprint/Last of the Wild
  - Natural Resource Management Index (NRMI)
- Infrastructure
  - Global Roads Open Access Data Set (gROADS)
  - Global Reservoirs and Dams (GRanD)
  - Nuclear Power Plants



News Media Use of SEDAC Data



The New York Times
nytimes.com

May 27, 2008

Turning Schools From Death Traps Into Havens

By ANDREW C. REVKIN

School-Age Children in Earthquake Zones

Estimated number of children, ages 5 to 14, who live in areas of relatively high earthquake risk.

rage, more
percent of
n ages 5 to 14
enrolled in
in Afghanistan,
n and Myanmar.

ersity. Population estir
n of China, which is t
Seismic Hazard Pro

Surveying the Scene

After a first earthquake rocked Sichuan province in China early Monday, at least 24 aftershocks—some nearly as powerful as the initial quake—continued to shake the region over the next 12 hours.

Earthquake and aftershocks

Impact of initial quake

Population density, people per square km
0 1 5 25 250 1,000+
Earthquake and aftershocks magnitude
7.9
5.5
4.9

Damage/intensity level*

- Heavy damage to large buildings
- Heavy in poorly built structures
- Light in well-built, but heavy in weak, structures
- Felt by all in area, some heavy furniture moved
- Felt by nearly all in area; dishes/windows broken
- Felt indoors; wakes sleepers; damages windows

*Based on the Modified Mercalli Intensity Scale

Sources: U.S. Geological Survey (earthquake location and intensity); Center for International Earth Science Information Network, Columbia University (population density)

CNN Video

nature REPORTS climate change
the news behind the science, the science behind the news

this issue
- Table of contents
- Previous article
- Next article

article navigation
- References
- Figures, schemes & tables

article tools
- Send to a friend
- Export citation
- Export references
- Rights and permissions
- Order commercial reprints
- Save this link

News Feature

Nature Reports Climate Change
Published online: 11 June 2009 | doi:10.1038/climate.2009.56

Report disperses migration myth

New research highlights the need for climate refugees to be considered in ongoing policy negotiations. Anna Barnett reports.

The first global survey of human migration driven by climate change suggests that most relocation will be regional rather than international in scale. The research, published 10 June as part of a report titled *In Search of Shelter*[1], dispels a common myth that the majority of climate refugees will arrive on the doorsteps of developed nations.

"There's been a bit of political rhetoric saying we're going to have waves of migrants at our doorsteps, rushing into Europe and North America," says Koko Warner, the report's lead author and an expert on migration and climate change at United Nations University in Bonn, Germany. "What we found is that the people whose livelihoods are most sensitive to the environment also tend to be the ones who may not have the means to move very far."

From 2007 to 2009, social scientists from six European universities conducted surveys and case studies at 23 sites on five continents where migration is already taking place and questioned 2,000 local people about their reasons for leaving or staying. The research was part of a two-year European Commission-funded programme called Environmental Change and Forced Migration Scenarios (EACH-FOR). The new report combines the results of the EACH-FOR surveys with maps of anticipated climate change impacts created by Columbia University scientists (see Fig. 1), and is published with a number of collaborating organisations including the nonprofit Cooperative for Assistance and Relief Everywhere and the World Bank.

Figure 1: Farmers face drought.

To accompany the EACH-FOR findings, researchers from Columbia University's Center for International Earth Science Information Network in New York created maps highlighting climate change impacts that could displace large numbers of people this century. (**a**) Pojected changes in runoff by 2080 in Mexico and Central America. Some of the hardest-hit areas coincide with major population centers (shown by superimposed shading). Purple

subject categories
- Society
- Adaptation

Search PubMed for
- Anna Barnett

naturejobs

Scientist in Cardio-Vascular Sciences for NRC-Laboratories in Beijing (China)
Nestle Research Center
Lausanne 1026 Switzerland

Pharmaceutical & Bio-analytical Sciences
Indian Institute of Chemical Biology
Kolkata India

More science jobs

deo/w
cyclone.
arch

WSJ.com THE WA

PAGE ONE

Earthquake Rocks China

Government Faces a Fresh Crisis as Thousands Are Killed

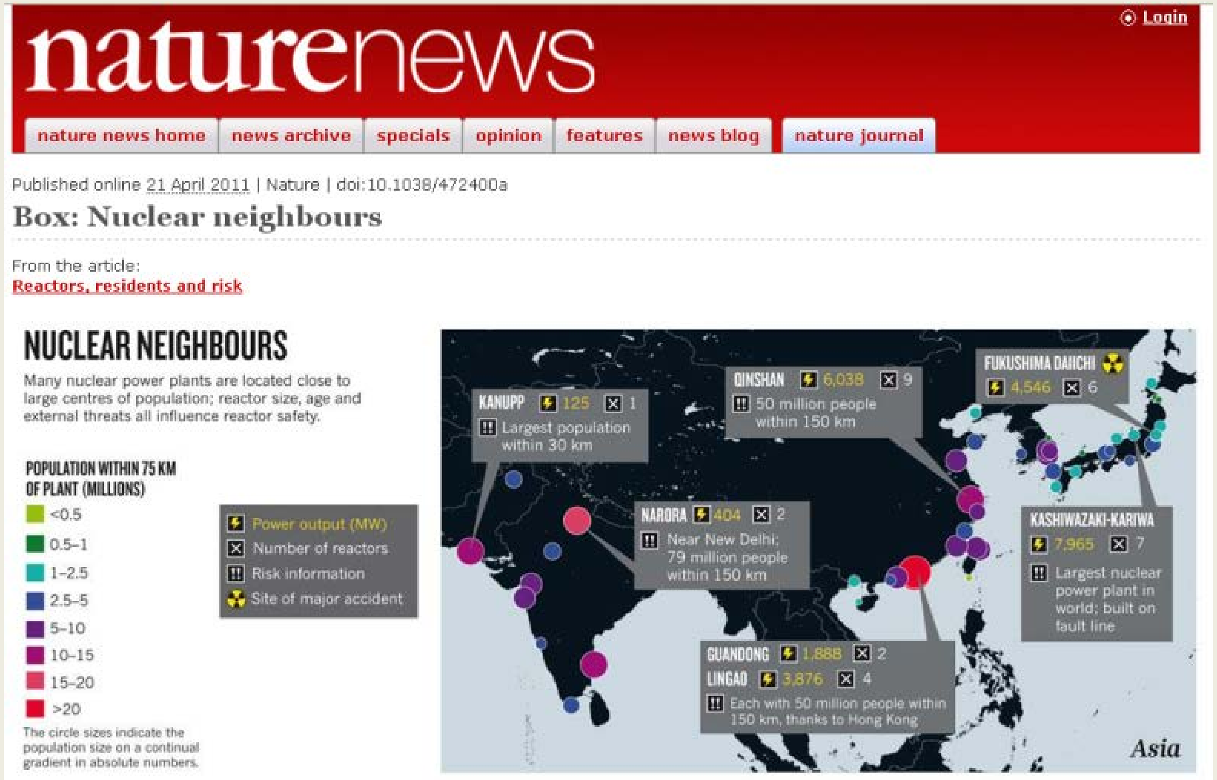
By GORDON FAIRCLOUGH in Dujiangyan, China, JASON DEAN in Beijing and LORETTA CHAO and JASON LEOW in Mianzhu, China

May 13, 2008; Page A1

Population Exposure Estimates in Proximity to Nuclear Power Plants

CIESIN
The Earth Institute at Columbia University

Country	Population within 150km (2010)
USA	180,329,616
China	169,811,536
Japan	94,793,752
India	94,382,912
UK	49,826,796
Germany	47,782,064
France	39,449,480
Spain	20,467,222
Pakistan	18,489,630
Korea	18,230,834
Argentina	18,199,546
Taiwan	15,904,056
Canada	15,070,809
Russia	14,983,205
Netherlands	12,967,174
Brazil	12,933,438
Belgium	10,697,587
Ukraine	7,179,316
Austria	7,012,218



naturenews

○ Login

nature news home | news archive | specials | opinion | features | news blog | nature journal

Published online 21 April 2011 | Nature | doi:10.1038/472400a

Box: Nuclear neighbours

From the article:
Reactors, residents and risk

NUCLEAR NEIGHBOURS

Many nuclear power plants are located close to large centres of population; reactor size, age and external threats all influence reactor safety.

POPULATION WITHIN 75 KM OF PLANT (MILLIONS)

- <0.5
- 0.5–1
- 1–2.5
- 2.5–5
- 5–10
- 10–15
- 15–20
- >20

The circle sizes indicate the population size on a continual gradient in absolute numbers.

- Power output (MW)
- ☒ Number of reactors
- !! Risk information
- Site of major accident

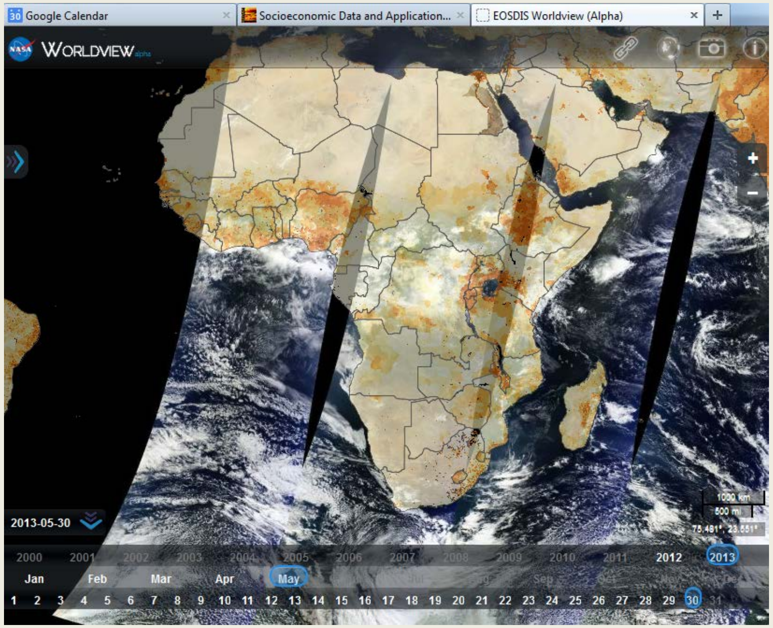
KANUPP 125 ☒ 1
!! Largest population within 30 km
QINSHAN 6,038 ☒ 9
!! 50 million people within 150 km
FUKUSHIMA DAIICHI
4,546 ☒ 6
NARORA 404 ☒ 2
!! Near New Delhi: 79 million people within 150 km
KASHIWAZAKI-KARIWA
7,965 ☒ 7
!! Largest nuclear power plant in world; built on fault line
GUANDONG 1,888 ☒ 2
LINGAO 3,876 ☒ 4
!! Each with 50 million people within 150 km, thanks to Hong Kong
Asia

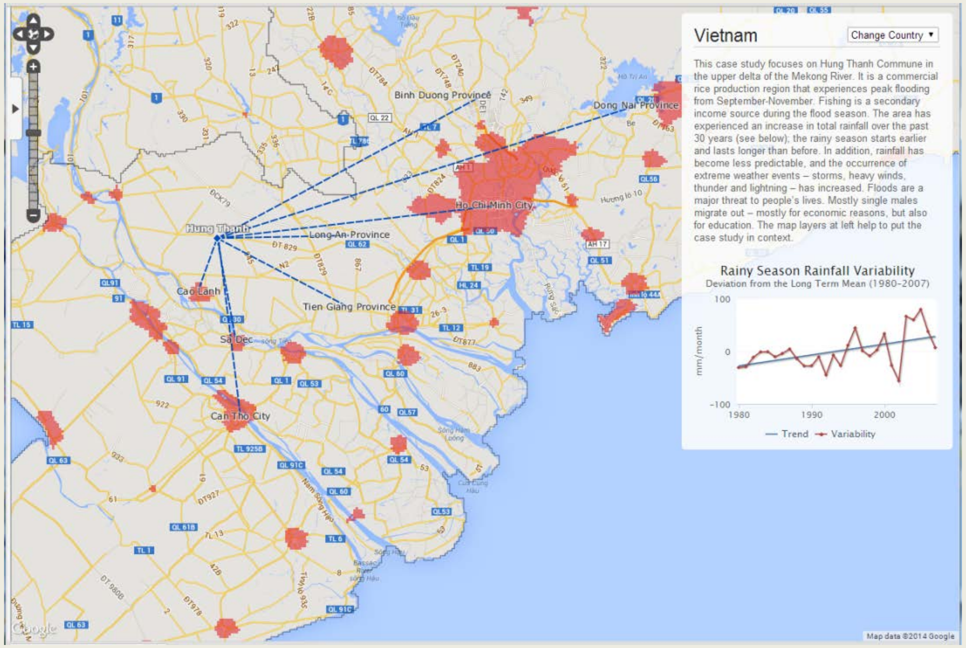
- Started with a news story
- Became a data set

Available at http://sedac.ciesin.columbia.edu/data/collection/energy

Map Clients

- NASA Worldview (http://earthdata.nasa.gov/labs/worldview/)
- NASA Climate and Health ANalysis for Global Education (CHANGE) Viewer (http://climatechangehumanhealth.org/changeviewer)
- SEDAC Map Client (http://sedac.ciesin.columbia.edu/maps/client)
- Where the Rainfalls: Rainfall variability and migration project (CARE, UNU & CIESIN) (http://ciesin.columbia.edu/care-mapclient/?iso=vnm)



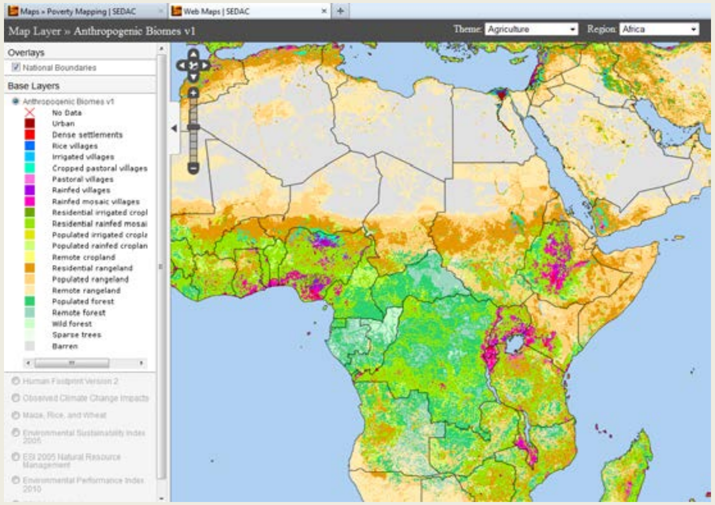


Vietnam

Change Country ▾

This case study focuses on Hung Thanh Commune in the upper delta of the Mekong River. It is a commercial rice production region that experiences peak flooding from September-November. Fishing is a secondary income source during the flood season. The area has experienced an increase in total rainfall over the past 30 years (see below); the rainy season starts earlier and lasts longer than before. In addition, rainfall has become less predictable, and the occurrence of extreme weather events – storms, heavy winds, thunder and lightning – has increased. Floods are a major threat to people's lives. Mostly single males migrate out – mostly for economic reasons, but also for education. The map layers at left help to put the case study in context.

