Designing An Information Commons for Sustainability Science: Lessons Learned from a World Data Center

Marc Levy
CIESIN
Earth Institute, Columbia University

Presentation to International Workshop on Designing Global Information Commons for Innovation in Frontier Sciences
8-10 November 2007
Tokyo, Japan

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

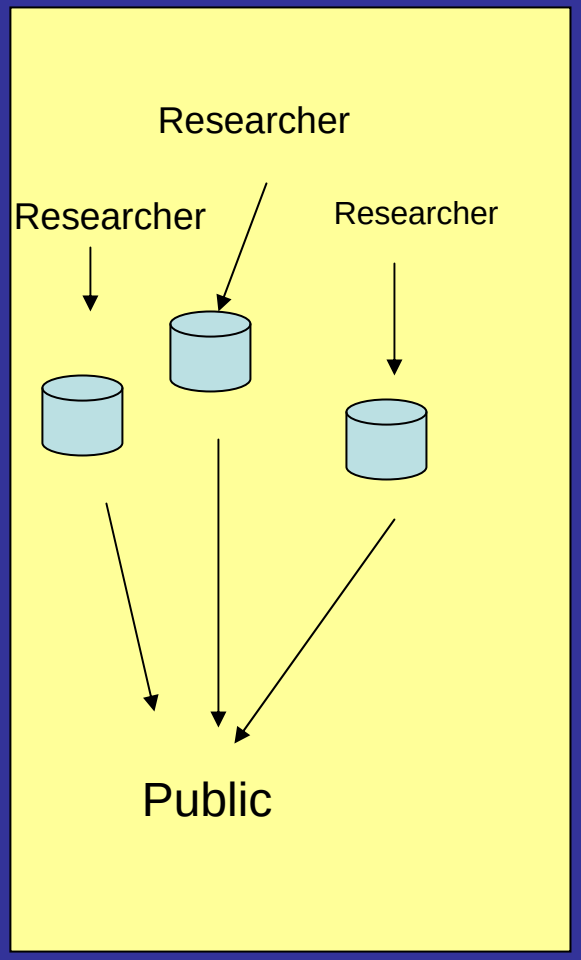
The Earth Institute
AT COLUMBIA UNIVERSITY

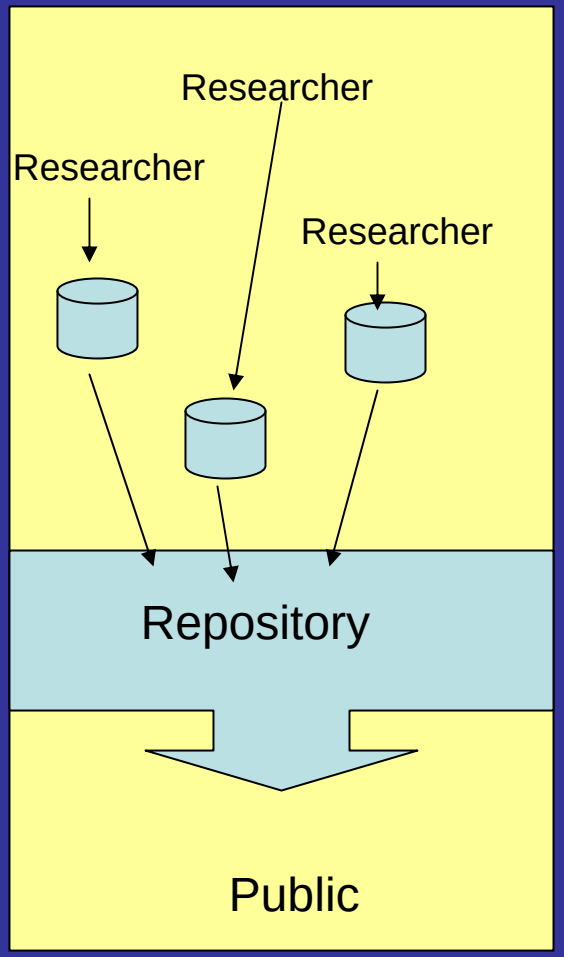
CIESIN
Columbia University

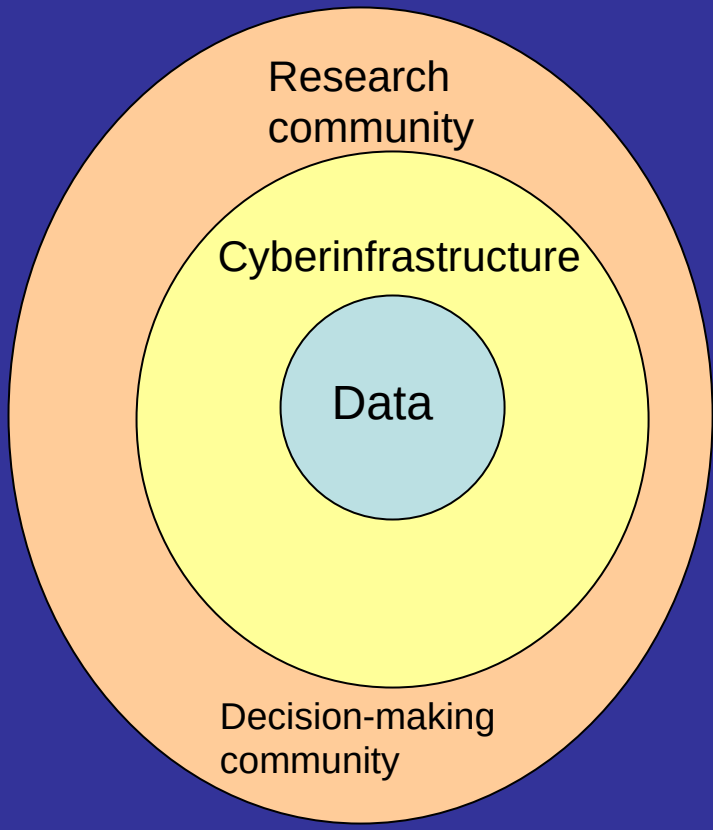
WDC
World Data Center for Human Interactions in the Environment

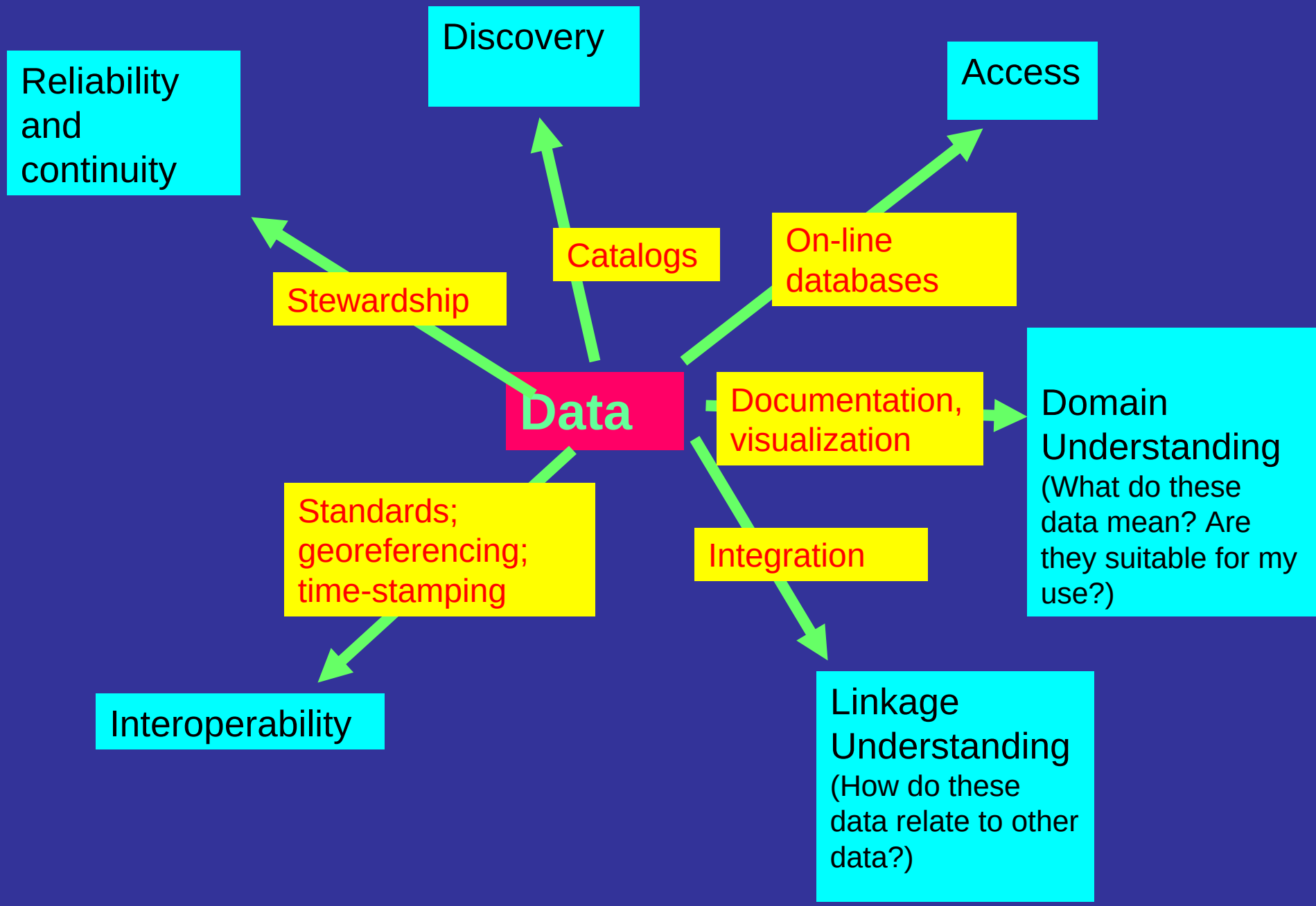
SEDAC

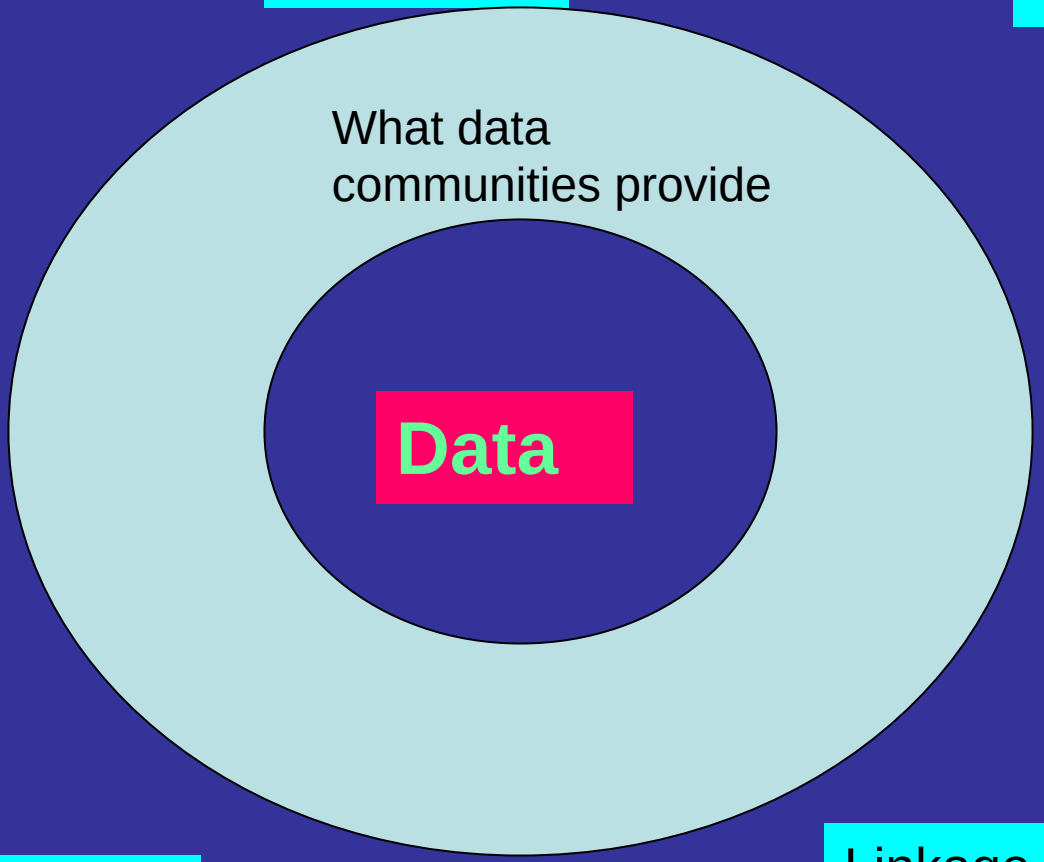
Evolution of the Data Provider Model








Reliability and continuity
Discovery
Access
Stewardship
Catalogs
On-line databases
Data
Documentation, visualization
Domain Understanding
(What do these data mean? Are they suitable for my use?)
Standards; georeferencing; time-stamping
Integration
Interoperability
Linkage Understanding
(How do these data relate to other data?)



