BIG DATA AND THE SDGS

Marcelo T. LaFleur

UN DESA Development Research Branch

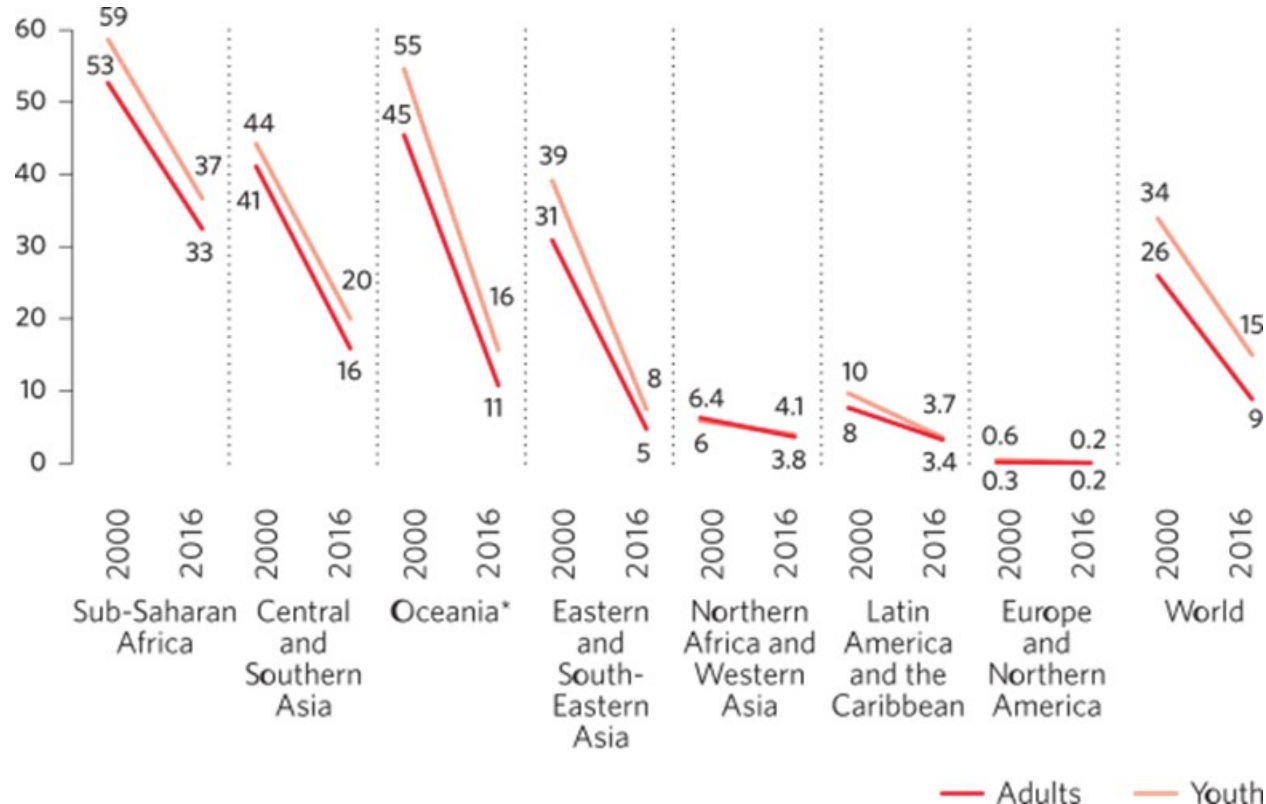Economic Analysis and Policy Division

17 integrated and indivisible Sustainable Development Goals and targets. 232 indicators
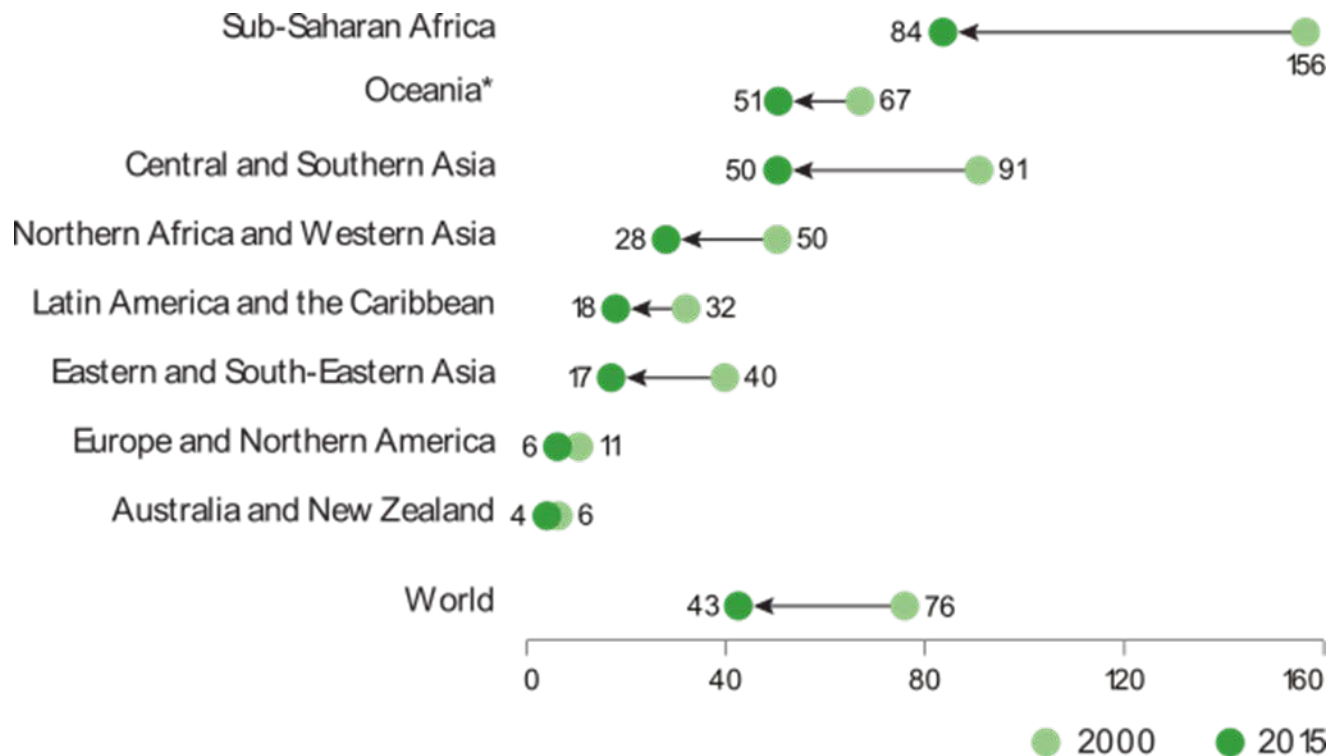
Poverty

Proportion of employed population living below 1.90 US dollars a day, youth and adults, 2000 and 2016 (percentage)

Child mortality

Under-5 mortality rate, 2000 and 2015 (deaths per 1,000 live births)

Monitoring of progress - UNSD

- Statistical Classifications
- Big data for official statistics
- Development indicators
- Geospatial Information
- Demographic and Social Statistics
- Gender Statistics
- Industry Statistics

- Trade Statistics
- Tourism Statistics
- National Accounts
- Energy Statistics
- Environment Statistics
- Environmental-Economic Accounting
- @Data4SDGs, #SDGInsights

Data methodology and coverage varies

Tier I	101	Conceptually clear, has an internationally established methodology and standards are available, and data are regularly produced by countries for <u>at least 50 per cent</u> of countries and of the population in every region where the indicator is relevant.
Tier II	91	Conceptually clear, has an internationally established methodology and standards are available, but <u>data are not regularly produced</u> by countries.
Tier III	34	<u>No internationally established methodology or standards are yet available</u> for the indicator, but methodology/standards are being (or will be) developed or tested.
Mixed	6	Multiple tiers.