Rainy Season Rainfall Variability
Deviation from the Long Term Mean (1980-2007)



Map Layer » Anthropogenic Biomes v1

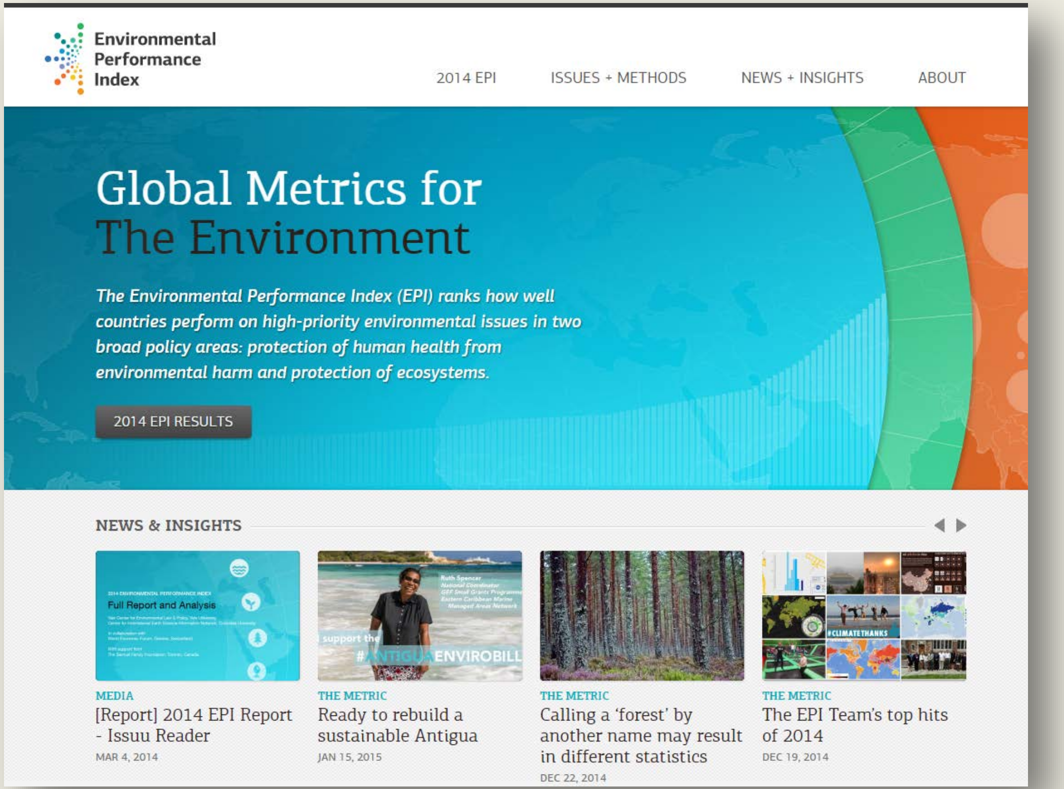
Overlays
☑ National Boundaries

Base Layers
● Anthropogenic Biomes v1
- ✕ No Data
- Urban
- Dense settlements
- Rice villages
- Irrigated villages
- Cropped pastoral villages
- Pastoral villages
- Rainfed villages
- Rainfed mosaic villages
- Residential irrigated cropl
- Residential rainfed mosai
- Populated irrigated cropla
- Populated rainfed croplan
- Remote cropland
- Residential rangeland
- Populated rangeland
- Remote rangeland
- Populated forest
- Remote forest
- Wild forest
- Sparse trees
- Barren

○ Human Footprint Version 2
○ Observed Climate Change Impacts
○ Maize, Rice, and Wheat
○ Environmental Sustainability Index 2005
○ ESI 2005 Natural Resource Management
○ Environmental Performance Index 2010

Yale/CIESIN Environmental Performance Index (EPI)

CIESIN
The Earth Institute at Columbia University



Environmental Performance Index

2014 EPI ISSUES + METHODS NEWS + INSIGHTS ABOUT

Global Metrics for The Environment

The Environmental Performance Index (EPI) ranks how well countries perform on high-priority environmental issues in two broad policy areas: protection of human health from environmental harm and protection of ecosystems.

2014 EPI RESULTS

NEWS & INSIGHTS

MEDIA
[Report] 2014 EPI Report - Issuu Reader
MAR 4, 2014

THE METRIC
Ready to rebuild a sustainable Antigua
JAN 15, 2015

THE METRIC
Calling a 'forest' by another name may result in different statistics
DEC 22, 2014

THE METRIC
The EPI Team's top hits of 2014
DEC 19, 2014

http://epi.yale.edu

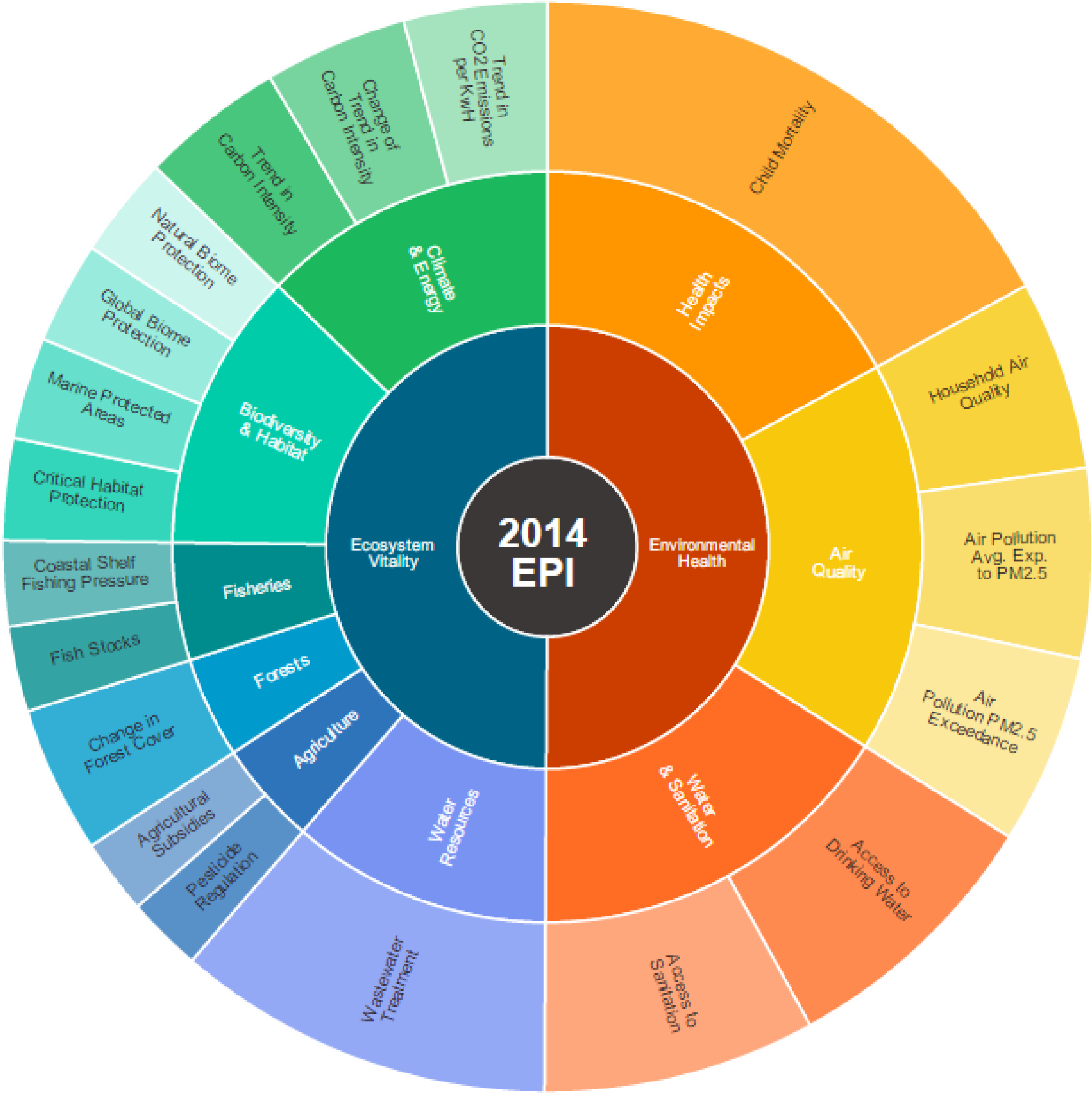
EPI Framework

ENVIRONMENTAL HEALTH

- Health Impacts
- Air Quality
- Water and Sanitation

ECOSYSTEM VITALITY

- Water Resources
- Agriculture
- Forests
- Fisheries
- Biodiversity and Habitat
- Climate and Energy



Spin offs

- National EPIs:
  - Abu Dhabi
  - China
  - Malaysia
  - Vietnam & India (forthcoming)
- Aquaculture Performance Index
- Ocean Health Index

ICSU
International Council for Science

ICSU
CODATA

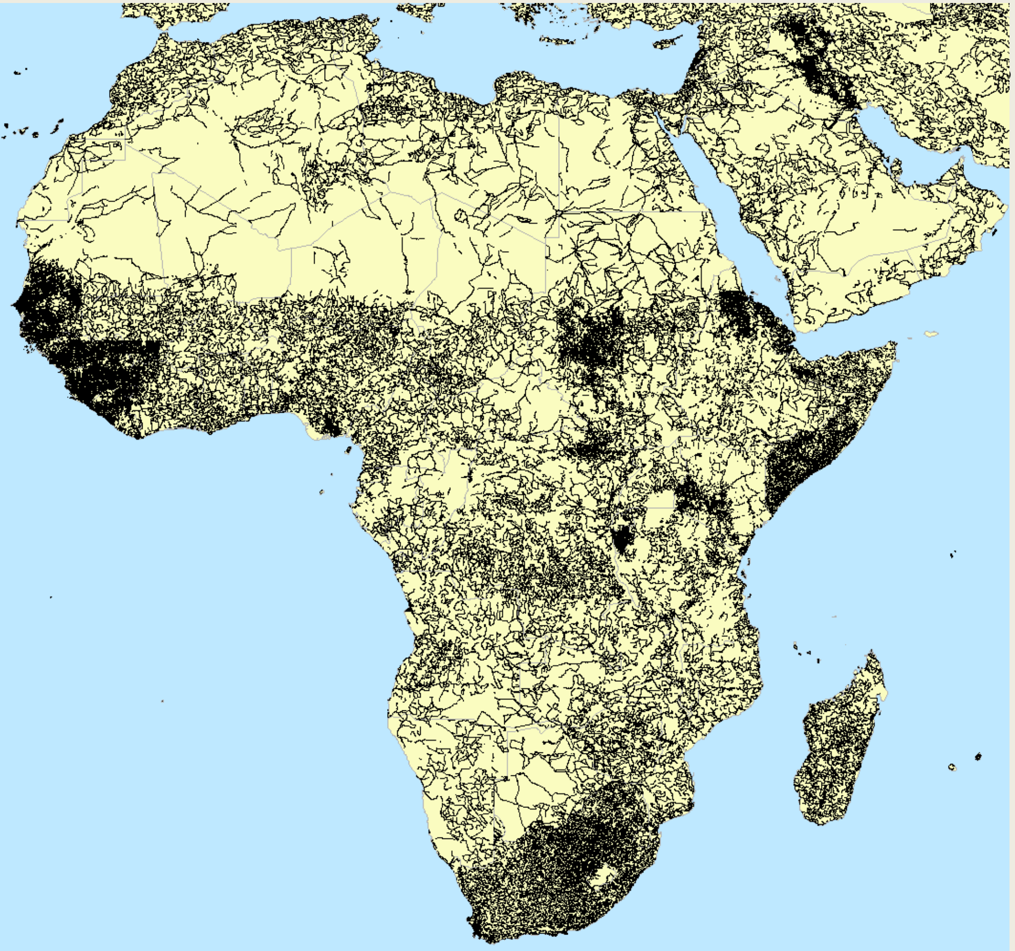
ICSU's Mission → **CODATA's Mission**

"Strengthen international science for the benefit of society ***by promoting improved scientific and technical data management and use.***"

CODATA Structure

- Governed by its General Assembly and Executive Committee
- Committees
- Working Groups
- National Members
- Co-operation with other organizations
- Task Groups
  - Advancing Informatics for Microbiology
  - Anthropometric Data and Engineering
  - Data at Risk
  - Data Citation Standards and Practices
  - Earth and Space Science Data Interoperability
  - Exchangeable Materials Data Representation to Support Scientific Research and Education
  - Fundamental Physical Constants
  - Global Information Commons for Science Initiative
  - Linked Open Data for Global Disaster Risk Research
  - Octopus: Mining Space and Terrestrial Data for Improved Weather, Climate and Agriculture Predictions
  - Global Roads Data Development
  - Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries (PASTD)

Global Roads Open Access Data Set (gROADS)



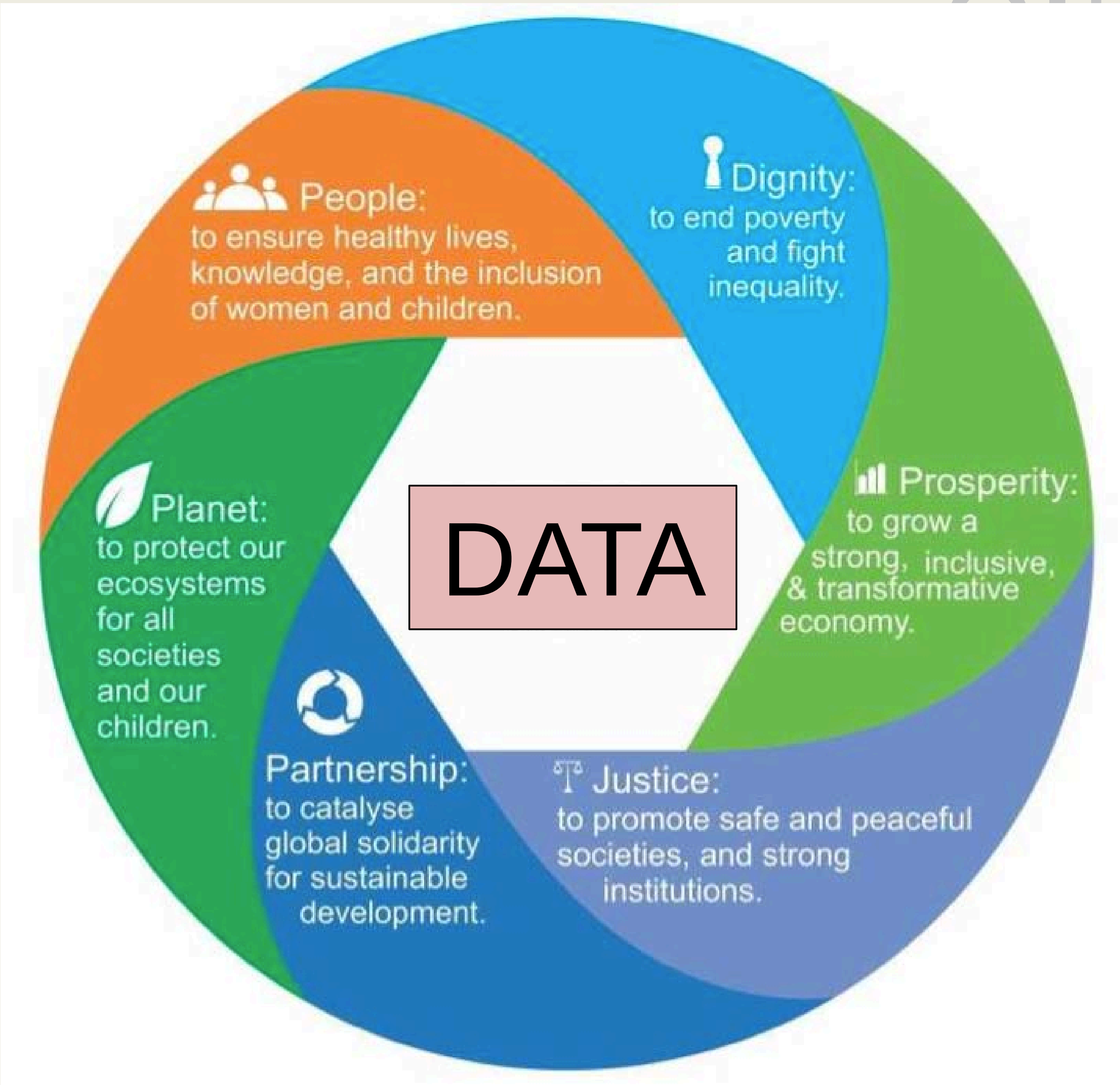
Goal: to develop a global data set (gROADS) that is:

1. spatially accurate (~50m positional accuracy)
2. focused on roads between settlements (not streets)
3. up-to-date and with the possibility of frequent updates
4. well documented
5. freely distributed (on attribution only basis)

2. WHAT ARE THE SDGs AND WHY ARE THEY IMPORTANT?

In September of this year, 2015, the world will agree upon a new global development framework. The Sustainable Development Goals (or SDGs) will succeed the Millennium Development Goals, forged in the year 2000. As agreed by the Open Working Group on the SDGs, the sustainable development goals will provide a holistic framework, applicable to all countries. Taken together, the goals will aim to eradicate poverty and deprivation, but also to grow our economies, to protect our environment and promote peace and good governance.