Discovery

Access

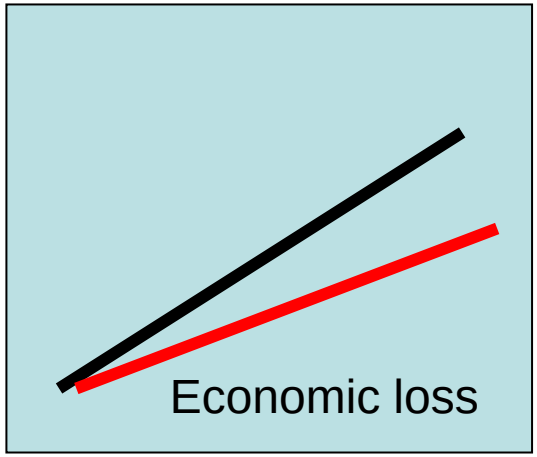
Reliability and continuity

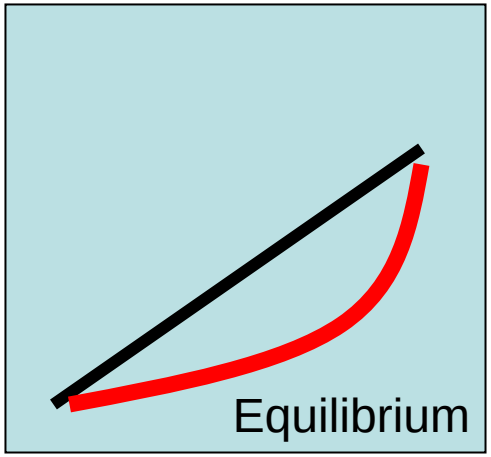
Domain Understanding
(What do these data mean? Are they suitable for my use?)

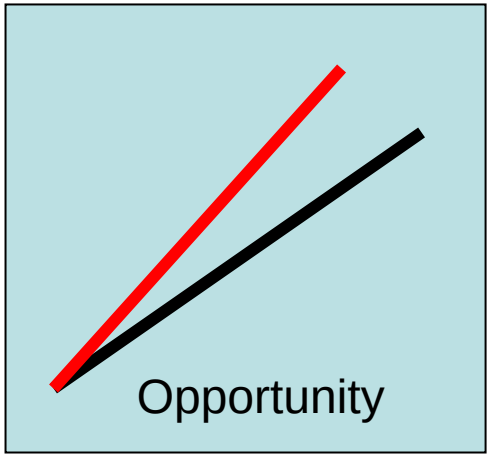
Linkage Understanding
(How do these data relate to other data?)

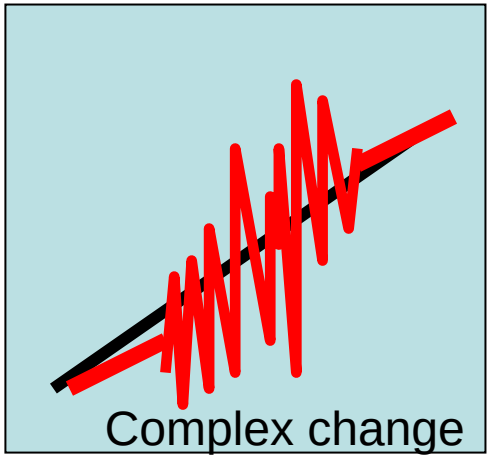
Interoperability

Dynamics of Human-Environment Interactions

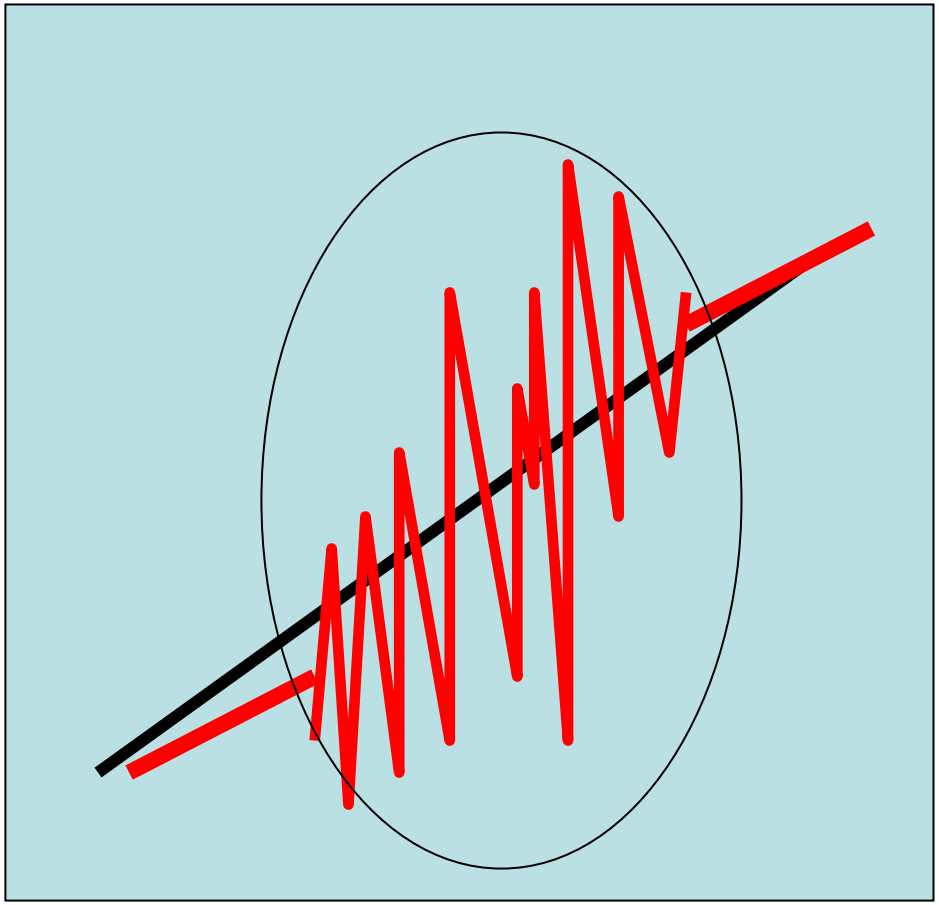








This is our world



- Multiple stresses
  - Economic
  - Demographic
  - Political
  - Change in land cover
  - Water scarcity
  - Soil fertility problems
- Complex outcomes
  - Livelihoods
  - Health
  - Equity
  - Security
  - Culture

Helping Users Make Wise Choices is Hard!

Traditional Documentation not enough

Multi-faceted approach required

- **Comparative Guides**
- **Visualizations**
- **Common Pitfalls**
- **Examples**
- **Citations**

SEDAC Citation Index

Browse references citing SEDAC data, applications, publications, and projects:

- Archive of Census Related Products (ACRP)
- Central American Vegetation / Land Cover Classification and Conservation Status Data (PROARCA)
- China Dimensions Data Collection
- Crop Climate Datasets
- Demographic Data Viewer (DDViewer)
- Environmental Treaties and Resource Indicators (ENTRI)
- Environmental Performance Index (EPI)
- Environmental Sustainability Index (ESI)
- Geographic Correspondence Engine (Geocorr)
- Georeference Population Data Sets of Mexico
- Global Rural-Urban Mapping Project (GRUMP)
- Gridded Population of the World (GPW)
- Human Footprint and Last of the Wild Datasets
- IPCC Socioeconomic Data Distribution Centre (DDC)
- Model Visualization and Analysis (MVA)
- Ozone and Human Health
- Population-Environment Research Network (PERN)
- Public Use Microdata Samples (PUMS)
- Remote Sensing and Environmental Treaties Workshop
- Thematic Guides

SEND US YOUR CITATIONS!

If you know of any publication citing data or information obtained from SEDAC, please **send us** the citation!

CITATION GUIDELINES

Guide to formatting citations.

http://sedac.ciesin.org/citations/CitationIndex.html

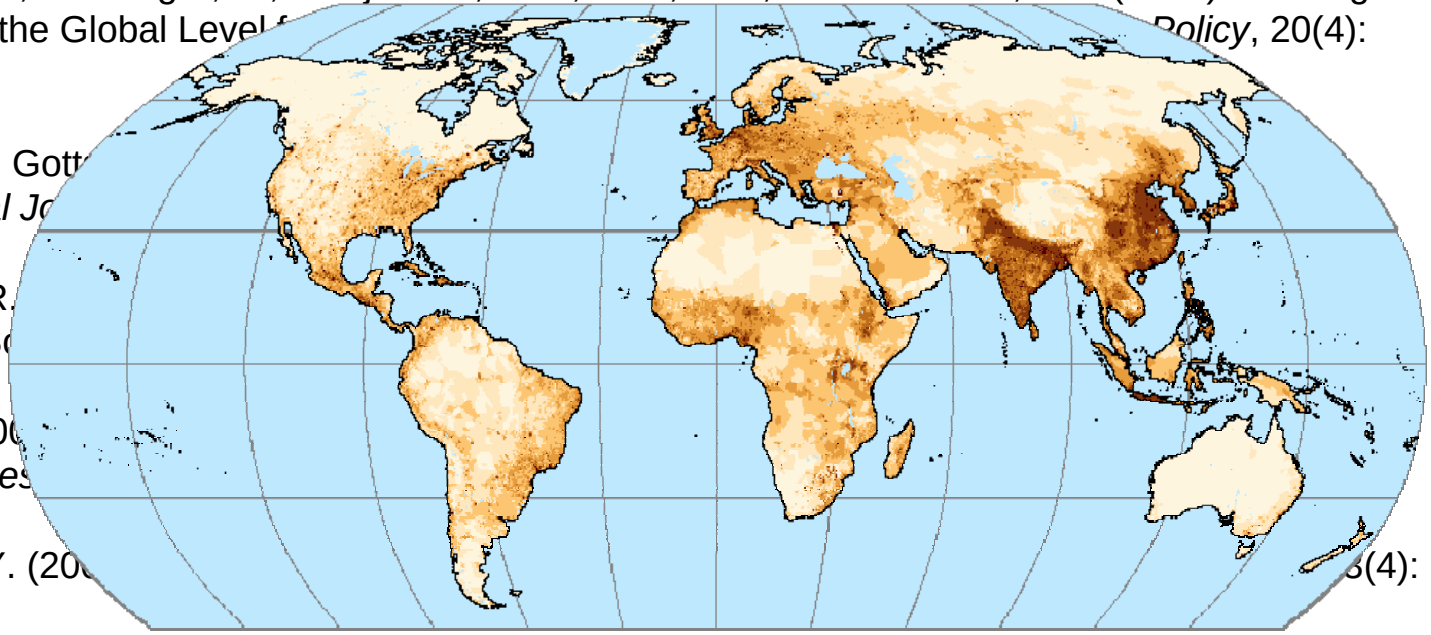
Poor Livestock Keepers at the Global Level [illegible] *Policy*, 20(4): 311-322.

Tobler, W., Deichmann, U., Gott[illegible] Quadrilaterals. *International Jo[illegible]*

van Lieshout, M., Kovats, R.[illegible] of the SRES Climate and So[illegible]

Vassolo, S. and Döll, P. (20[illegible] Water Use. *Water Resources[illegible]*

Verburg, P.H., and Chen, Y. (20[illegible] 3(4): 369-385.