Available data does not match our ambition

- We know little about poor countries and even less about the poor people who live in these countries.
- Econometric models used for testing hypothesis are much more sophisticated than the quality of the data would justify.
- Representative at large scales
  - Surveys of households are mostly conducted at wide scales, often national or across large administrative regions.
- Low frequency, high cost
  - Samples occur across days, weeks or sometimes months
  - Cannot track seasonal fluctuations from agricultural cycles or seasonal construction
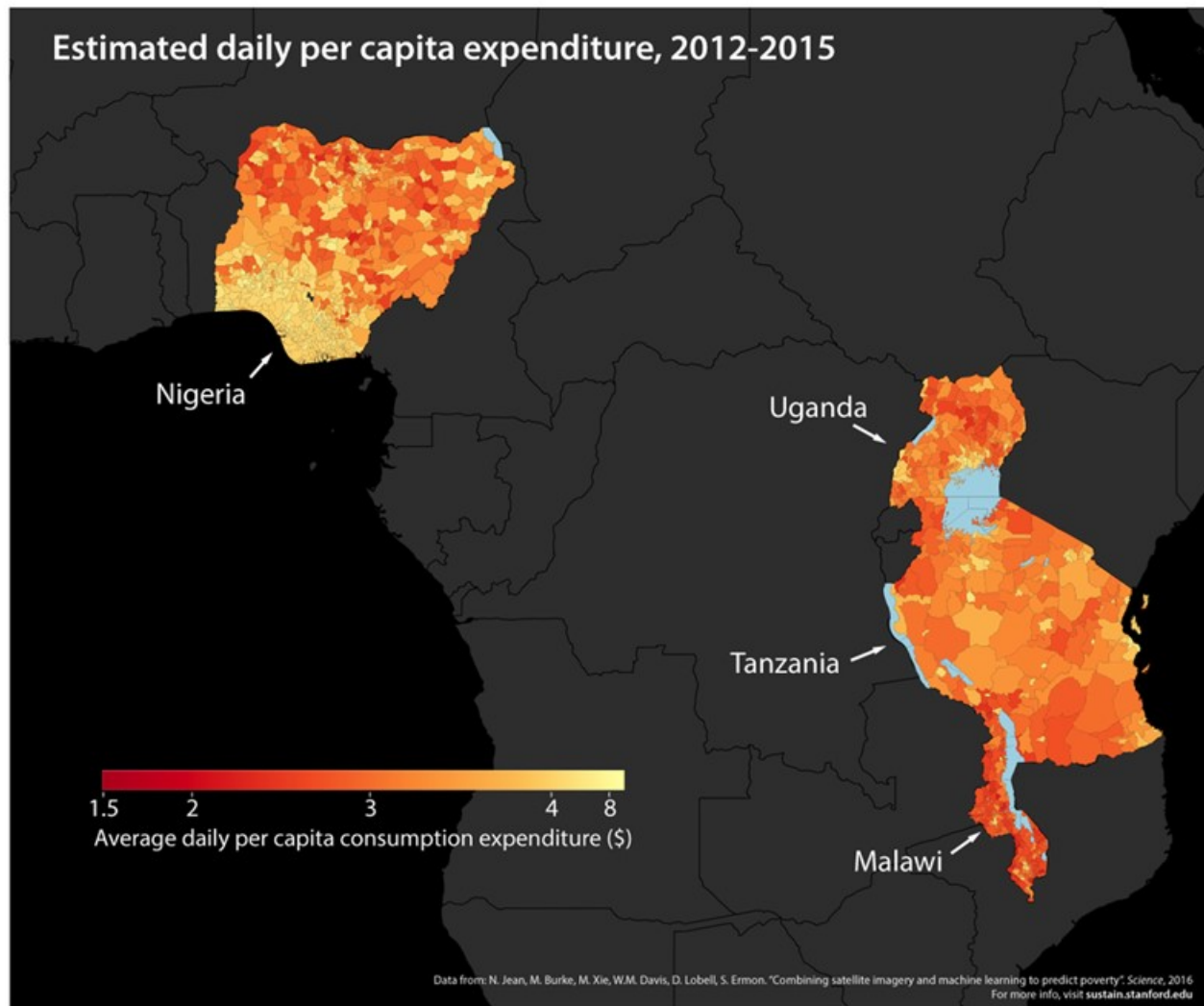
How to improve data coverage and quality?

□ We need more detailed data for individuals and households.

□ We need innovation in measuring data, combining data sets, and analyzing relationships.

Nightlights paper

- ☐ Focus on predicting poverty
- ☐ Use the intensity of light illumination as proxy for poverty
- ☐ Train ML algorithm to associate physical features with light intensity (poverty)
- ☐ Validate model with existing wealth surveys
- ☐ Now can predict poverty and distribution

Estimated daily per capita expenditure, 2012-2015

Nigeria

Uganda

Tanzania

Malawi

Average daily per capita consumption expenditure ($)

1.5 2 3 4 8

Data from: N. Jean, M. Burke, M. Xie, W.M. Davis, D. Lobell, S. Ermon. "Combining satellite imagery and machine learning to predict poverty". Science, 2016
For more info, visit sustain.stanford.edu

The data is out there…

- Imagine what else may be possible once creative and clever researchers have access to the vast amounts of data from many different fields?

- What policy options can be discovered as we pursue the 2030 Agenda?

#SDGinsights

How might we combine household survey data with other georeferenced datasets to provide insights into the status and progress of the SDGs?

UNDESA USAID unite IDEAS

| Personal data / Big Data / Internet of Things | Capture data | Store and organize | Analyze | Report | Impact |

Capture data
- Internet of Things
- Big Data
- New sources

Store and organize
- Harmonization
- API for sharing data
- Data science and development
- Cloud providers