The SDGs represent a comprehensive effort to set targets for, and to monitor, the planet's social and environmental systems



DATA

Dignity: to end poverty and fight inequality.

People: to ensure healthy lives, knowledge, and the inclusion of women and children.

Prosperity: to grow a strong, inclusive, & transformative economy.

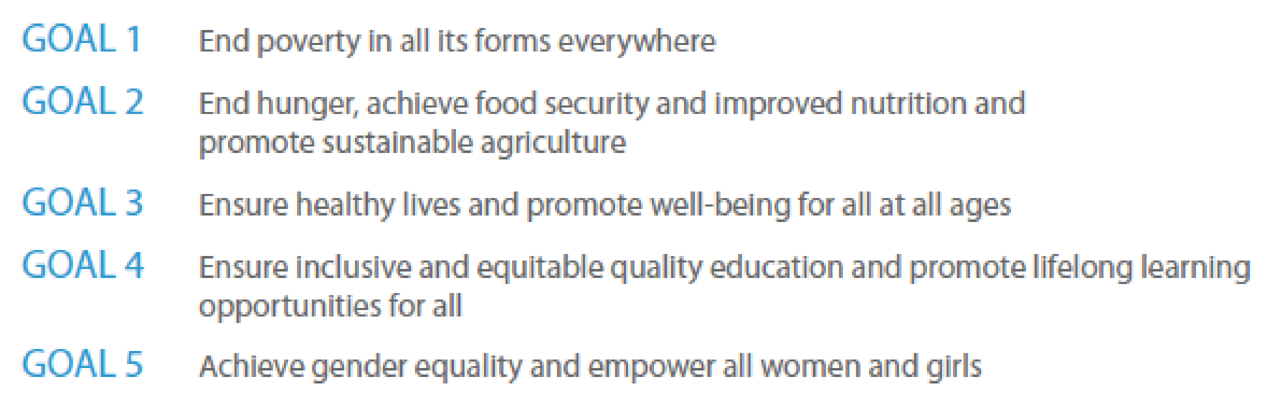
Planet: to protect our ecosystems for all societies and our children.

Partnership: to catalyse global solidarity for sustainable development.

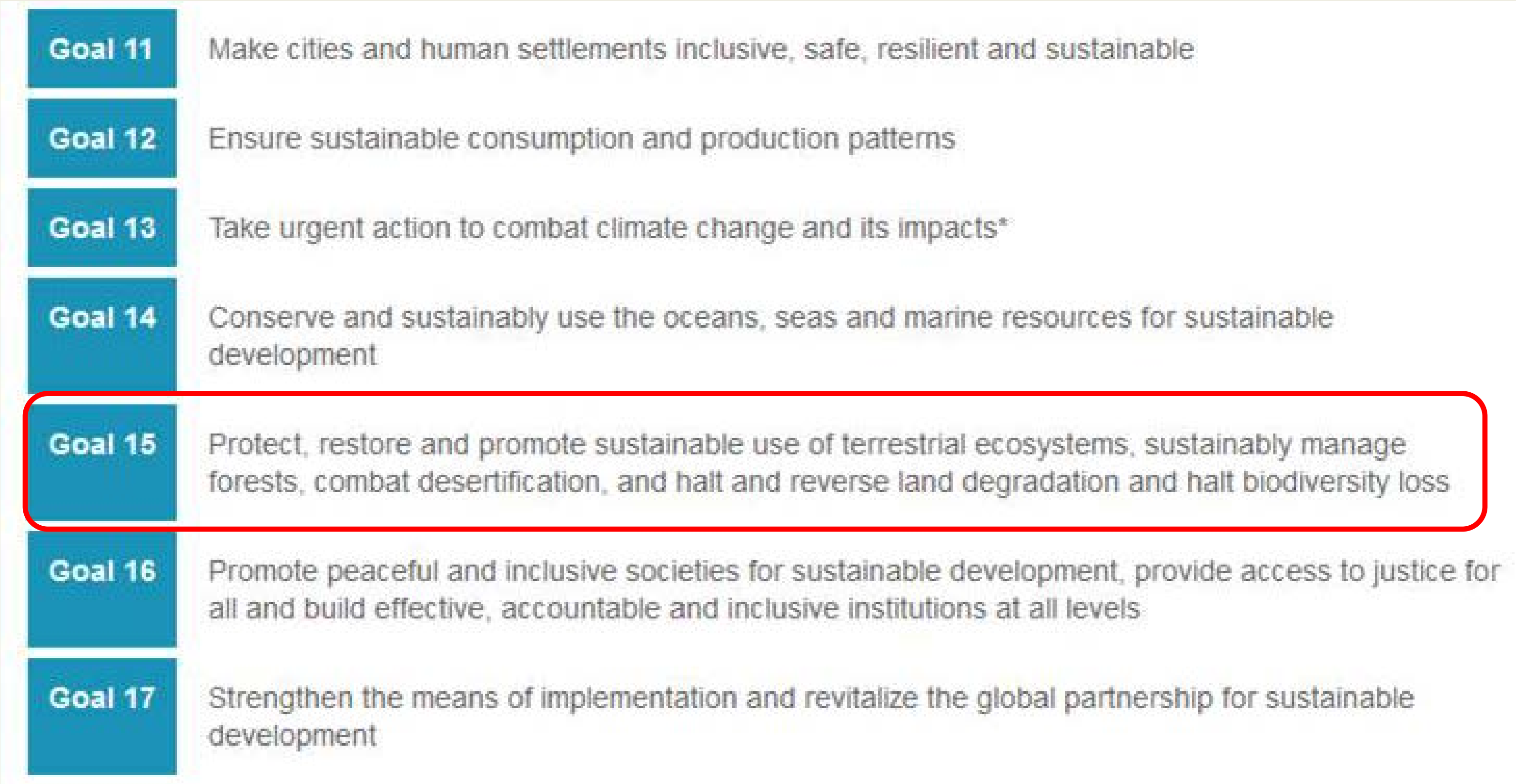
Justice: to promote safe and peaceful societies, and strong institutions.

Source: UN Secretary General, *Road to Dignity* synthesis report

Sample SDGs



GOAL 1	End poverty in all its forms everywhere
GOAL 2	End hunger, achieve food security and improved nutrition and promote sustainable agriculture
GOAL 3	Ensure healthy lives and promote well-being for all at all ages
GOAL 4	Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all
GOAL 5	Achieve gender equality and empower all women and girls

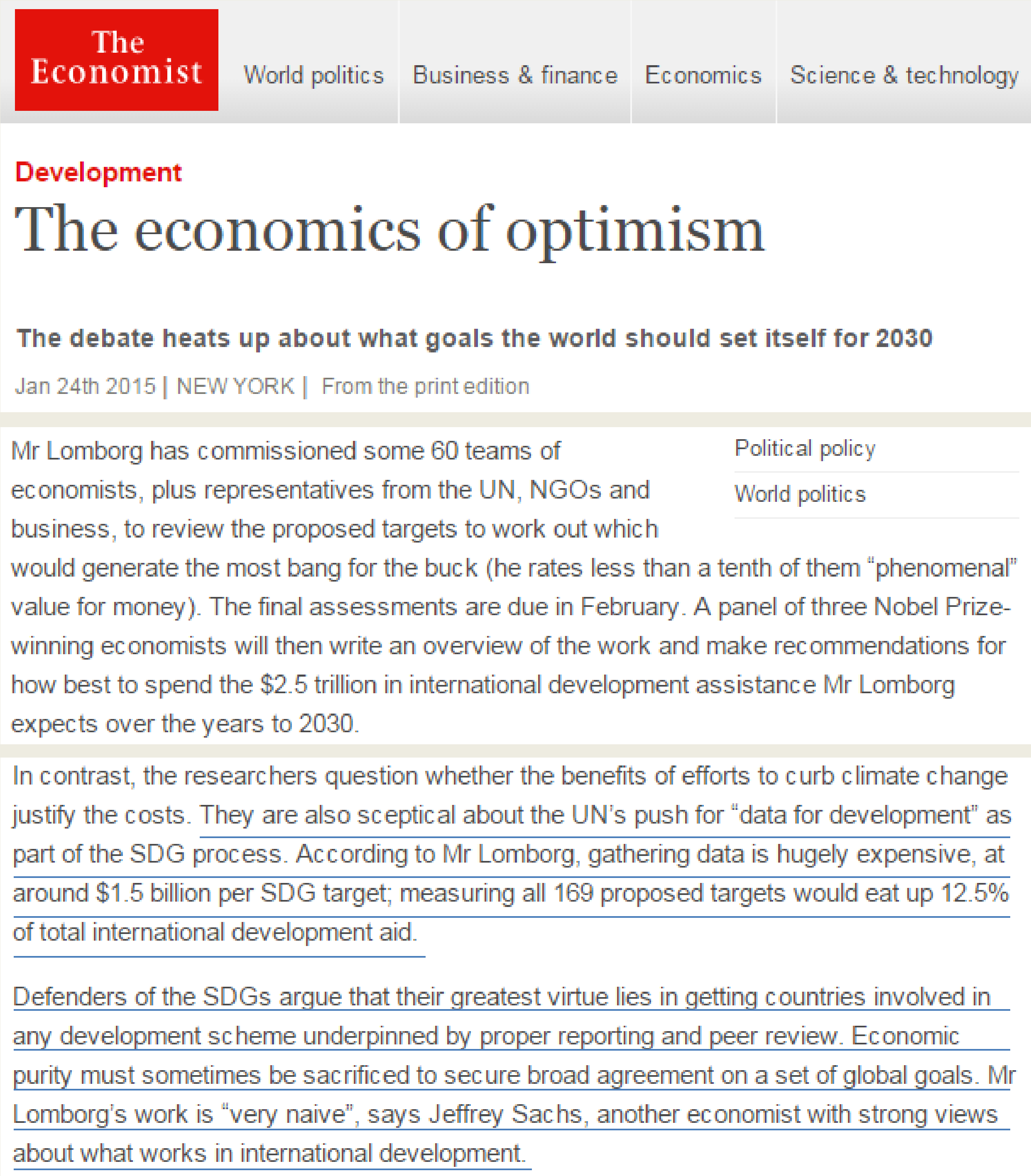


Goal 11	Make cities and human settlements inclusive, safe, resilient and sustainable
Goal 12	Ensure sustainable consumption and production patterns
Goal 13	Take urgent action to combat climate change and its impacts*
Goal 14	Conserve and sustainably use the oceans, seas and marine resources for sustainable development
Goal 15	Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss
Goal 16	Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels
Goal 17	Strengthen the means of implementation and revitalize the global partnership for sustainable development

Goal 15 Proposed Indicators

- Indicator 84: Annual change in forest area and land under cultivation
- Indicator 85: Area of forest under sustainable forest management as a percentage of forest area
- Indicator 86: Red List Index
- Indicator 87: Protected areas overlay with biodiversity

> “Common statistical tools include: census, household and facility surveys, civil registration and vital statistics, environmental statistics, administrative data, financial and economic statistics…. [There] should be a set of common principles for the production, dissemination and use of development data.”
>
> - SDSN, 2014. “A Needs Assessment for SDG Monitoring”



Development

The economics of optimism

The debate heats up about what goals the world should set itself for 2030

Jan 24th 2015 | NEW YORK | From the print edition

Mr Lomborg has commissioned some 60 teams of economists, plus representatives from the UN, NGOs and business, to review the proposed targets to work out which would generate the most bang for the buck (he rates less than a tenth of them "phenomenal" value for money). The final assessments are due in February. A panel of three Nobel Prize-winning economists will then write an overview of the work and make recommendations for how best to spend the $2.5 trillion in international development assistance Mr Lomborg expects over the years to 2030.

In contrast, the researchers question whether the benefits of efforts to curb climate change justify the costs. They are also sceptical about the UN's push for "data for development" as part of the SDG process. According to Mr Lomborg, gathering data is hugely expensive, at around $1.5 billion per SDG target; measuring all 169 proposed targets would eat up 12.5% of total international development aid.

Defenders of the SDGs argue that their greatest virtue lies in getting countries involved in any development scheme underpinned by proper reporting and peer review. Economic purity must sometimes be sacrificed to secure broad agreement on a set of global goals. Mr Lomborg's work is "very naive", says Jeffrey Sachs, another economist with strong views about what works in international development.

SDSN's needs assessment project suggests that what is required to lay the foundations of basic statistical systems is more like $600 million per annum, which is less that 0.5% of ODA

3. WHY ARE DATA AND MONITORING IMPORTANT TO THE SDGs?

Why are data important to the SDGs?