Viviroli, D. and Weingartner, R. (2004) The Hydrological Significance of Mountains: from Regional to Global Scale. *Hydrology and Earth System Science,* 8(6): 1016-1029.

Vorosmarty, C.J., Green, P., Salisbury, J., and Lammers, R.B. (2000) Global Water Resources: Vulnerability From Climate Change Acid Population Growth. *Science*, 289(5477): 284-288.

Vorosmarty, C.J., and Sahagian, D. (2000) Anthropogenic Disturbance of the Terrestrial Water Cycle. *Bioscience*, 50(9): 753-765.

White, M.A., Hoffman, F., Hargrove, W.W., and Nemani, R.R. (2005) A Global Framework for Monitoring Phenological Responses to Climate Change. *Geophysical Research Letters,* 32(L04705): 4pp.

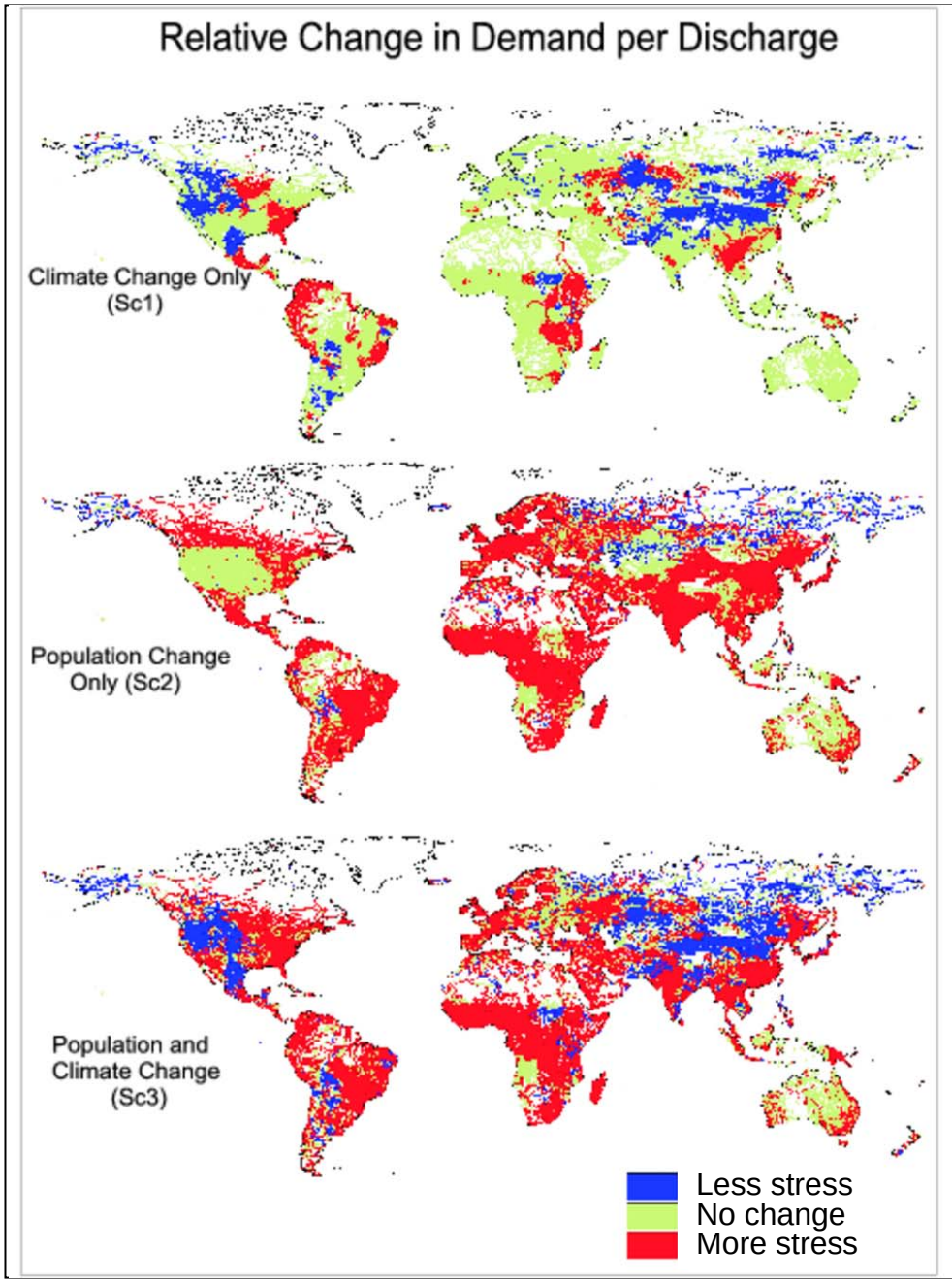
White, M.A., Nemani, R.R., Thornton, P.E., and Running, S.W. (2002) Satellite Evidence of Phenological Differences Between Urbanized and Rural Areas of the Eastern United States Deciduous Broadleaf Forest. *Ecosystems*, 5(3): 260-273.

Wilson, S.J., Steenhuisen, F., Pacnya, J.M., and Pacnya, E.G. (2006) Mapping the Spatial Distribution of Global

Water Stress Changes to 2025

- 80% of future stress from **population** & **development**, not **climate change**!
- Future distortions of the water cycle are inevitable
- Issue gaining momentum in global policy fora

(e.g. Millennium Assessment, World Water Assessment Programme, MDGs)