Analyze
- Analytics package
- Machine learning
- Algorithm development

Report
- Visualization
- User interface
- Investigation and discovery

Expand into new functions

Expand into new stages

Global Fundamental Geospatial Data Themes

Global Geodetic
Reference Frame

Addresses

Buildings and
Settlements

Elevation and
Depth

Functional Areas

Geographical
Names

Geology and
Soils

Land Cover
and Use

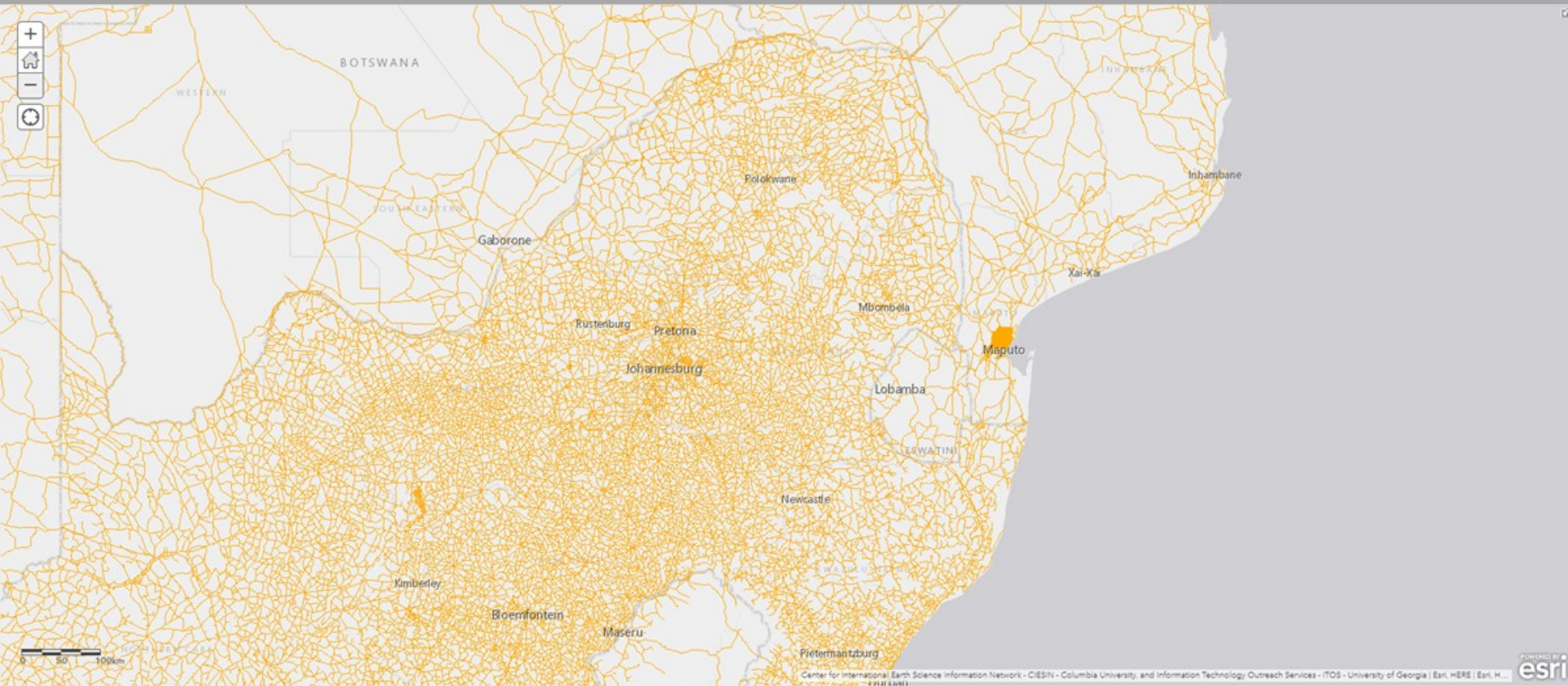Land Parcels

Orthoimagery

Physical
Infrastructure
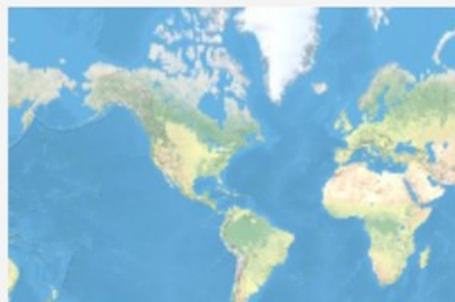
Population
Distribution

Transport
Networks

Water

World Population

This map features estimated global population modeled using road intersection density, imagery, urban footprints, and 1.6 million census polygons

Landsat GLS (NDVI)

Contains 30 meter resolution Landsat imagery from 1990-2010. Normalized Difference Vegetation Index (NDVI) is a compilation of visible

Landsat Time Enabled Imagery

This is a map of the time enabled Landsat services and can be used to explore change over time in web browsers

Under-five mortality rate - 2000

Under-five mortality rate - 2015

Under-five mortality rate - 2015

Under-five mortality rate - 2015

Indicator 1.2.1 - Proportion of total popula...

Proportion of total population living below the national poverty line

Indicator 6.5.1 - Degree of integrated wat...

Degree of integrated water resources management implementation

Global Roads

A HTTPS map of global roads.

Thank you