1. Data as a management tool
   - Data and indicators:
     - **D**escribe what’s going on, reducing complexity in policy relevant ways
     - **D**iagnose problems and **D**iscover patterns you didn’t know were there
     - Aid **D**eliberation and promote **D**ebates in society
     - Guide **D**ecision-making
     - **D**rive action and evaluation of policy effects
2. Data for democracy
   - Increase transparency
     - Government accountability
     - Private sector accountability
   - Citizen engagement

Reporting frequency: average time gap in reporting

Many critical data are collected infrequently or not at all

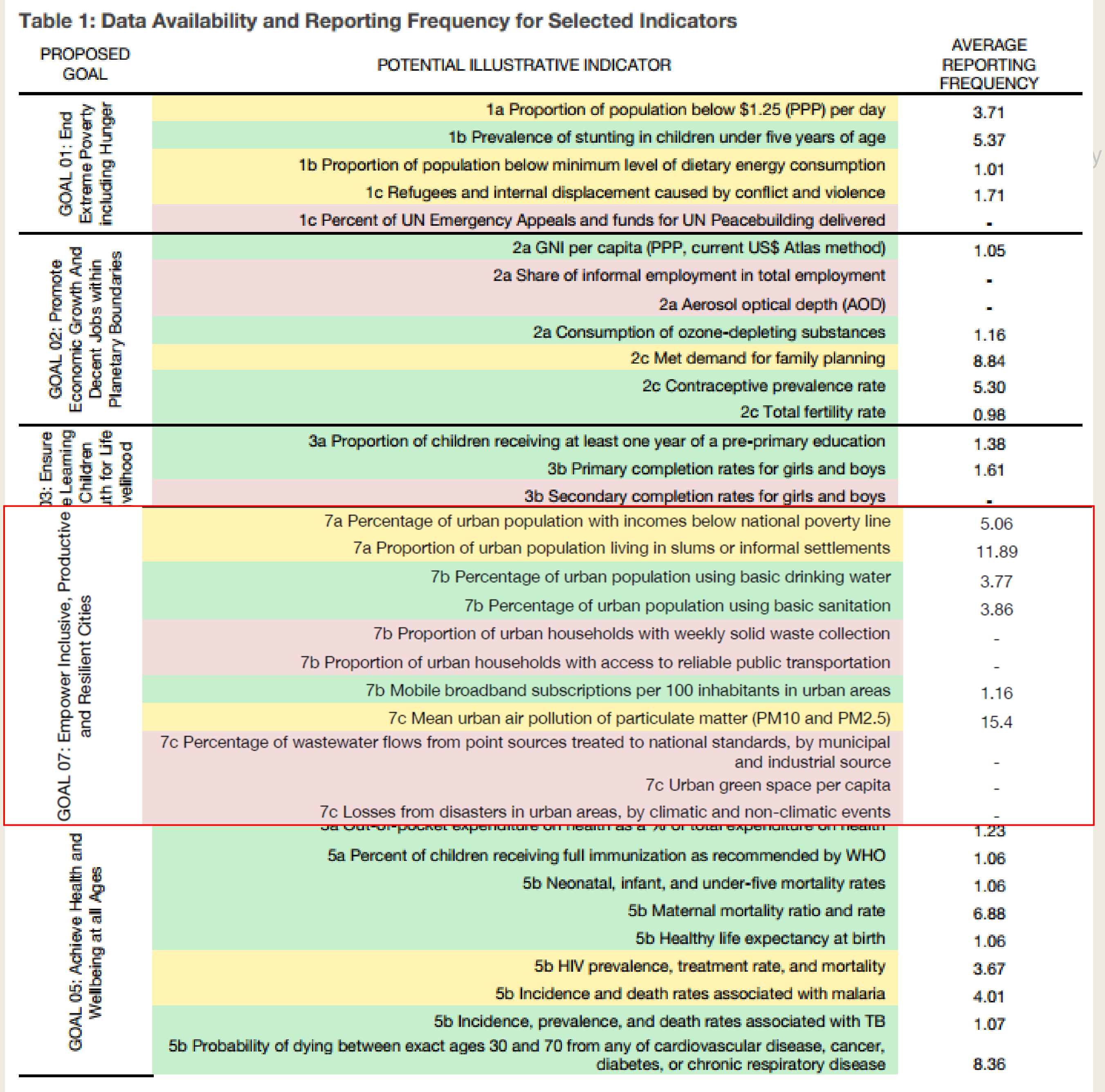


Table 1: Data Availability and Reporting Frequency for Selected Indicators

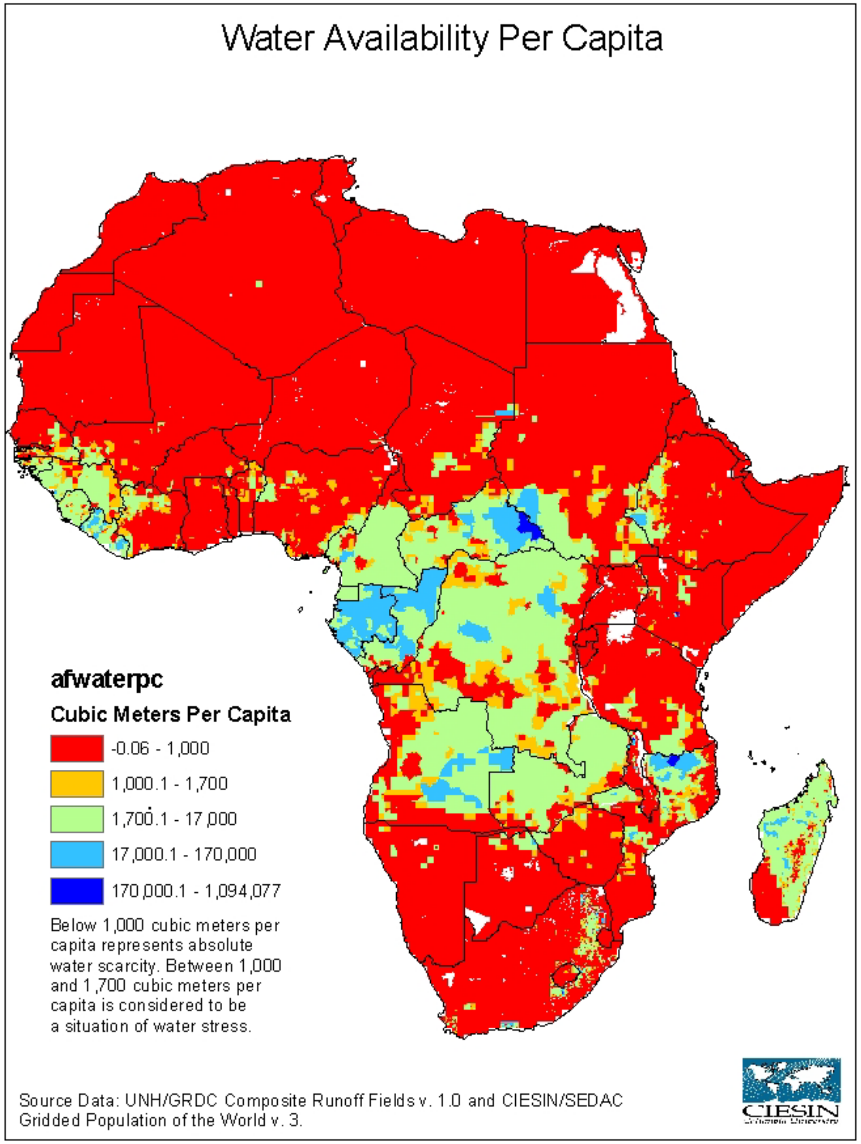
PROPOSED GOAL	POTENTIAL ILLUSTRATIVE INDICATOR	AVERAGE REPORTING FREQUENCY
GOAL 01: End Extreme Poverty including Hunger	1a Proportion of population below $1.25 (PPP) per day	3.71
	1b Prevalence of stunting in children under five years of age	5.37
	1b Proportion of population below minimum level of dietary energy consumption	1.01
	1c Refugees and internal displacement caused by conflict and violence	1.71
	1c Percent of UN Emergency Appeals and funds for UN Peacebuilding delivered	-
GOAL 02: Promote Economic Growth And Decent Jobs within Planetary Boundaries	2a GNI per capita (PPP, current US$ Atlas method)	1.05
	2a Share of informal employment in total employment	-
	2a Aerosol optical depth (AOD)	-
	2a Consumption of ozone-depleting substances	1.16
	2c Met demand for family planning	8.84
	2c Contraceptive prevalence rate	5.30
	2c Total fertility rate	0.98
03: Ensure e Learning Children th for Life velihood	3a Proportion of children receiving at least one year of a pre-primary education	1.38
	3b Primary completion rates for girls and boys	1.61
	3b Secondary completion rates for girls and boys	-
GOAL 07: Empower Inclusive, Productive and Resilient Cities	7a Percentage of urban population with incomes below national poverty line	5.06
	7a Proportion of urban population living in slums or informal settlements	11.89
	7b Percentage of urban population using basic drinking water	3.77
	7b Percentage of urban population using basic sanitation	3.86
	7b Proportion of urban households with weekly solid waste collection	-
	7b Proportion of urban households with access to reliable public transportation	-
	7b Mobile broadband subscriptions per 100 inhabitants in urban areas	1.16
	7c Mean urban air pollution of particulate matter (PM10 and PM2.5)	15.4
	7c Percentage of wastewater flows from point sources treated to national standards, by municipal and industrial source	-
	7c Urban green space per capita	-
	7c Losses from disasters in urban areas, by climatic and non-climatic events	-
GOAL 05: Achieve Health and Wellbeing at all Ages	5a Out-of-pocket expenditure on health as a % of total expenditure on health	1.23
	5a Percent of children receiving full immunization as recommended by WHO	1.06
	5b Neonatal, infant, and under-five mortality rates	1.06
	5b Maternal mortality ratio and rate	6.88
	5b Healthy life expectancy at birth	1.06
	5b HIV prevalence, treatment rate, and mortality	3.67
	5b Incidence and death rates associated with malaria	4.01
	5b Incidence, prevalence, and death rates associated with TB	1.07
	5b Probability of dying between exact ages 30 and 70 from any of cardiovascular disease, cancer, diabetes, or chronic respiratory disease	8.36

Source: SDSN (2014), *Assessing Gaps in Indicator Availability and Coverage*.

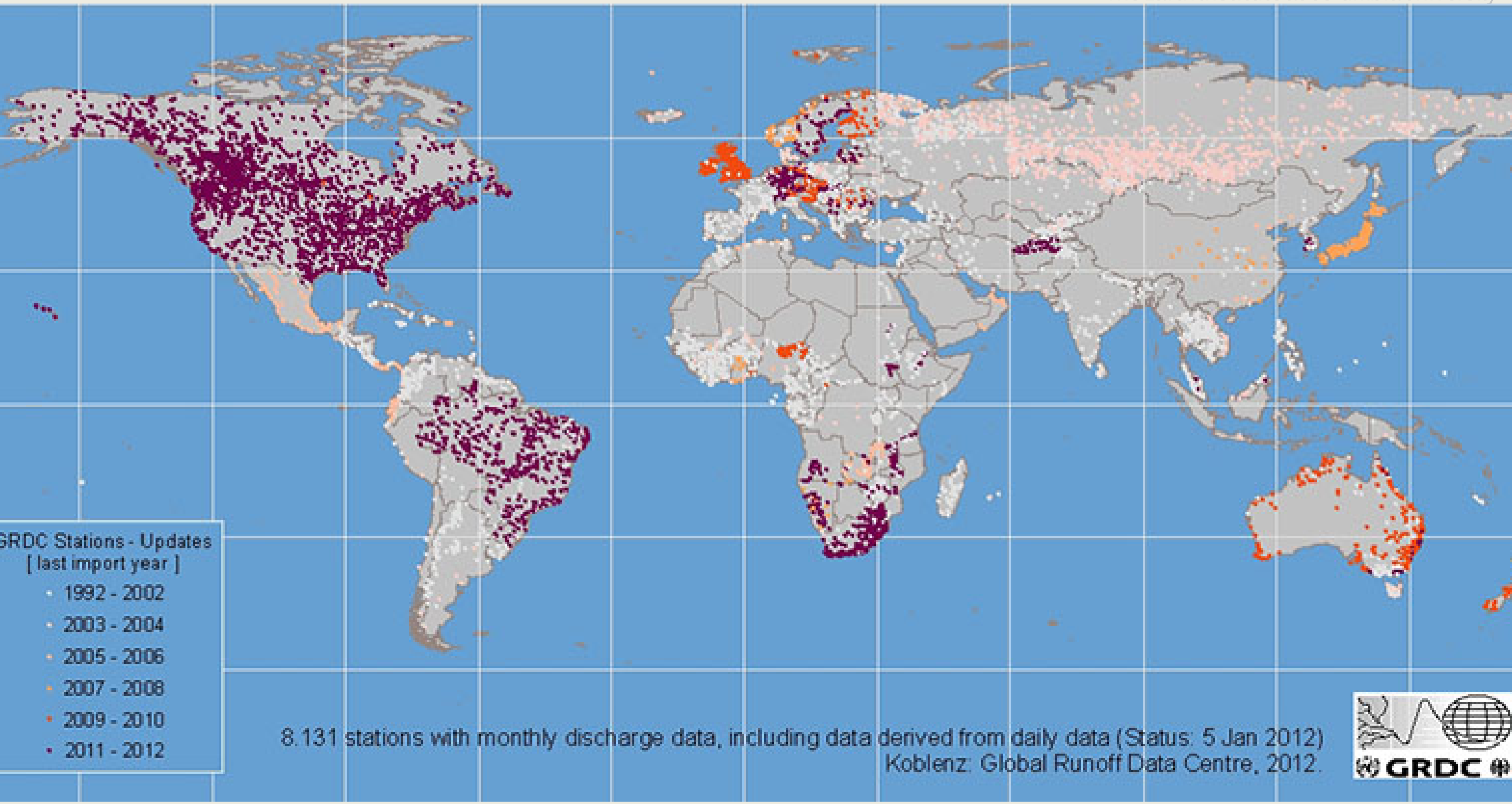
To create this map requires 1,000s of runoff monitoring stations

“The Global Runoff Data Center (GRDC) is **overstressed with respect to its ability to collect unbroken time series of discharge data**, which is the mainstay of GTN-H’s capacity to monitor the state of world water resources. The problem arises from several factors: **the basic logistical challenge of having to contact numerous data holders in any one country** (all with varying commitments to share data); staffing shortages at GRDC; **decline in the number of operating stations**, especially in the developing world; **prohibitions against data release, forced by agency cost-recovery concerns** and national policy based on **strategic concerns**; and, even where the data are available, substantial delays in data processing and release. “

- CIESIN (2008), *Land Degradation Indicator Profiles For the KM:Land Project*


Water Availability Per Capita
afwaterpc
Cubic Meters Per Capita
🟥 -0.06 - 1,000
🟨 1,000.1 - 1,700
🟩 1,700.1 - 17,000
🟦 17,000.1 - 170,000
🟦 170,000.1 - 1,094,077
Below 1,000 cubic meters per capita represents absolute water scarcity. Between 1,000 and 1,700 cubic meters per capita is considered to be a situation of water stress.
CIESIN

Source Data: UNH/GRDC Composite Runoff Fields v. 1.0 and CIESIN/SEDAC Gridded Population of the World v. 3.