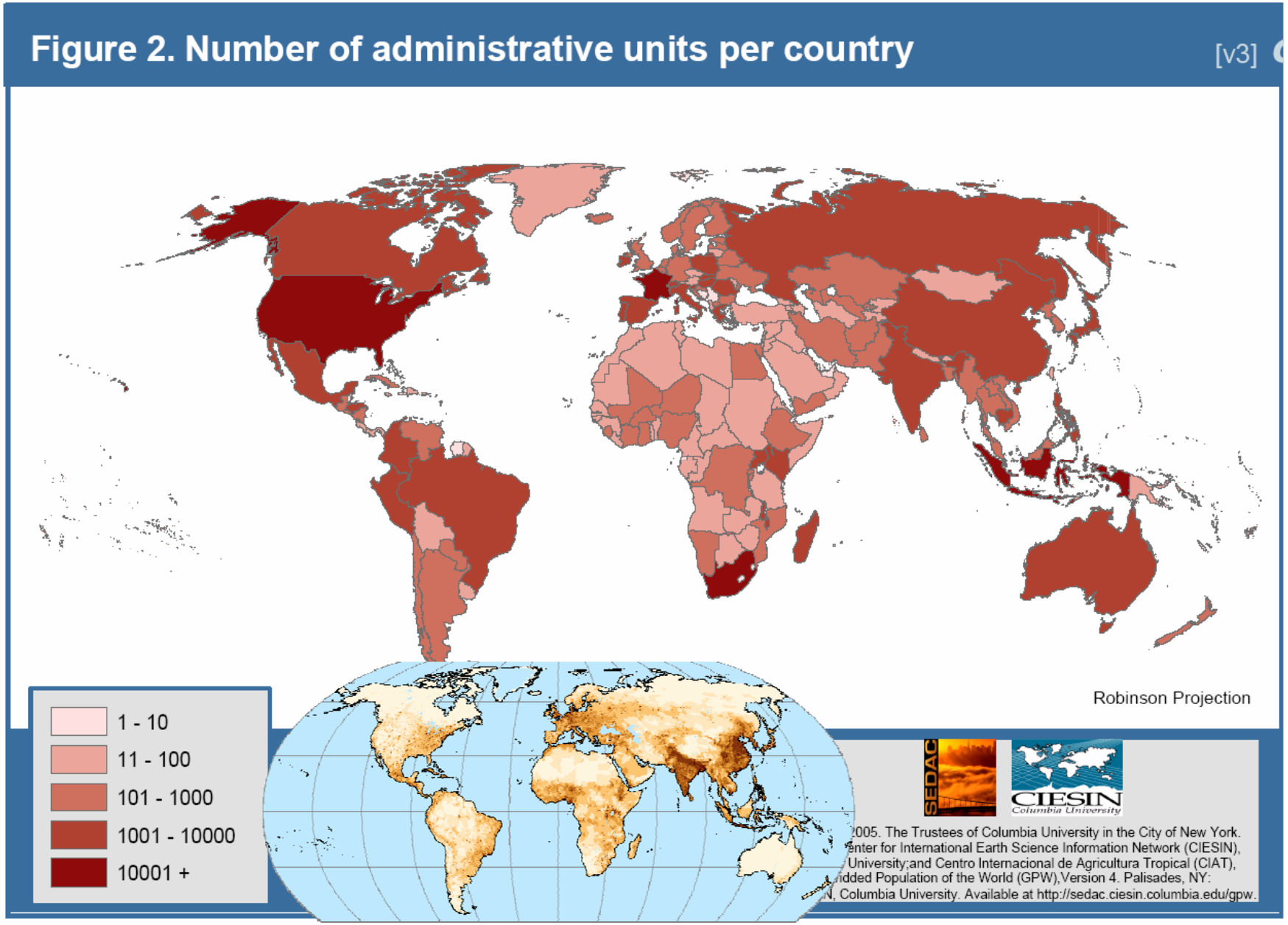
Source: Vörösmarty et al. 2000

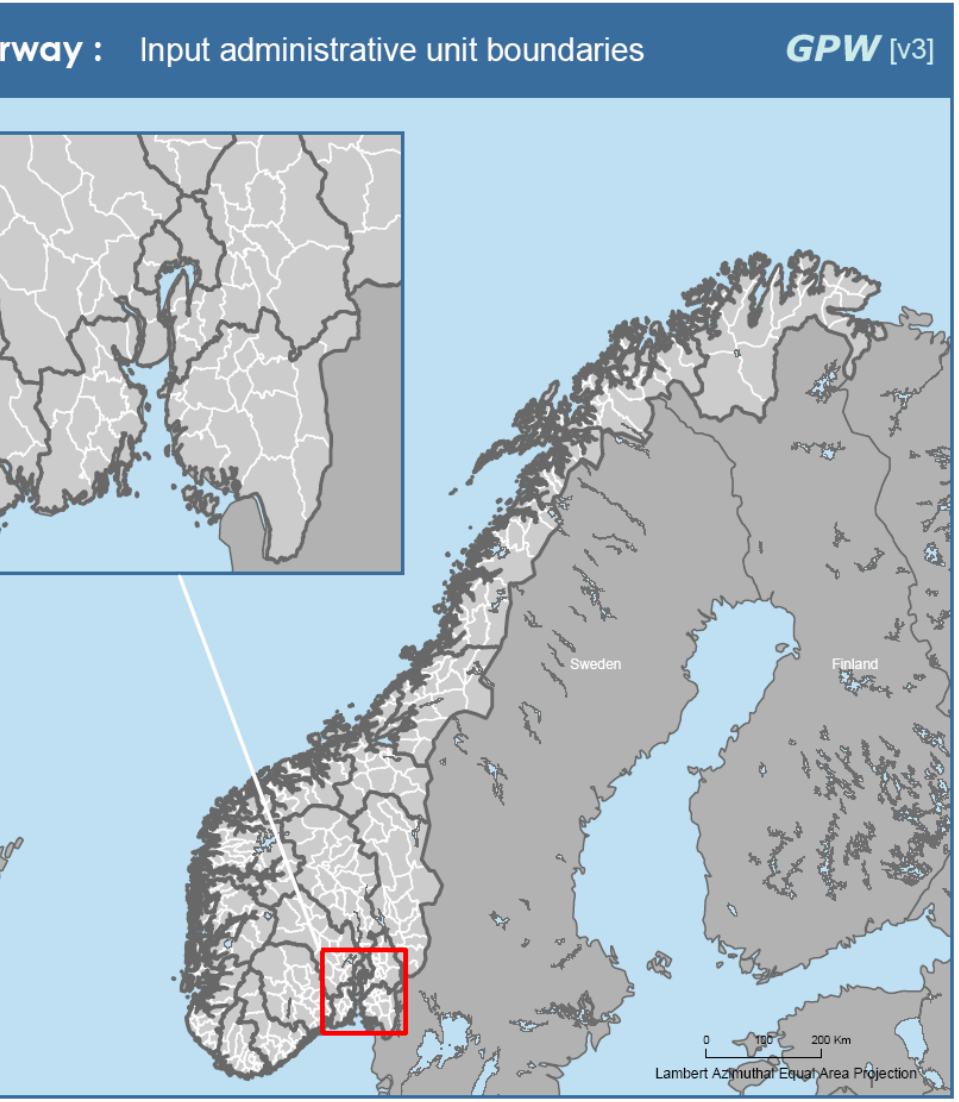
UNH

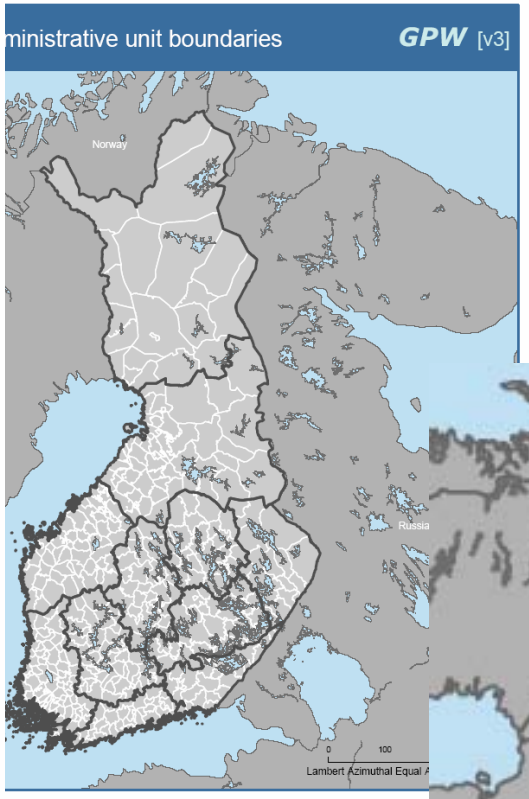
Table 5.2. Population Growth within MA Systems, 1990–2000

System	Change In Population (million)	Net Change In Population (percent)	Change In Population per Square Kilometer
Cultivated	505.7	14.1	14.3
Dryland	329.6	18.5	5.5
Inland Water	203.5	17.0	7.0
Mountain	171.0	16.3	5.4
Forest	142.1	13.5	3.4
Coastal	140.3	15.9	23.3
Island	67.0	12.3	9.5
Polar	–117.9	–6.5	0.0

Information on data quality is critical to judging goodness of fit







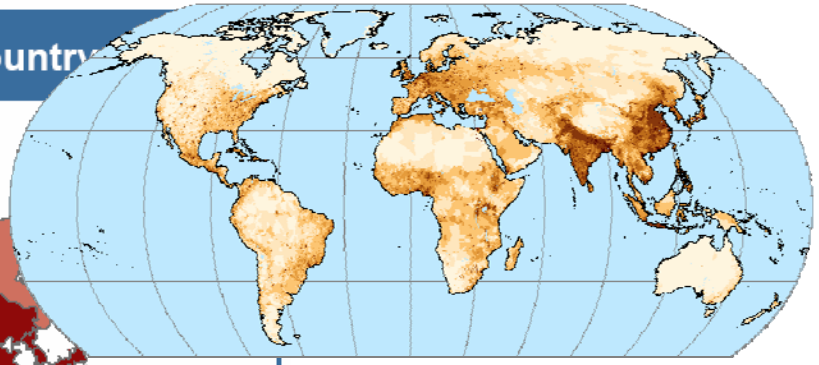
nistrative Boundaries
Level 1: Fylke
Level 2: Kommune

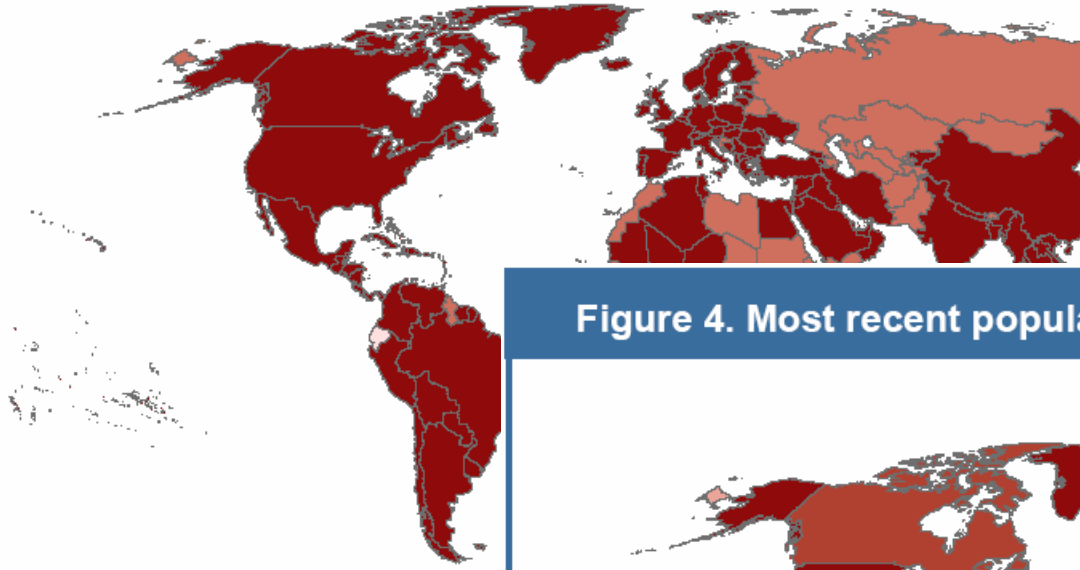
ndaries indicate level used in gridding.

Copyright 2005. The Trustees of Columbia University in the City of New York.
Source: Center for International Earth Science Information Network (CIESIN), Columbia University; Centro Internacional de Agricultura Tropical (CIAT); Gridded Population of the World (GPW), Version 3. Palisades, NY: CIESIN, Columbia University. Available at: http://sedac.ciesin.columbia.edu/gpw.

It shouldn't even be this hard

Figure 3. Number of population data reference years per country





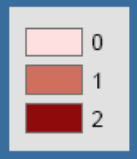
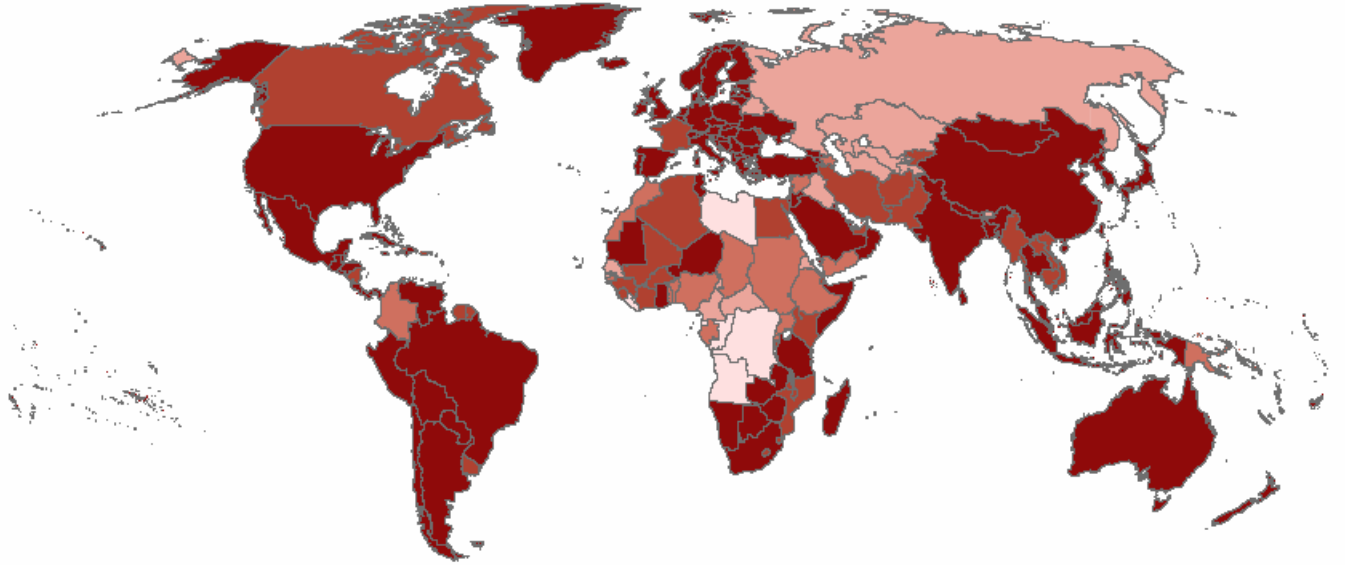
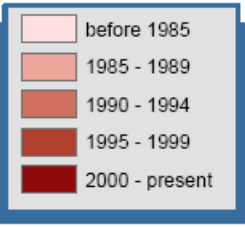


Figure 4. Most recent population data year

GPW



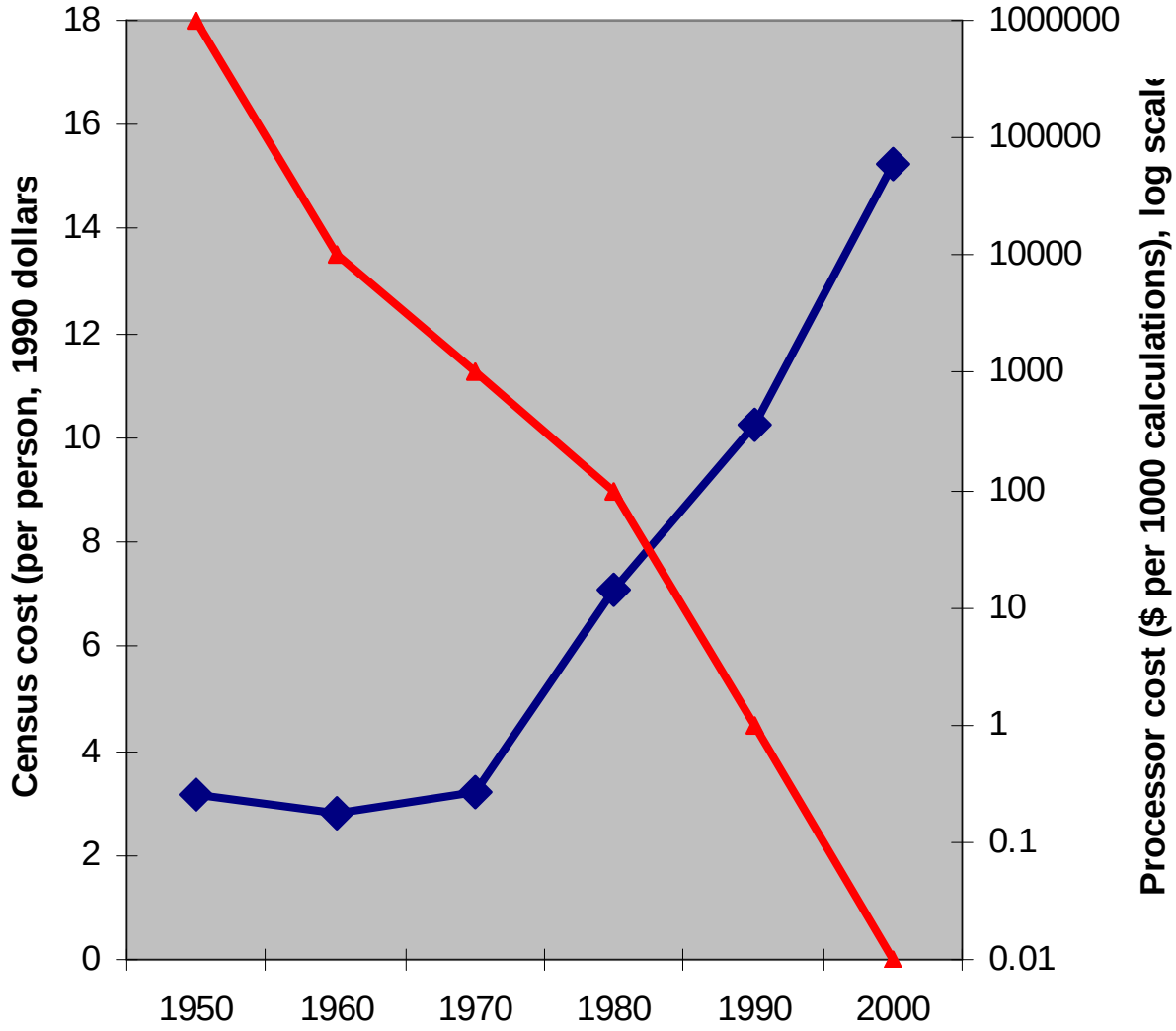
Robinson Projection



Copyright 2005. The Trustees of Columbia University in the City of New York.
Source: Center for International Earth Science Information Network (CIESIN), Columbia University;and Centro Internacional de Agricultura Tropical (CIAT), 2004. Gridded Population of the World (GPW),Version 4. Palisades, NY: CIESIN, Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw.

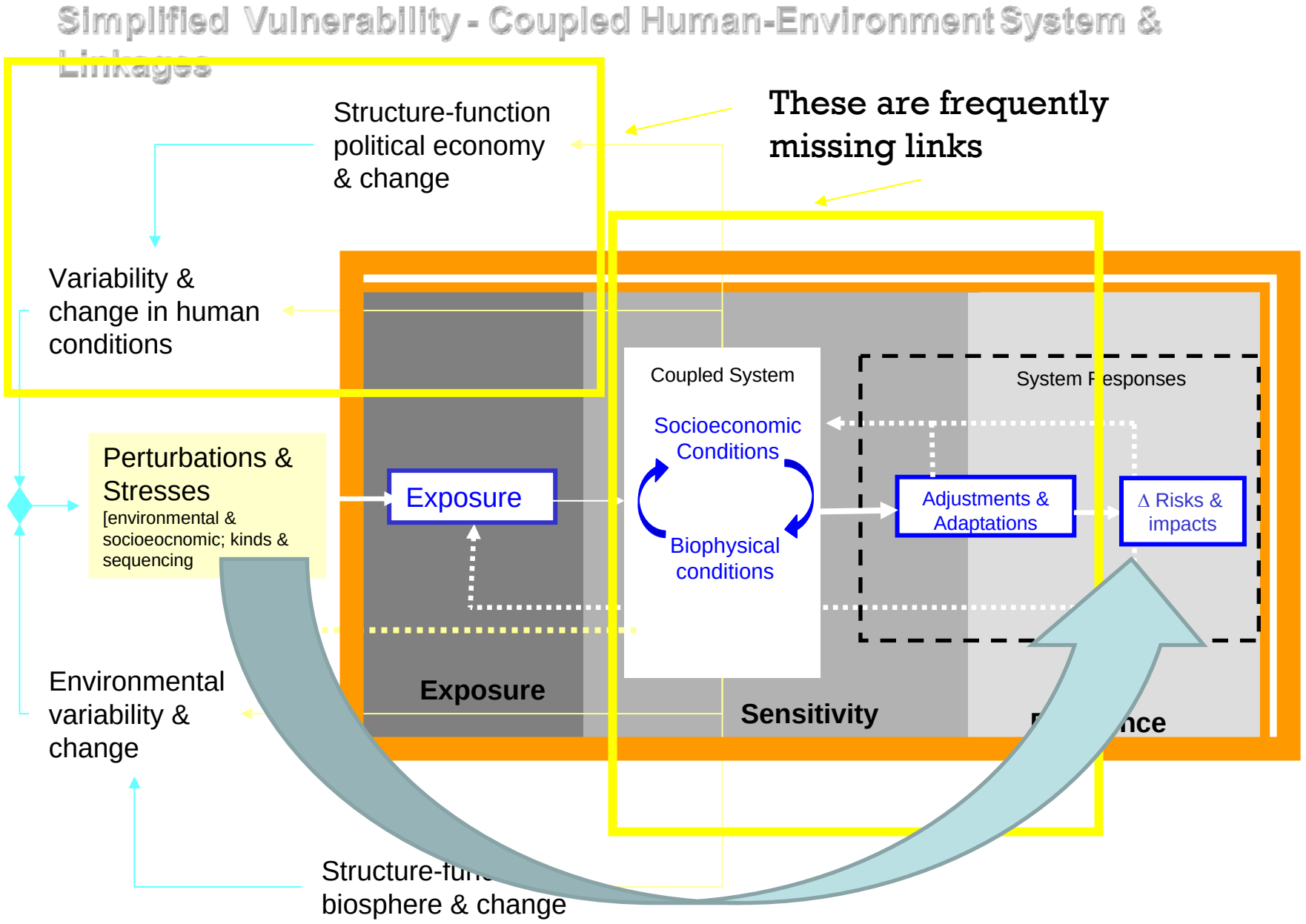
Helping users make wise choices is a community-building and community-strengthening task

Moore's Law Benefits Data Collection Processes Unequally



Pace of progress across data domains is very uneven.

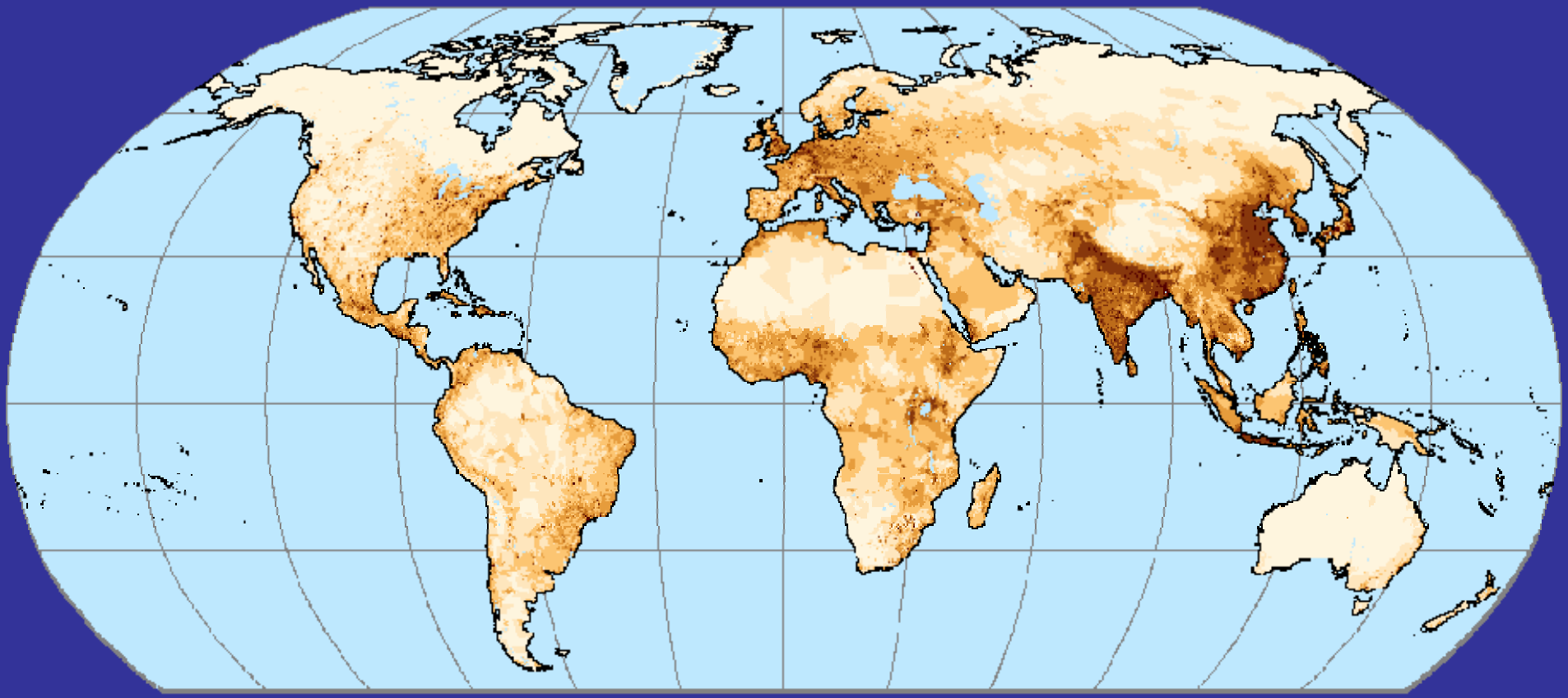
Greater the divergence, greater the ***Integration Frustration***



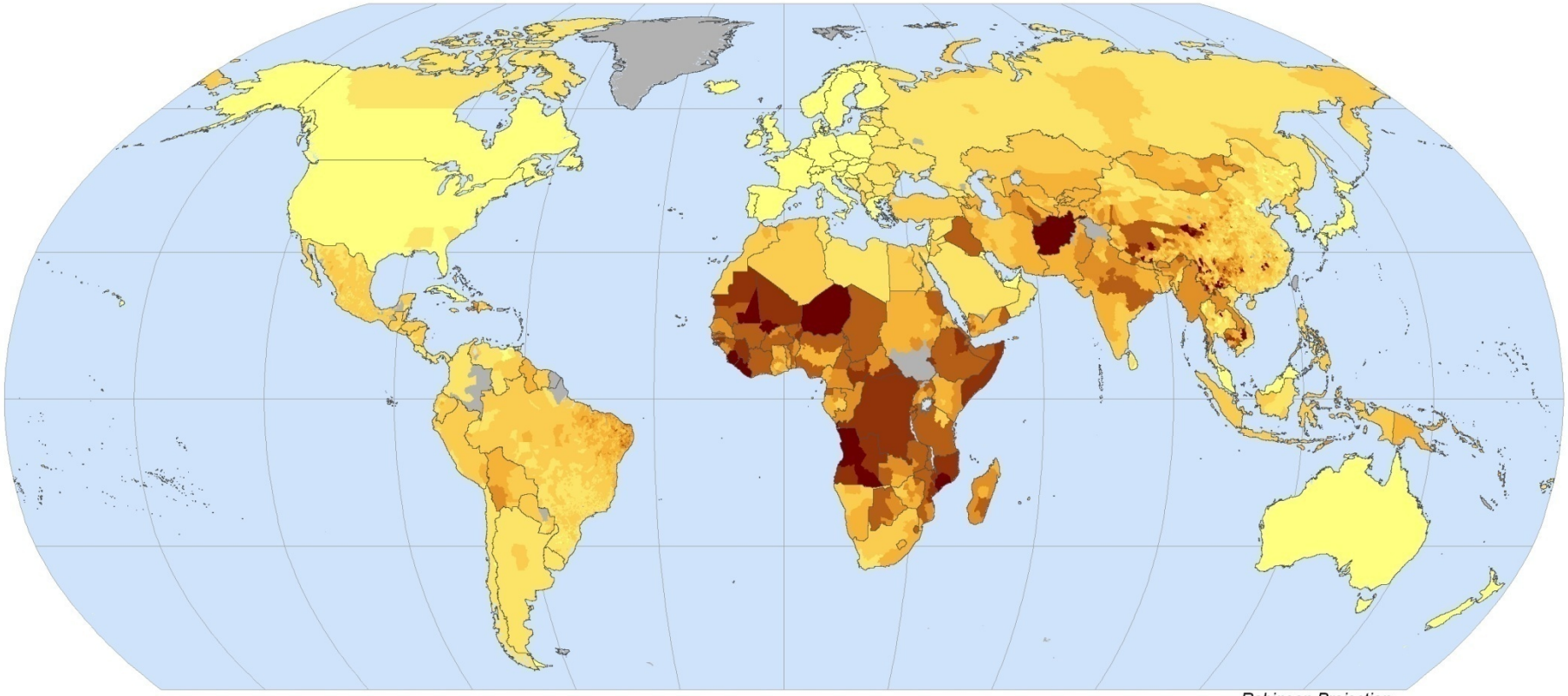
Turner, Matson, Kasperson and Kasperson, 2002

Temporal and spatial dynamics omitted

Identify and Fill Gaps!