GRDC Stations - Updates
[last import year]
1992 - 2002
2003 - 2004
2005 - 2006
2007 - 2008
2009 - 2010
2011 - 2012
8.131 stations with monthly discharge data, including data derived from daily data (Status: 5 Jan 2012)
Koblenz: Global Runoff Data Centre, 2012.
GRDC

Source: Global Runoff Data Center:
http://www.bafg.de/cln_031/nn_266934/GRDC/EN/01__GRDC/03__Database/database__node.html?__nnn=true

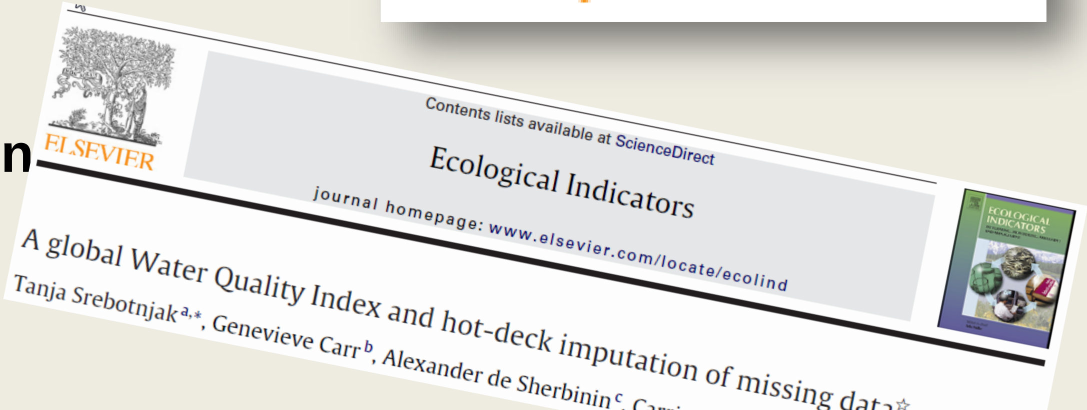
CIESIN
The Earth Institute at Columbia University

The Water Quality Index (WQI)

Environmental Performance Index

Born: 2006
Died: 2010
Parents: UNEP-GEMS, CIESIN, Yale
Country coverage: 85

Cause of Death: Lack of consistent time series data on water quality for a sufficient number of monitoring stations in all countries


ELSEVIER

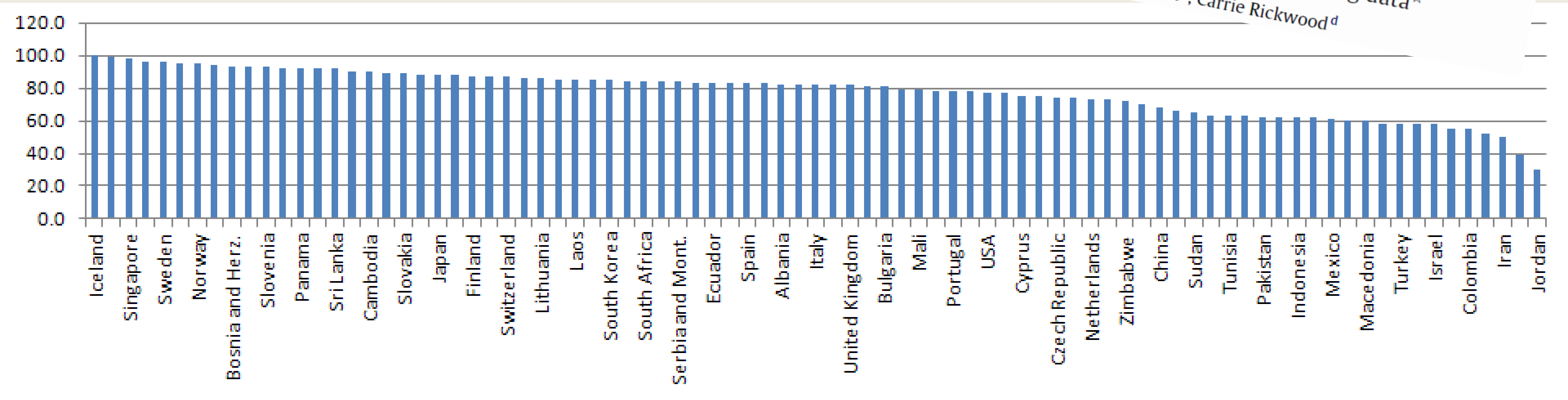
Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

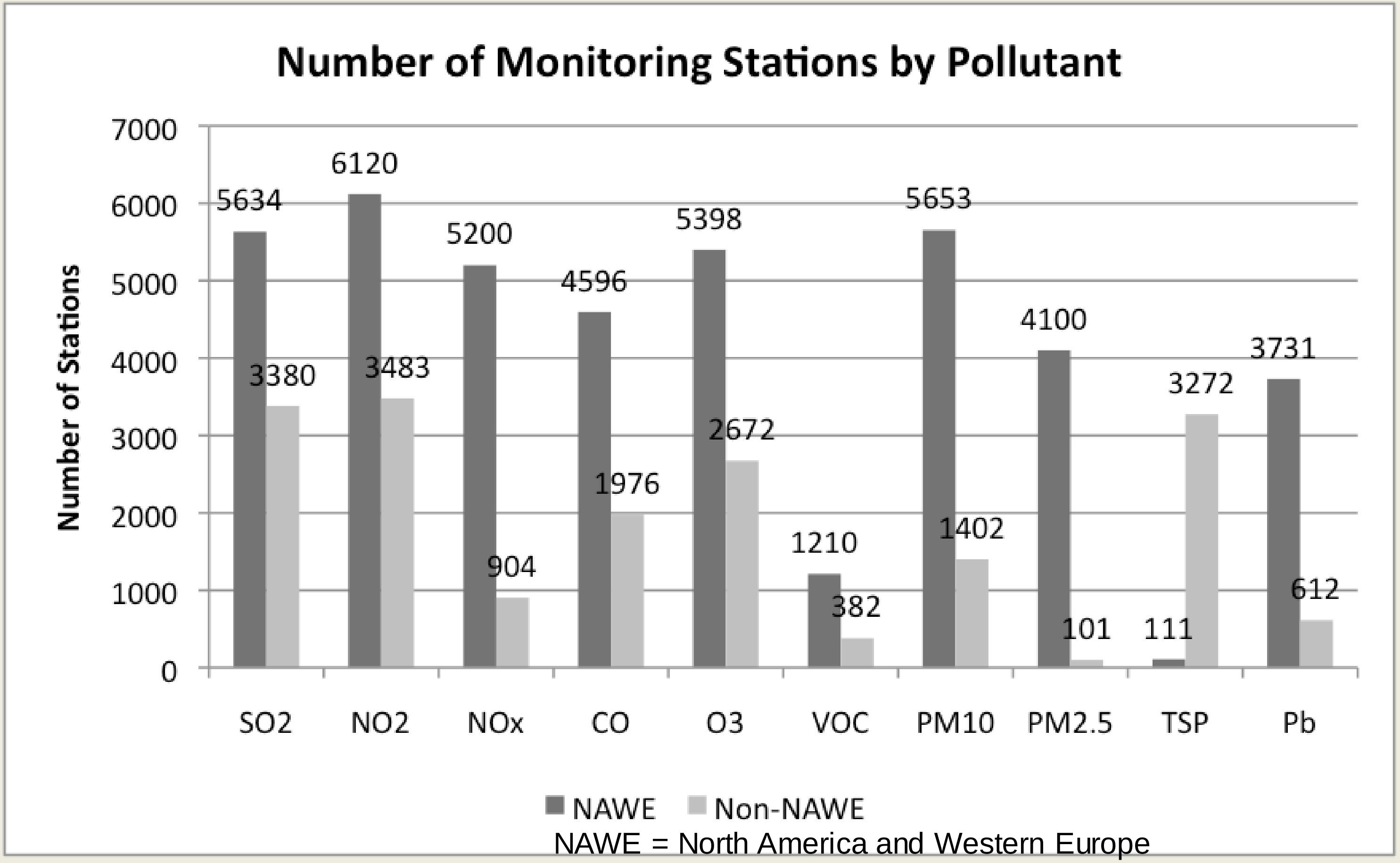
A global Water Quality Index and hot-deck imputation of missing data☆

Tanja Srebotnjak[a,*], Genevieve Carr[b], Alexander de Sherbinin[c], Carrie Rickwood[d]



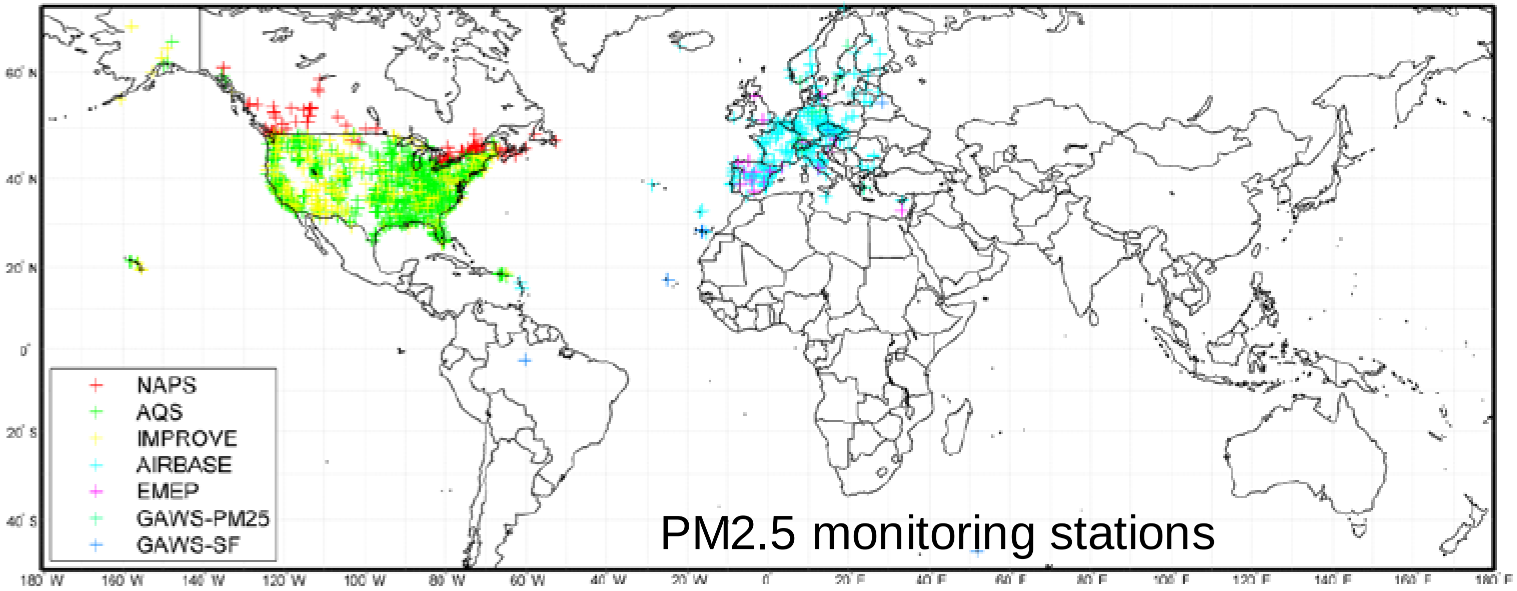
Disparities in ground-level air quality monitoring

CIESIN
The Earth Institute at Columbia University



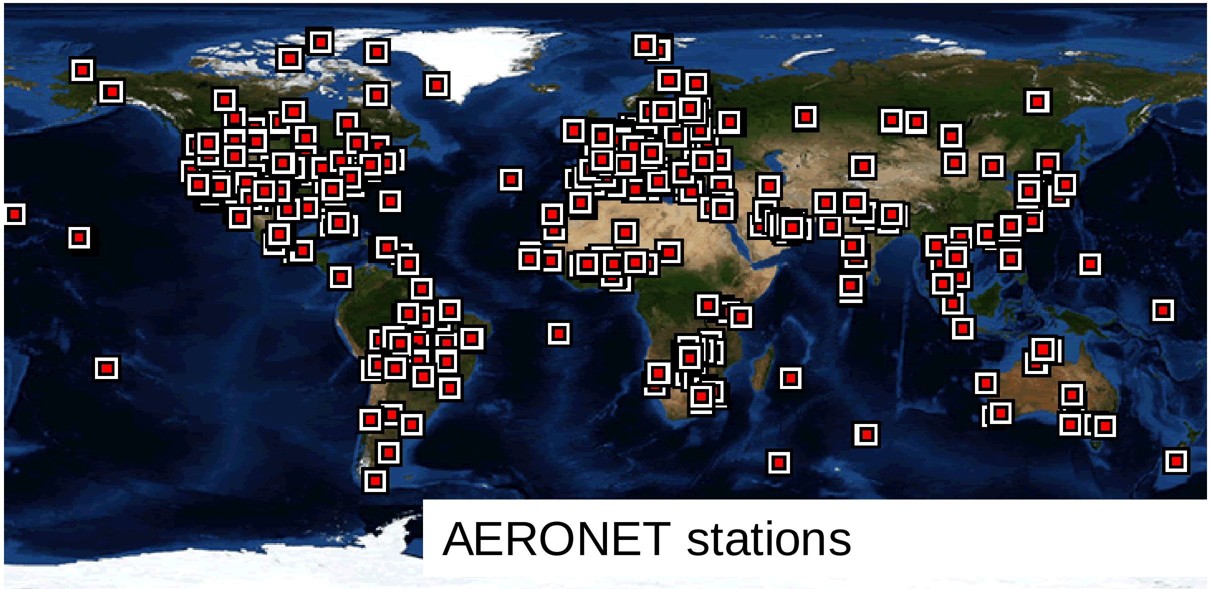
NAWE = North America and Western Europe

Source: Compiled by Rudy Husar (Washington University in St. Louis) and Stefan Falke (Northrop Grumman Corp), Funded by NASA, for GEO Task US-09-01a