CIESIN, Gridded Population of the World, 350,000 census input units



Robinson Projection

Measures of Poverty

Infant Mortality Rates

By Subnational Administrative Unit

Subnational mortality rates are adjusted to 2000 using national trend data.
Original data for 96% of countries are from 1995 or later. All data are from 1990 or later.

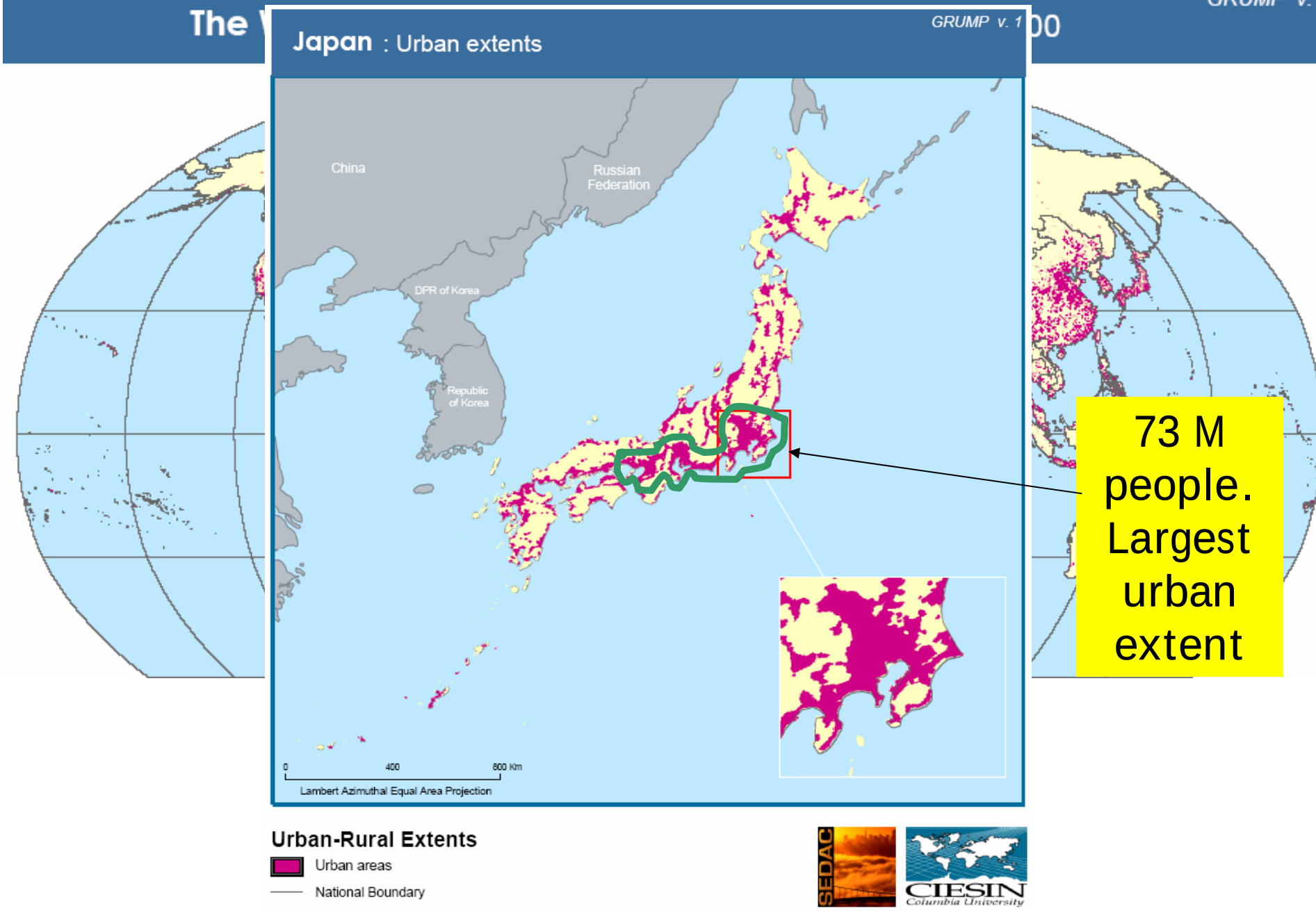
Infant mortality rate, 2000
(deaths per 1000 live births)

- < 9.1
- 9.1 - 25.0
- 25.1 - 50.0
- 50.1 - 75.0
- 75.1 - 100.0
- 100.1 - 125.0
- 125 - 150.0
- 150.0 <
- no data

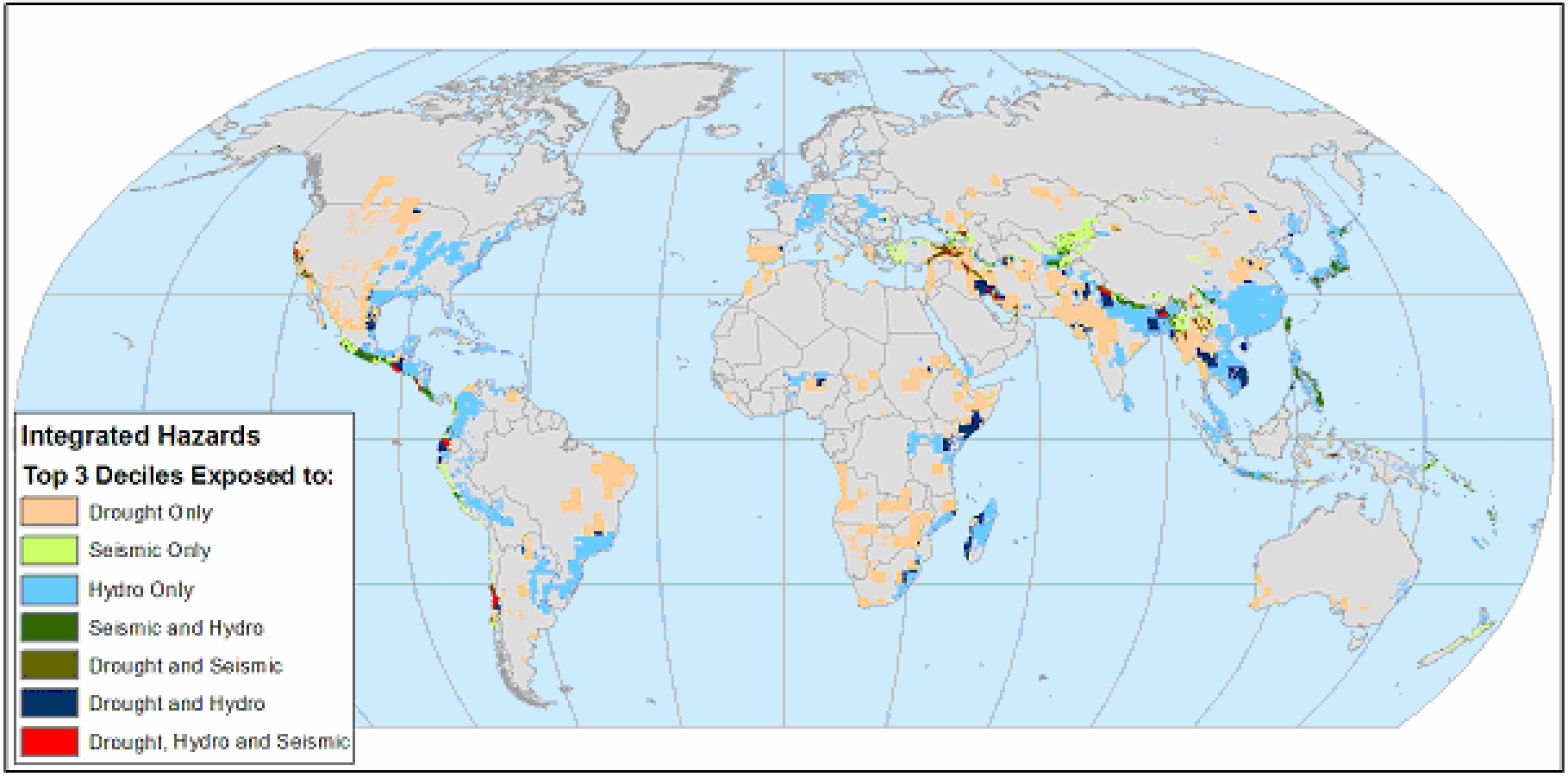
National Boundary

Subnational boundaries have been removed from countries for clarity.

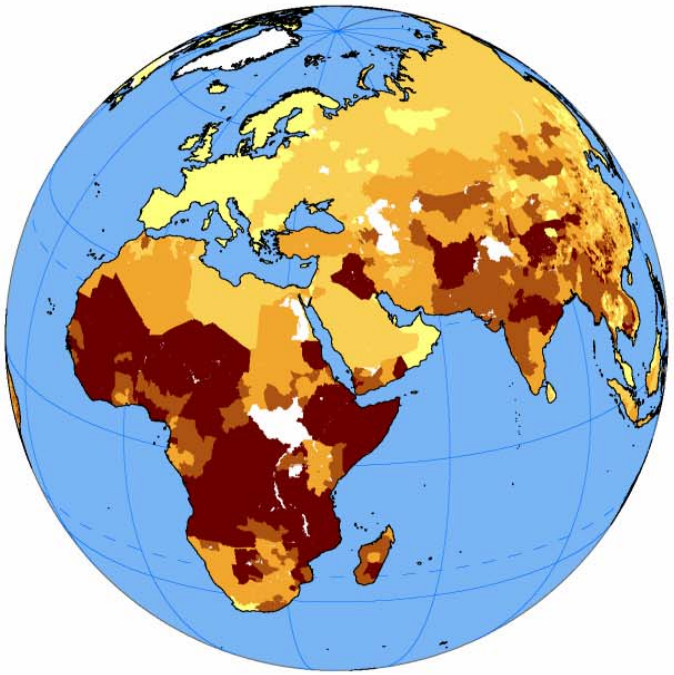
Copyright 2005. The Trustees of Columbia University in the City of New York.
Source: Center for International Earth Science Information Network (CIESIN).
Columbia University. Global subnational infant mortality rates; maps and
further documentation available at: http://www.ciesin.columbia.edu/povmap


Japan : Urban extents
GRUMP v. 1
China
Russian Federation
DPR of Korea
Republic of Korea
0
400
800 Km
Lambert Azimuthal Equal Area Projection
73 M people. Largest urban extent
Urban-Rural Extents
Urban areas
National Boundary
SEDAC
CIESIN
Columbia University

Exposure to Multiple Natural Hazards



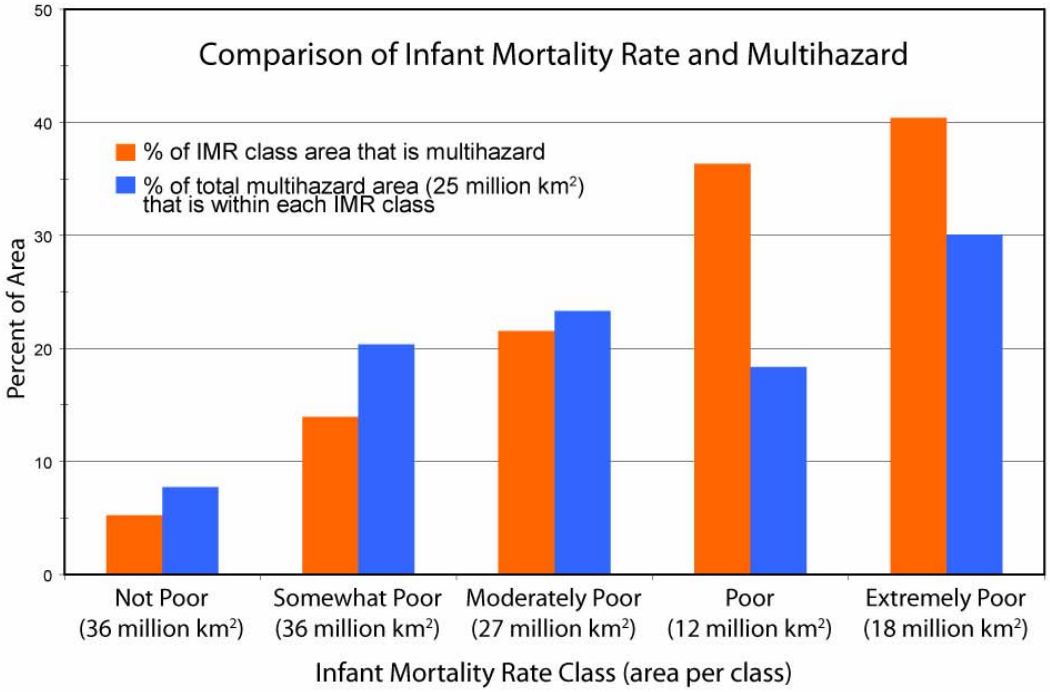
Seismic hazards include earthquakes and volcanoes; *hydrological hazards* include floods, cyclones, and landslides