PM2.5 monitoring stations

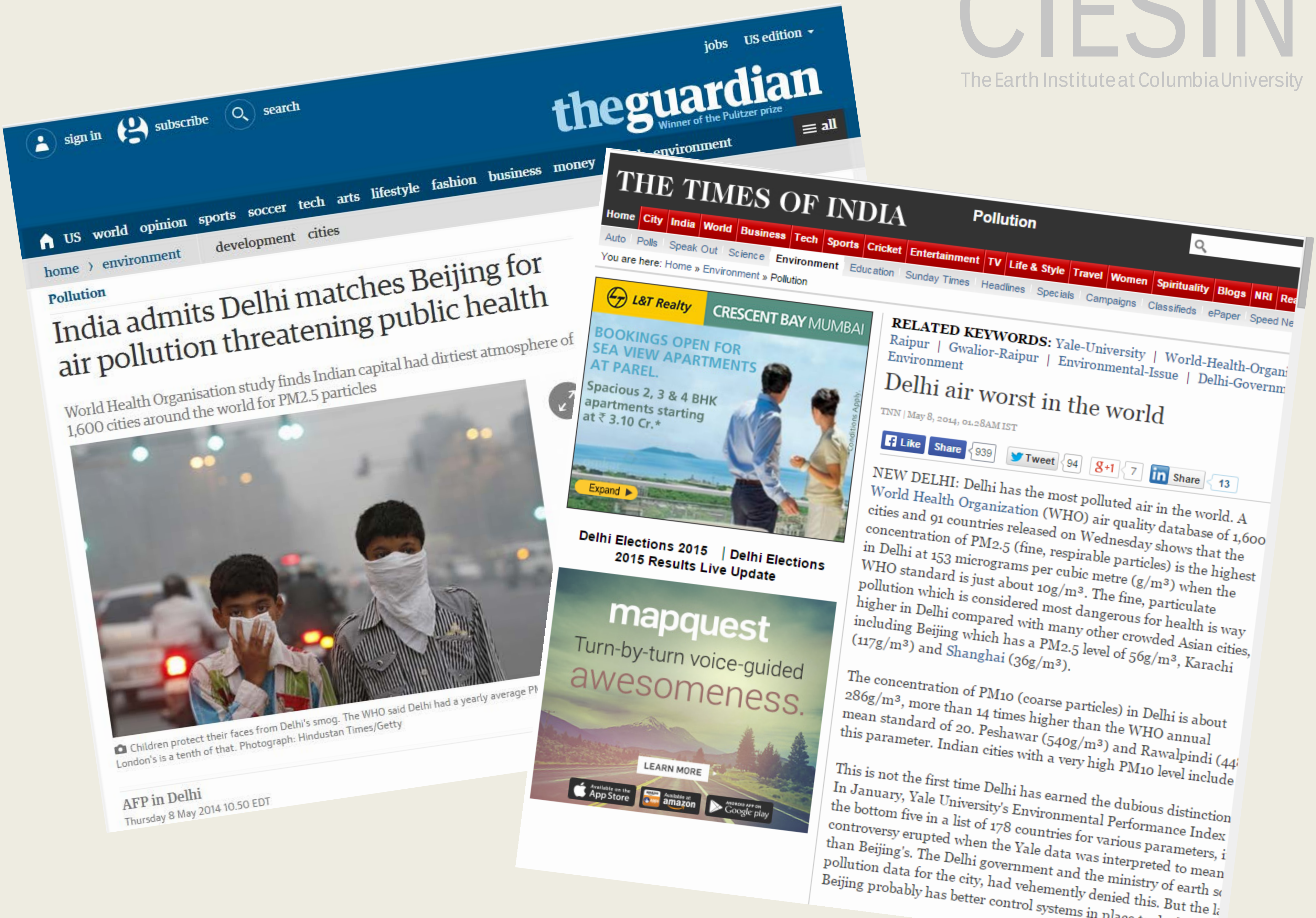
To add new $PM_{2.5}$ monitoring stations costs ~$14-18k per station, and $5k per year to operate.



AERONET stations

CIESIN

The Earth Institute at Columbia University



Pollution

India admits Delhi matches Beijing for air pollution threatening public health

World Health Organisation study finds Indian capital had dirtiest atmosphere of 1,600 cities around the world for PM2.5 particles

Children protect their faces from Delhi's smog. The WHO said Delhi had a yearly average PM London's is a tenth of that. Photograph: Hindustan Times/Getty

AFP in Delhi
Thursday 8 May 2014 10.50 EDT

RELATED KEYWORDS: Yale-University | World-Health-Organi Raipur | Gwalior-Raipur | Environmental-Issue | Delhi-Governm Environment

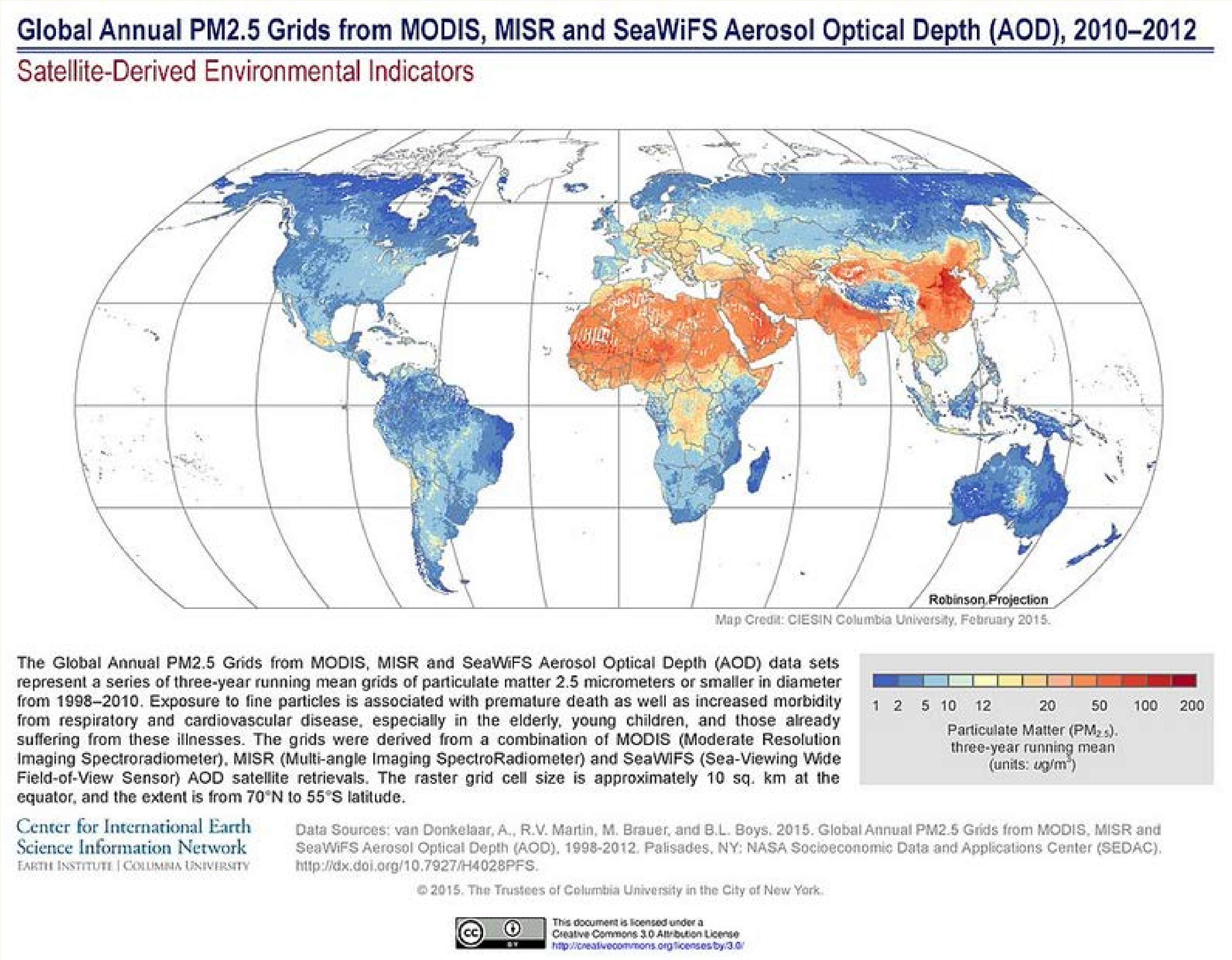
Delhi air worst in the world

TNN | May 8, 2014, 01.28AM IST

NEW DELHI: Delhi has the most polluted air in the world. A World Health Organization (WHO) air quality database of 1,600 cities and 91 countries released on Wednesday shows that the concentration of PM2.5 (fine, respirable particles) is the highest in Delhi at 153 micrograms per cubic metre (g/m³) when the WHO standard is just about 10g/m³. The fine, particulate pollution which is considered most dangerous for health is way higher in Delhi compared with many other crowded Asian cities, including Beijing which has a PM2.5 level of 56g/m³, Karachi (117g/m³) and Shanghai (36g/m³).

The concentration of PM10 (coarse particles) in Delhi is about 286g/m³, more than 14 times higher than the WHO annual mean standard of 20. Peshawar (540g/m³) and Rawalpindi (44 this parameter. Indian cities with a very high PM10 level include

This is not the first time Delhi has earned the dubious distinction In January, Yale University's Environmental Performance Index the bottom five in a list of 178 countries for various parameters, i controversy erupted when the Yale data was interpreted to mean than Beijing's. The Delhi government and the ministry of earth s pollution data for the city, had vehemently denied this. But the l Beijing probably has better control systems in place



Global Annual PM2.5 Grids from MODIS, MISR and SeaWiFS Aerosol Optical Depth (AOD), 2010–2012

Satellite-Derived Environmental Indicators

Map Credit: CIESIN Columbia University, February 2015.

The Global Annual PM2.5 Grids from MODIS, MISR and SeaWiFS Aerosol Optical Depth (AOD) data sets represent a series of three-year running mean grids of particulate matter 2.5 micrometers or smaller in diameter from 1998–2010. Exposure to fine particles is associated with premature death as well as increased morbidity from respiratory and cardiovascular disease, especially in the elderly, young children, and those already suffering from these illnesses. The grids were derived from a combination of MODIS (Moderate Resolution Imaging Spectroradiometer), MISR (Multi-angle Imaging SpectroRadiometer) and SeaWiFS (Sea-Viewing Wide Field-of-View Sensor) AOD satellite retrievals. The raster grid cell size is approximately 10 sq. km at the equator, and the extent is from 70°N to 55°S latitude.

Center for International Earth Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Data Sources: van Donkelaar, A., R.V. Martin, M. Brauer, and B.L. Boys. 2015. Global Annual PM2.5 Grids from MODIS, MISR and SeaWiFS Aerosol Optical Depth (AOD), 1998-2012. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4028PFS.

Remote sensing data require ground validation

- Without ground monitoring systems, we are unable to monitor planetary health
- Satellites and novel streams of “big data” cannot completely replace ground measurement (Part 5)
- We also need to support the custodians who aggregate such data

Data for democracy and open data

- Open data are data that are accessible for free or at negligible cost, and with minimal limitations on its use, transformation, and distribution
- Open data may be machine-readable and an input for machine learning
  - Computer algorithms that have the ability to 'learn' to make better predictions and decisions based on what was experienced in the past (e.g., spam filtering)

> "Big Data and open data movements will be the two main pillars of a larger 'data revolution'. Both rise against a background of increased public demand for more openness, agility, transparency and accountability for public data and actions. The political overtones — so easily forgotten — are clear."
>
> - Letouzé, 2014

Data democracy:
The "data revolution for development"

- More diverse, integrated, timely and trustworthy information can lead to better decision-making and real-time citizen feedback. This in turn enables individuals, public and private institutions, and companies to make choices that are good for them and for the world they live in.
- Too often, existing data remain unused because they are released too late or not at all, not well documented and harmonized, or not available at the level of detail needed for decision-making

Source: Independent Expert Advisory Group (IEAG)

Environmental Development 3 (2012) 39–51

Contents lists available at SciVerse ScienceDirect

Environmental Development

journal homepage: www.elsevier.com/locate/envdev

ELSEVIER

Seeking truth from facts: The challenge of environmental indicator development in China

Angel Hsu[a,*], Alex de Sherbinin[b], Han Shi[c]

[a] Yale School of Forestry and Environmental Studies; 205 Prospect Street, New Haven, CT 06511, United States
[b] Center for International Earth Science Information Network, The Earth Institute, Columbia University. P.O. Box 1000 (61 Route 9W), Palisades, NY 10964, United States
[c] Department of Public and Social Administration, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, China

Towards a China Environmental Performance Index

Yale Center for Environmental Law and Policy, Yale University

Center for International Earth Science Information Network, Columbia University

In collaboration with: Chinese Academy for Environmental Planning

City University of Hong Kong

- China publishes reams of “official” statistics
- Accessing raw data behind the statistics is very difficult
- When the US embassy in Beijing published PM2.5 measures on its Web site it created a diplomatic incident
- Freedom of information type legislation is on the books
- Lack of access to accurate data stifles civil society engagement and promotes apathy

Two Models

Old School

- Information is power
- Information has a price
- Data producers must recover costs from users
- User unable to redistribute value-added products

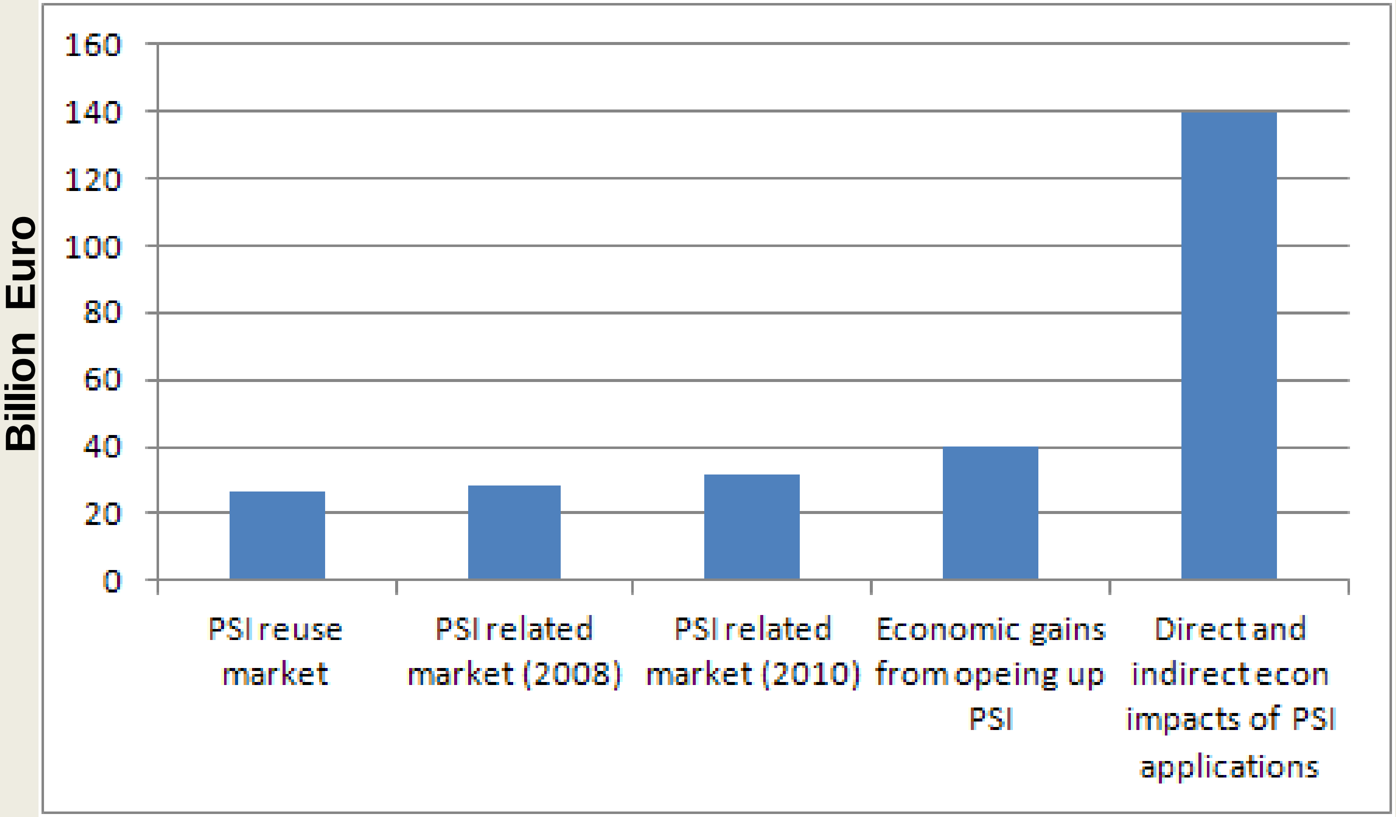
Open Data

- Information is provided free of charge and without copyright restriction
- Society is better informed
- Lower costs to industry
- Information sector is spawned and grows
- Taxes on this sector funds data
- Move from data to services

Investing in Open Data

CIESIN
The Earth Institute at Columbia University

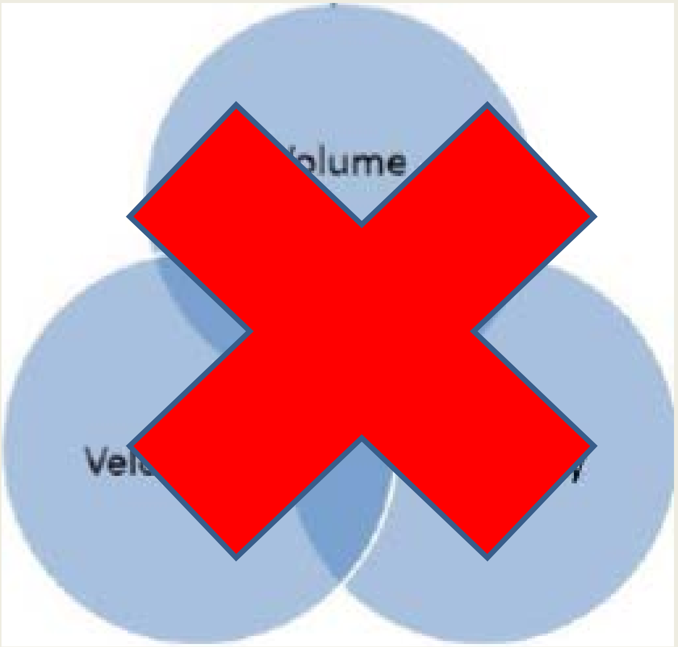
Economic benefits of open public sector information (PSI) in the EU27



Source: Vickery (2011), "Review of Recent Studies on PSI Re-Use and Related Market Developments"

4. CAN “BIG DATA” FILL ALL OUR DATA GAPS?

What are big data?

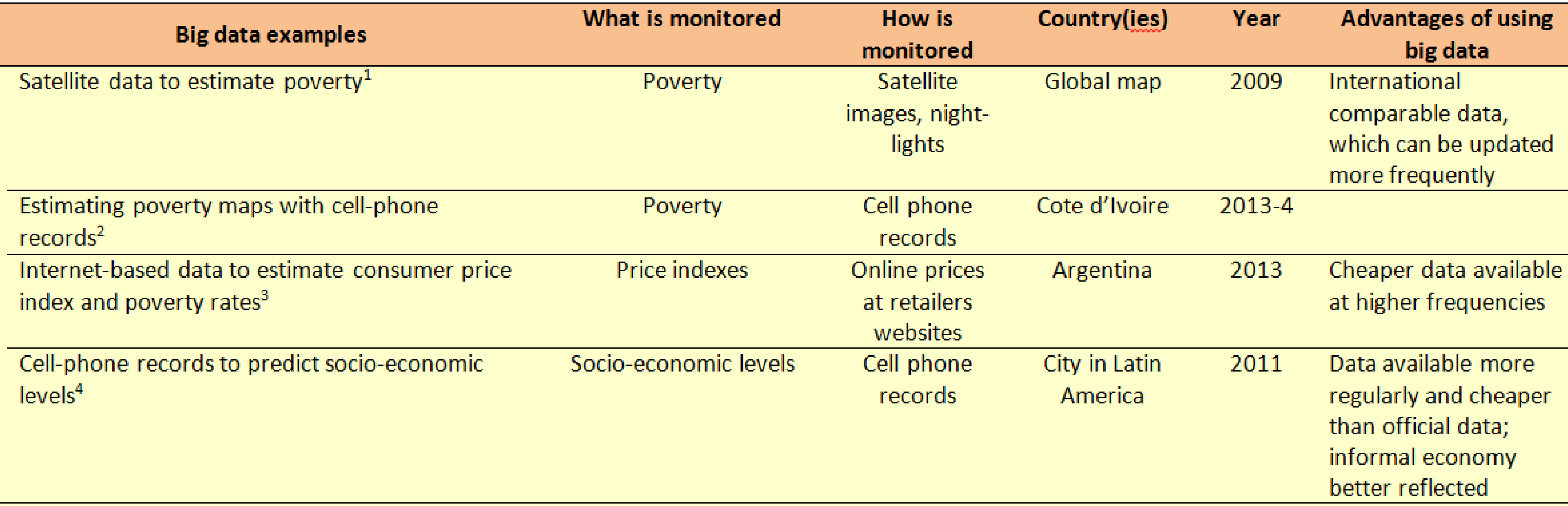


> “Even a small ‘Big Data’ dataset is Big Data because it doesn't stem from fully controlled processes like surveys and statistical imputations undertaken by official bodies”
>
> - Letouzé, 2014

- Crumbs, Capacities and Communities
  - Digital **Crumbs**: generated about and by people, often as the by-product of their use of digital devices
    - Call detail records, transaction data
    - Social media
    - Remotely sensed or citizen sensed
    - Crowd sourced data
  - **Capacities**: powerful computers and machine-learning techniques are able to look for and unveil patterns and trends in vast amounts of complex data
  - **Communities**: A ‘movement’ of individual and institutional actors that operate largely outside of traditional policy and research spheres working to turn data into decision making (e.g., hack-a-thons, app competitions)

Big Data and the SDGs

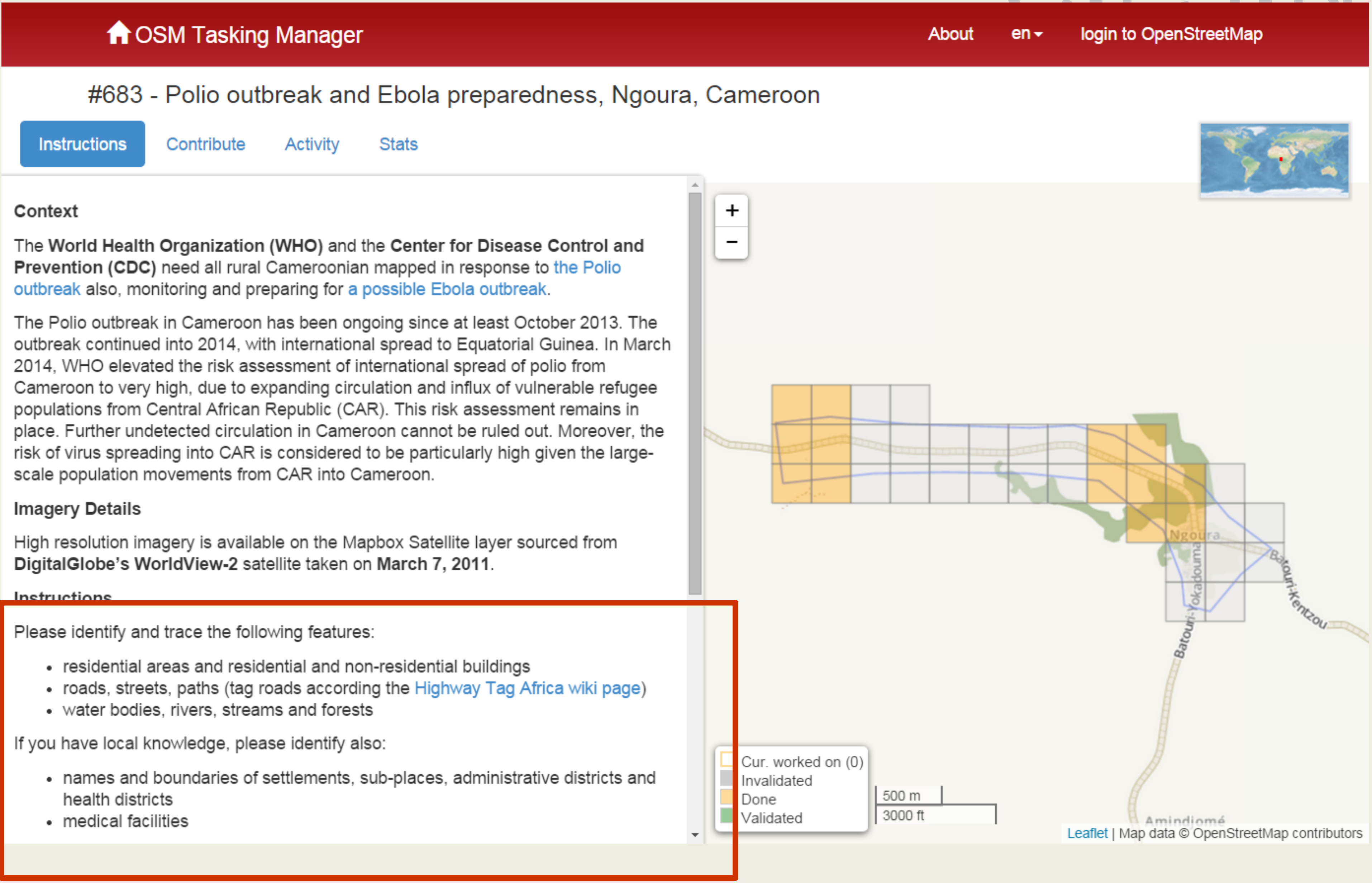
SDG 1: Poverty Eradication



Big data examples	What is monitored	How is monitored	Country(ies)	Year	Advantages of using big data
Satellite data to estimate poverty[1]	Poverty	Satellite images, night-lights	Global map	2009	International comparable data, which can be updated more frequently
Estimating poverty maps with cell-phone records[2]	Poverty	Cell phone records	Cote d'Ivoire	2013-4	
Internet-based data to estimate consumer price index and poverty rates[3]	Price indexes	Online prices at retailers websites	Argentina	2013	Cheaper data available at higher frequencies
Cell-phone records to predict socio-economic levels[4]	Socio-economic levels	Cell phone records	City in Latin America	2011	Data available more regularly and cheaper than official data; informal economy better reflected

Crowd Sourcing Roads / Settlements / Infrastructure Data

- Applications
  - Humanitarian response
  - Development planning
  - Community development
- Data and methods
  - Members of the public use high resolution remote sensing imagery to identify objects using an online tool, or collect data with GPS units
- Pros
  - Low costs of production because it is decentralized
  - Possible to types of residential areas (e.g. slums, high income)
  - With some simple modeling it is possible to estimate the number of people per dwelling and possibly their wealth or other characteristics
- Cons
  - The “crowd” may only be motivated at the time of a crisis (e.g., post 2010 earthquake Haiti, current Ebola outbreak)
  - Large areas go unmapped until a crisis erupts
  - Reticence by traditional bureaucracies to use unvalidated data



#683 - Polio outbreak and Ebola preparedness, Ngoura, Cameroon

Context

The **World Health Organization (WHO)** and the **Center for Disease Control and Prevention (CDC)** need all rural Cameroonian mapped in response to the Polio outbreak also, monitoring and preparing for a possible Ebola outbreak.

The Polio outbreak in Cameroon has been ongoing since at least October 2013. The outbreak continued into 2014, with international spread to Equatorial Guinea. In March 2014, WHO elevated the risk assessment of international spread of polio from Cameroon to very high, due to expanding circulation and influx of vulnerable refugee populations from Central African Republic (CAR). This risk assessment remains in place. Further undetected circulation in Cameroon cannot be ruled out. Moreover, the risk of virus spreading into CAR is considered to be particularly high given the large-scale population movements from CAR into Cameroon.

Imagery Details

High resolution imagery is available on the Mapbox Satellite layer sourced from **DigitalGlobe's WorldView-2** satellite taken on **March 7, 2011**.

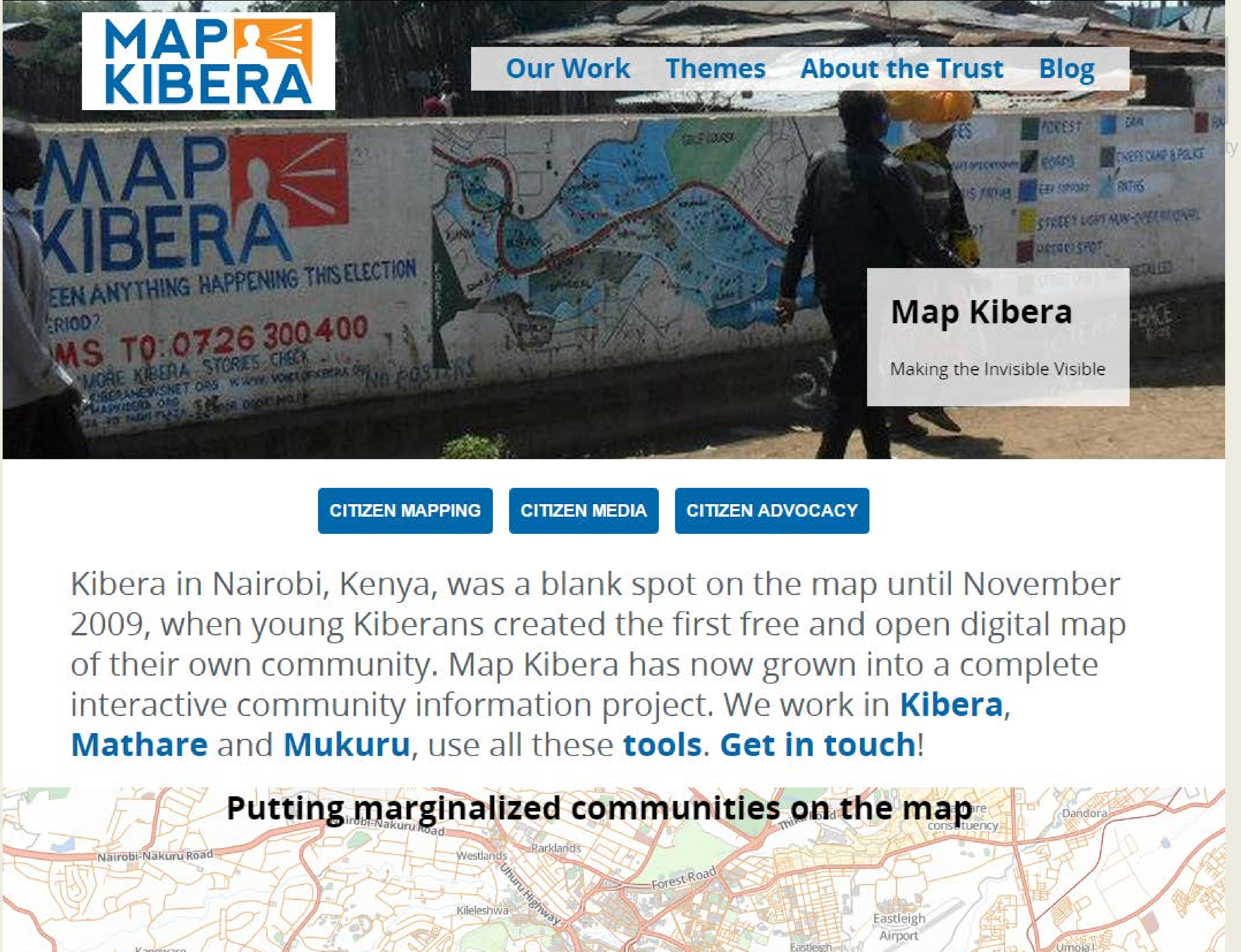
Instructions

Please identify and trace the following features:

- residential areas and residential and non-residential buildings
- roads, streets, paths (tag roads according the Highway Tag Africa wiki page)
- water bodies, rivers, streams and forests

If you have local knowledge, please identify also:

- names and boundaries of settlements, sub-places, administrative districts and health districts
- medical facilities



Map Kibera

Making the Invisible Visible

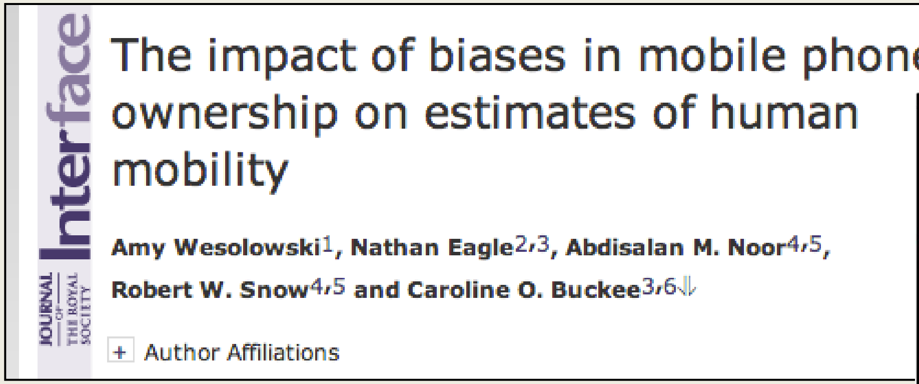
Kibera in Nairobi, Kenya, was a blank spot on the map until November 2009, when young Kiberans created the first free and open digital map of their own community. Map Kibera has now grown into a complete interactive community information project. We work in **Kibera**, **Mathare** and **Mukuru**, use all these **tools**. **Get in touch**!