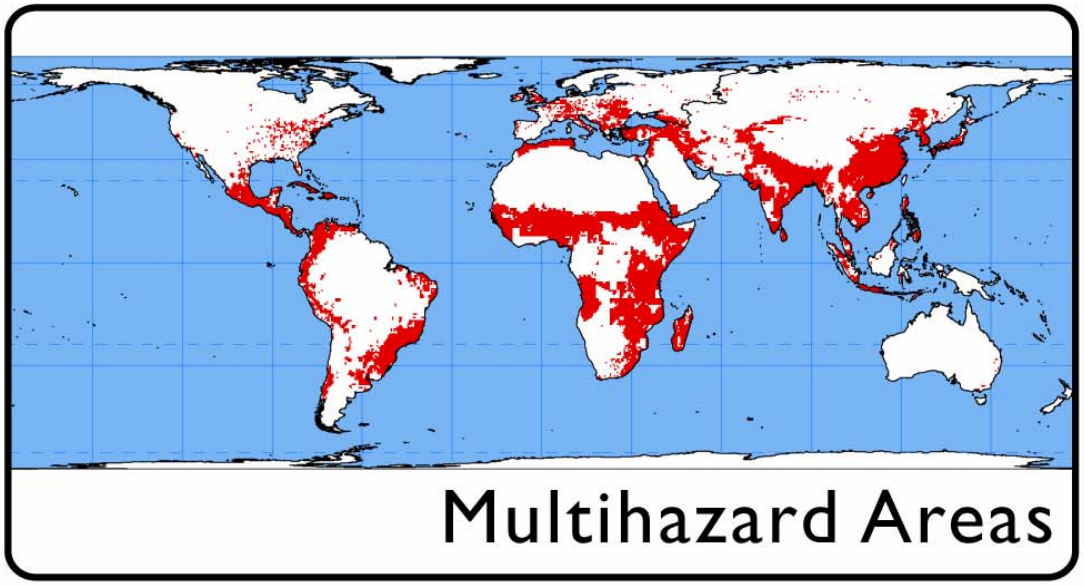


Infant Mortality Rates

(per 1,000)

Not Poor	1.9 - 15
Somewhat Poor	15.1 - 32
Moderately Poor	32.1 - 65
Poor	65.1 - 100
Extremely Poor	> 100

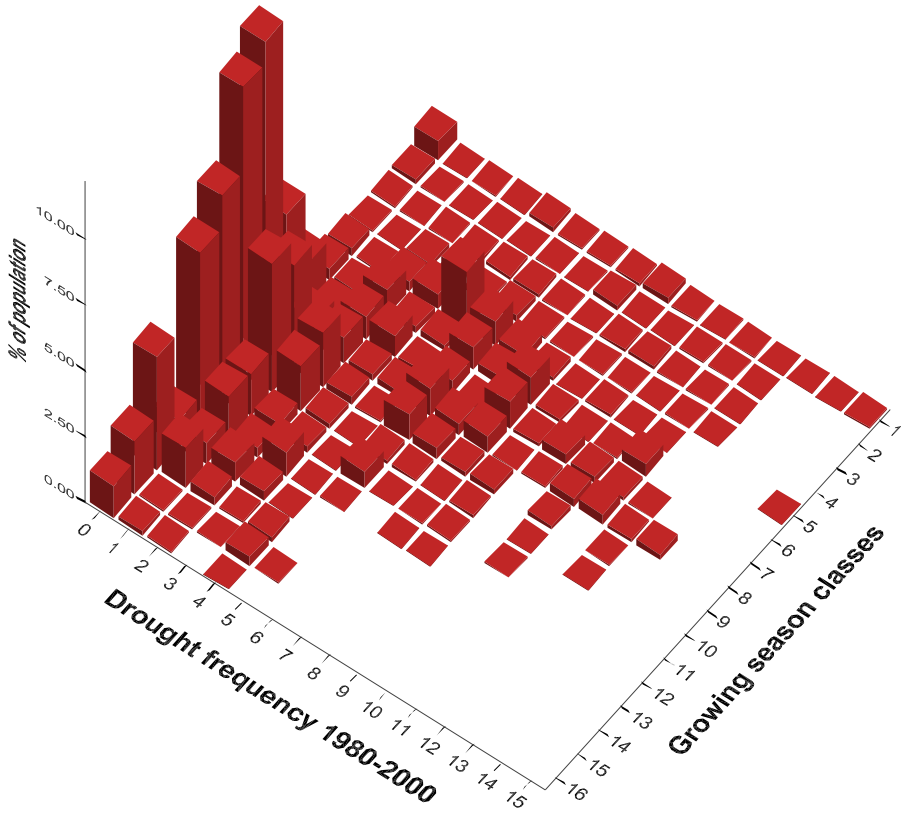




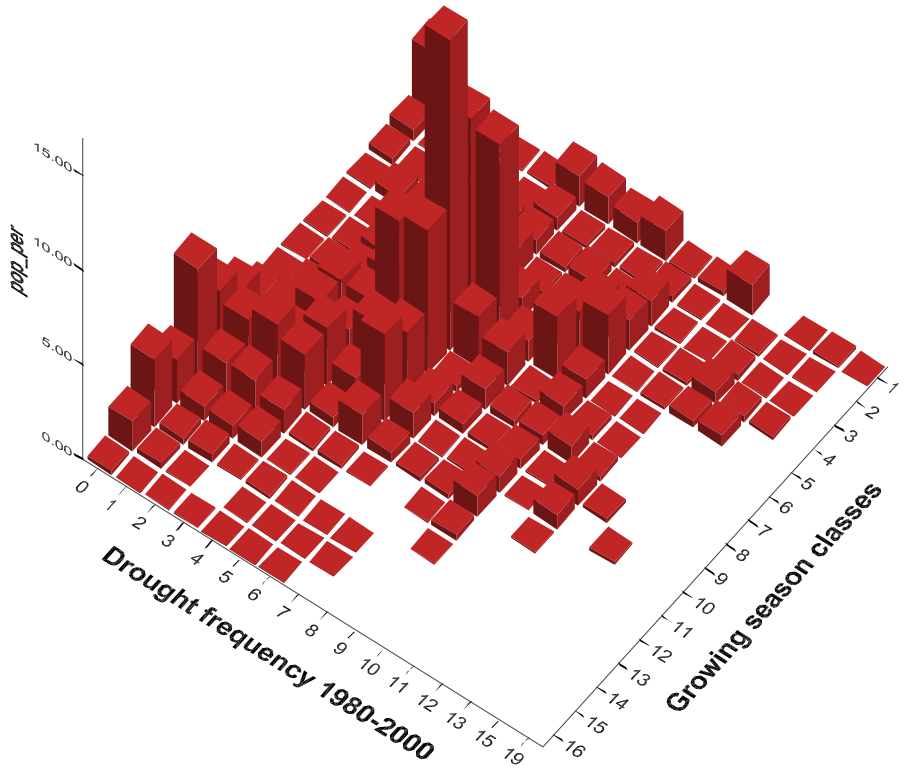
Multihazard Areas

Growing Season and Drought

Distribution of non-poor population



Distribution of poor and extremely poor population



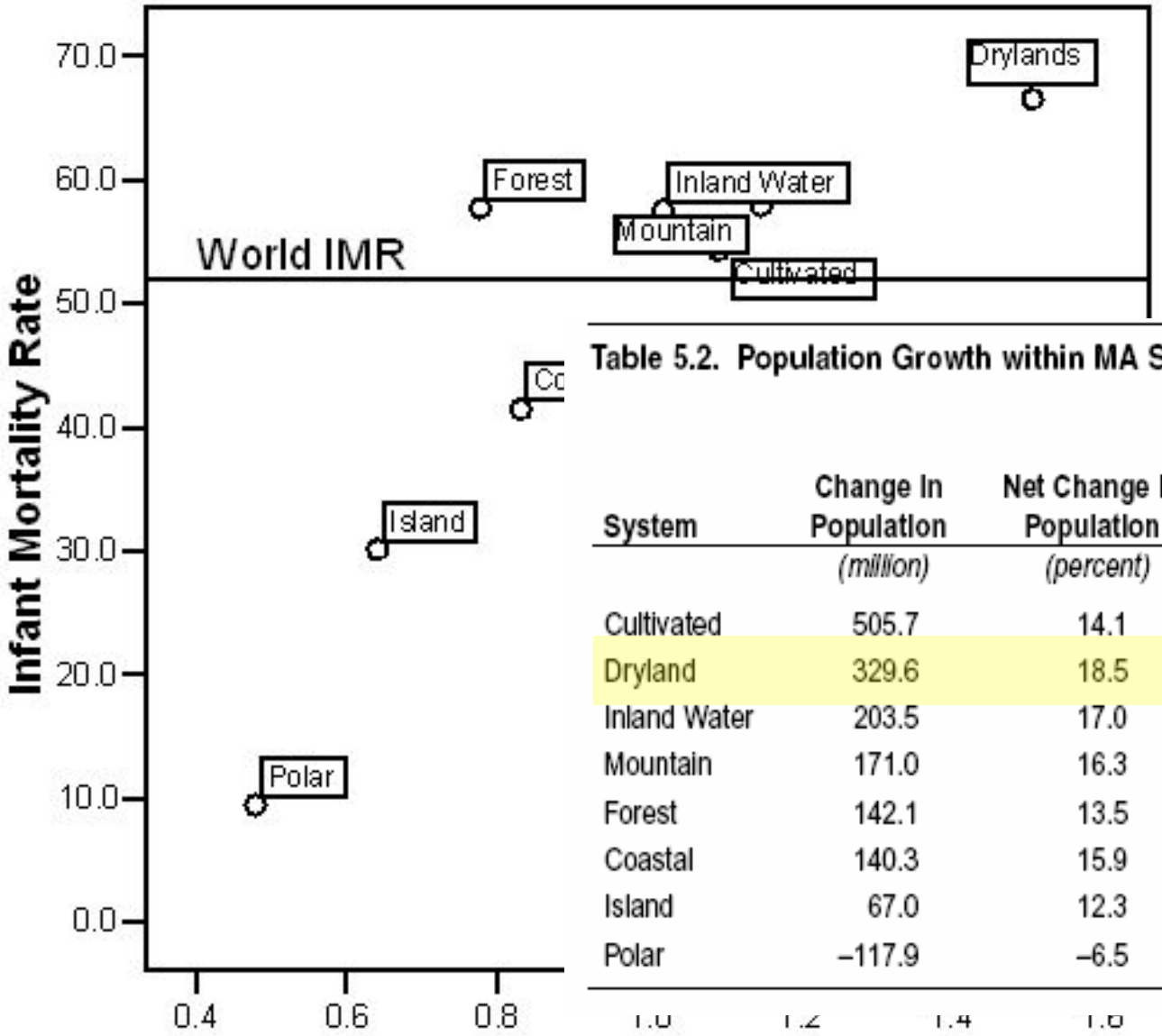


Table 5.2. Population Growth within MA Systems, 1990–2000

System	Change in Population (million)	Net Change in Population (percent)	Change in Population per Square Kilometer
Cultivated	505.7	14.1	14.3
Dryland	329.6	18.5	5.5
Inland Water	203.5	17.0	7.0
Mountain	171.0	16.3	5.4
Forest	142.1	13.5	3.4
Coastal	140.3	15.9	23.3
Island	67.0	12.3	9.5
Polar	–117.9	–6.5	0.0

Millennium Ecosystem Assessment, 2005

Many more gaps to fill!