Putting marginalized communities on the map

Cell Phone Call Data Records (CDRs)

- Measured phenomena
  - NRT population movements (daily commutes, disaster displacement)
  - Proxy for ambient population (day time, night time, etc.) for health/ hazard exposure studies
- Data and methods
  - Call data records come in many different formats depending on the carrier
- Pros
  - Provides very detailed data on population movements in/to urban areas
  - Data are abundant, timely, and require no dedicated survey or collection effort
- Cons
  - In rural areas with sparse arrays of cell towers, it is much harder to localize people
  - Fragmentation of telecom markets limits work to single countries
  - Proprietary data, unique data formats = data cleaning is labor intensive
  - Selectivity issues: making population-level inferences is complicated by differential ownership of phones among different demographic groups

Mobile Data for Socioeconomic/Poverty Mapping

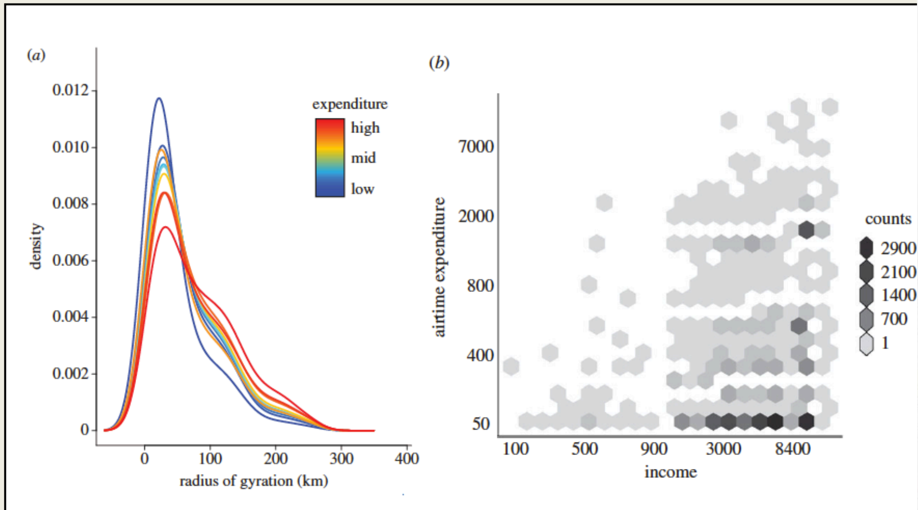


Interface JOURNAL OF THE ROYAL SOCIETY

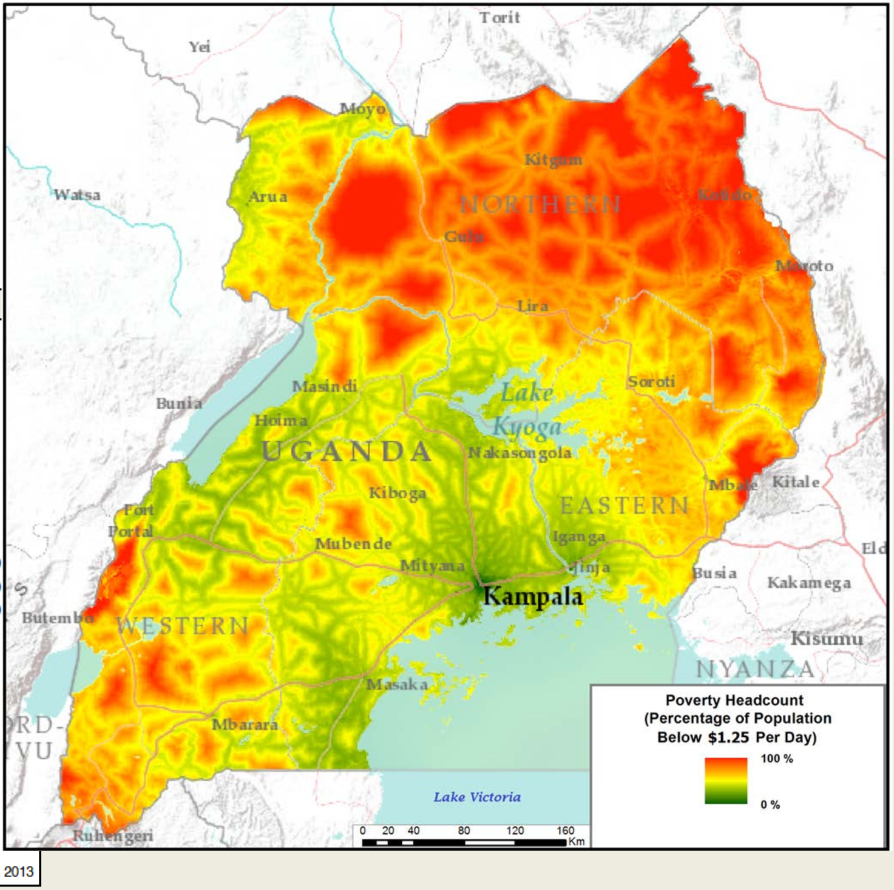
The impact of biases in mobile phone ownership on estimates of human mobility

Amy Wesolowski[1], Nathan Eagle[2,3], Abdisalan M. Noor[4,5], Robert W. Snow[4,5] and Caroline O. Buckee[3,6]

+ Author Affiliations



Wesolowski, A. Journal of the Royal Society, Interface / the Royal Society. 2013

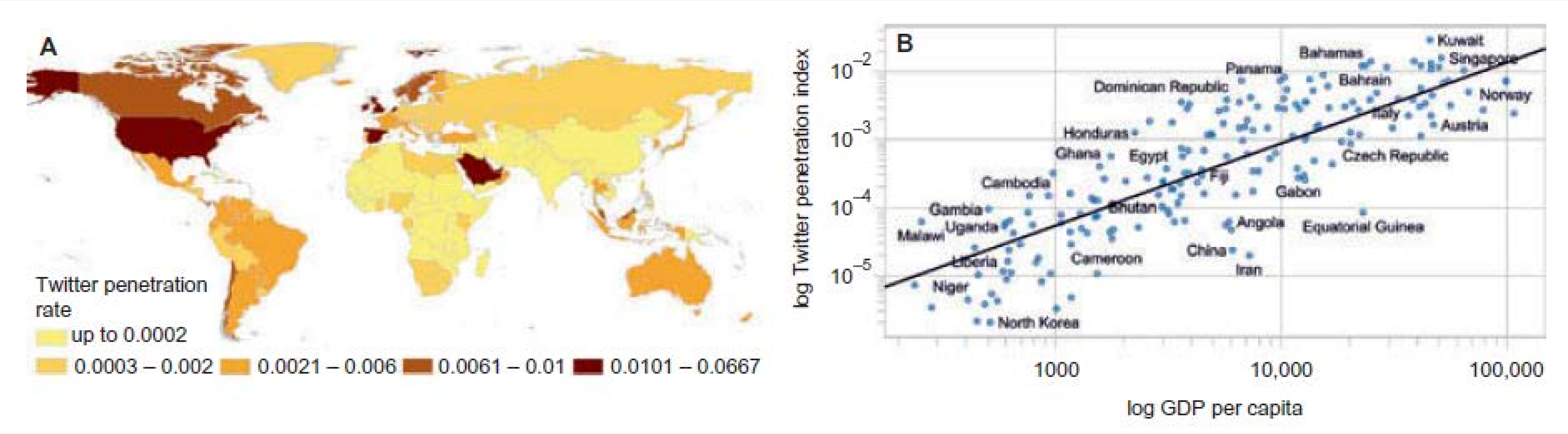


Poverty Headcount
(Percentage of Population
Below $1.25 Per Day)

Slide courtesy Erik Wetter, FLOWMINDER.ORG

Social Media (Twitter, SMS, Instagram, Facebook)

- Application areas
  - Assess short-term population movements (tourist flows, population displacements, humanitarian crises)
  - Health outbreaks or trends
  - Anti-terrorism monitoring
- Data and methods
  - Mapping of geo-located tweets
  - Location references in SMS (e.g. Ushahidi)
- Pros
  - Timeliness: rapid data development in humanitarian crises
  - SMS is available to a broad demographic
  - Twitter used in almost all countries unlike telecoms / CDRs
  - Geo-located tweets 1% of total feed *but* growing rapidly
- Cons
  - Possible biases in social media users (age, gender, income, location)
  - People displaced by conflict and in remote areas unlikely to text or tweet



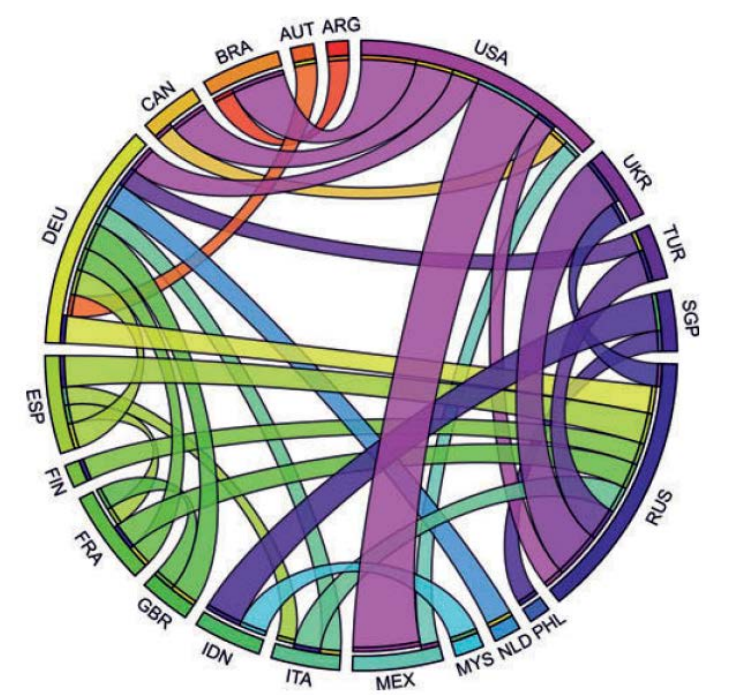


Figure 9. Top 30 country-to-country estimated flows of visitors. Colors of the ribbons correspond to the destination of a trip; the country of origin is marked with a thin stripe at the end of a ribbon (visualization method based on Krzywinski et al. 2009).

Source: Hawelka et al. 2014. Geo-located Twitter as a proxy for global mobility patterns. *Cartography & Geog. Info. Science.* 41(3):260

"The problems we identify are not limited to [Google Flu Trends]. Research on whether search or social media can predict *x* has become commonplace and is often put in sharp contrast with traditional methods and hypotheses. Although these studies have shown the value of these data, we are far from a place where they can supplant more traditional methods or theories."

- Lazar et al., 2014. "The Parable of Google Flu: Traps in Big Data Analysis" *Science.*

"A few generations ago, people grew up in and were comfortable with big organizations – the army, corporations and agencies. They organized huge construction projections in the 1930s, gigantic industrial mobilization during World War II,… [and, dare I say, censuses and surveys]. Now nobody wants to be an Organization Man. We like start-ups, disrupters, and rebels. Creativity is honored more than the administrative execution. Post-Internet, many people assume that big problems can be solved by swarms of small, loosely networked nonprofits and social entrepreneurs. Big hierarchical organizations are dinosaurs."

- David Brooks, "Goodbye, Organization Man", *International Herald Tribune*, 9/17/14

Pros and Cons of Big Data

- There are a diversity of data streams and methods, implying varying degrees of expertise and investments in data cleaning
- Big data applications can raise privacy and confidentiality concerns
- Using novel data can be especially useful in:
  - data-poor developing country environments
  - Research areas plagued by poor data
- Most novel data streams are still validated against traditional census and survey data
- *Ergo*, we still need to keep investing in traditional data collection systems!

5. BRINGING IT ALL TOGETHER

Summary

- The SDGs represent a comprehensive attempt to set targets for and monitor the planet's social and environmental systems
- Open access data are an important component of data democracy, promoting civil society engagement
- Novel data streams hold promise but can't always substitute for traditional data collection approaches
- There need to be continued (and even increased) investments in the core data systems
  - National statistical agencies
  - Ground-based monitoring systems
- Similar investments need to be made to support data custodians (for documentation and dissemination)

“Good data can help progressive leaders and civil servants make their case and implement better policies. And, ideally, you need facts to know whether your policies are working – it is no good just having good intentions.

“On the other, there is a danger of getting overexcited by datasets and infographics. Where rapid poverty reduction has taken place in recent years, is better data behind it? Are we continuing to pollute our planet and make it more unequal because we lack data? … Do we need to know precisely how many people are hungry, and how hungry they are, before making sure there is enough food for them?...

“Just like all other development interventions, the question that matters most is this: are poor and marginalised communities more powerful than before?”

- Jonathan Glennie “A development data revolution
needs to go beyond the geeks and bean-counters”, *The Guardian,* 3 October 2013

Feel free to contact me:

Alex de Sherbinin
adesherbinin@ciesin.columbia.edu

THANK YOU!