- Roads
- Migration
- Time-series spatial d
  urbanziation
- Spatial economic da
- Soil fertility
- Spatial health data

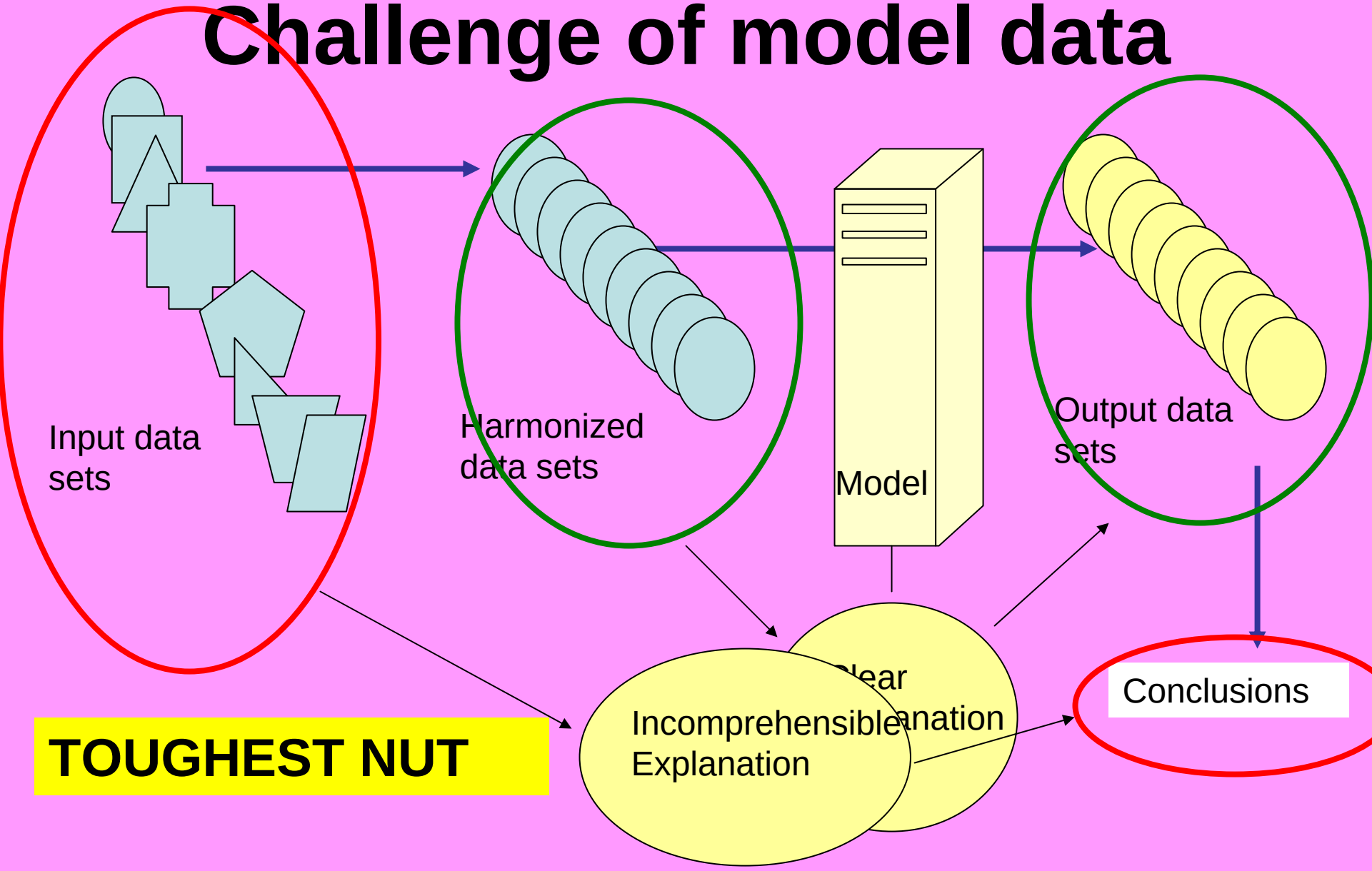
Prioritize

Assign roles

Be transparent

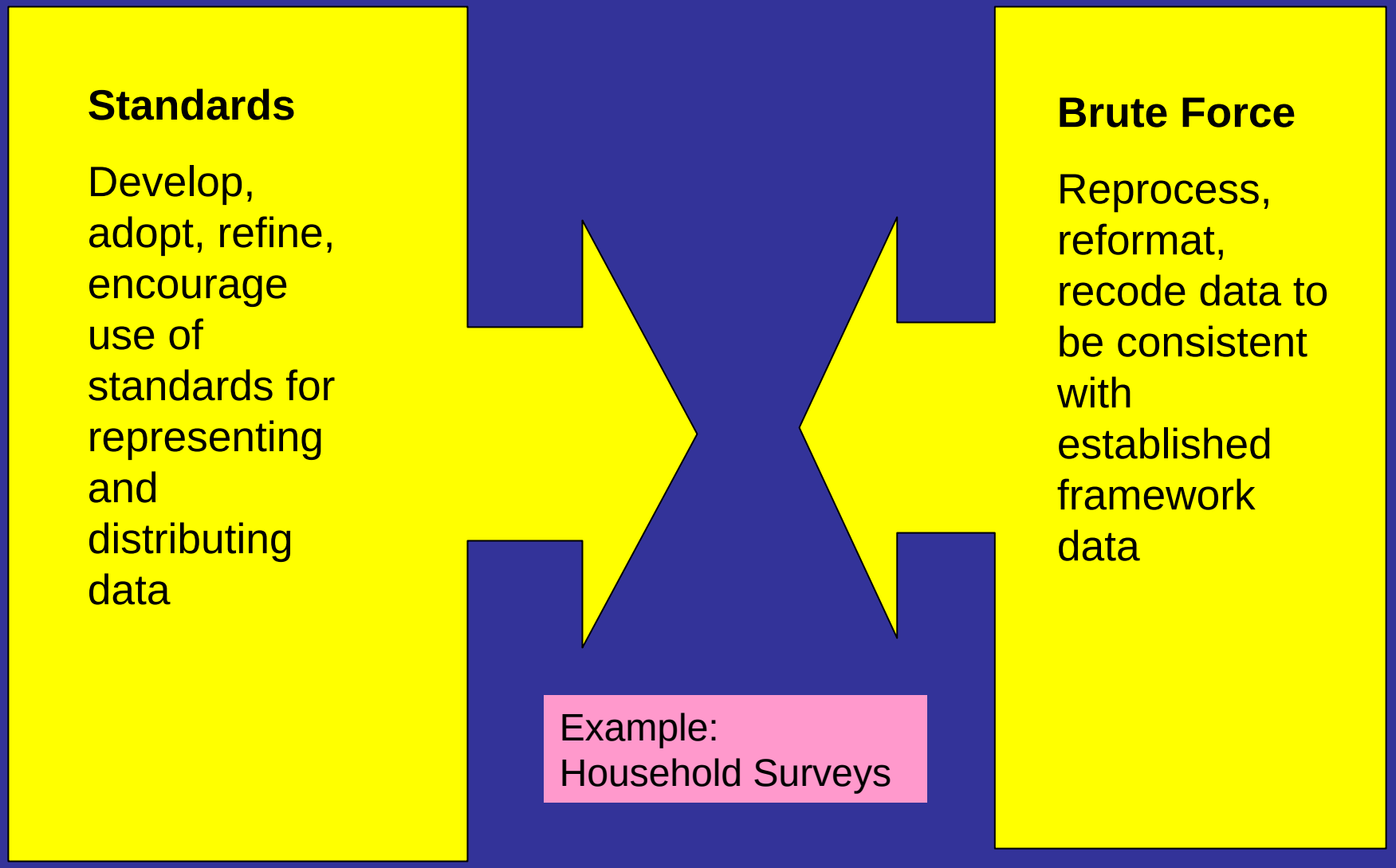
Persevere!

Challenge of model data



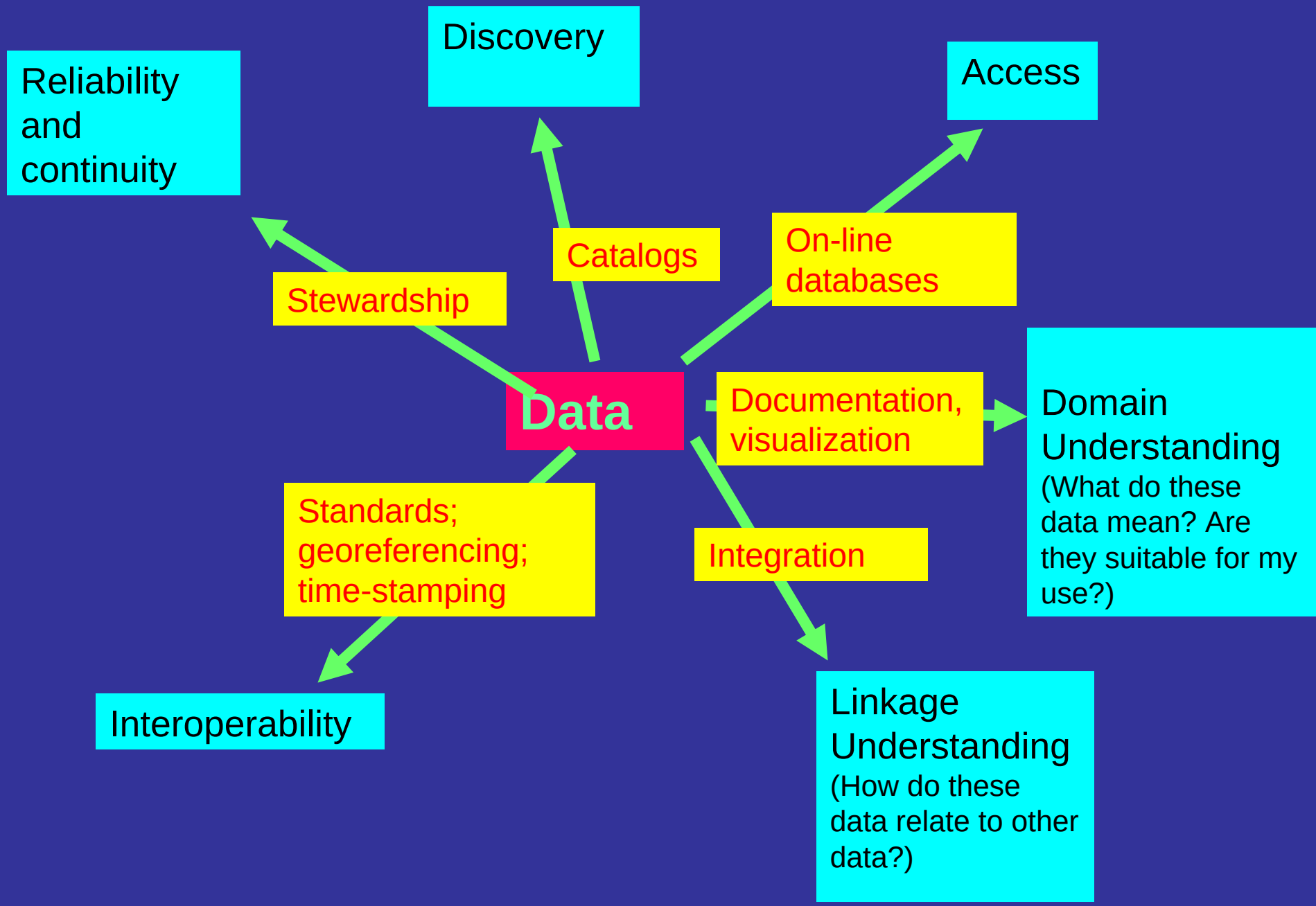
TOUGHEST NUT

Interoperability

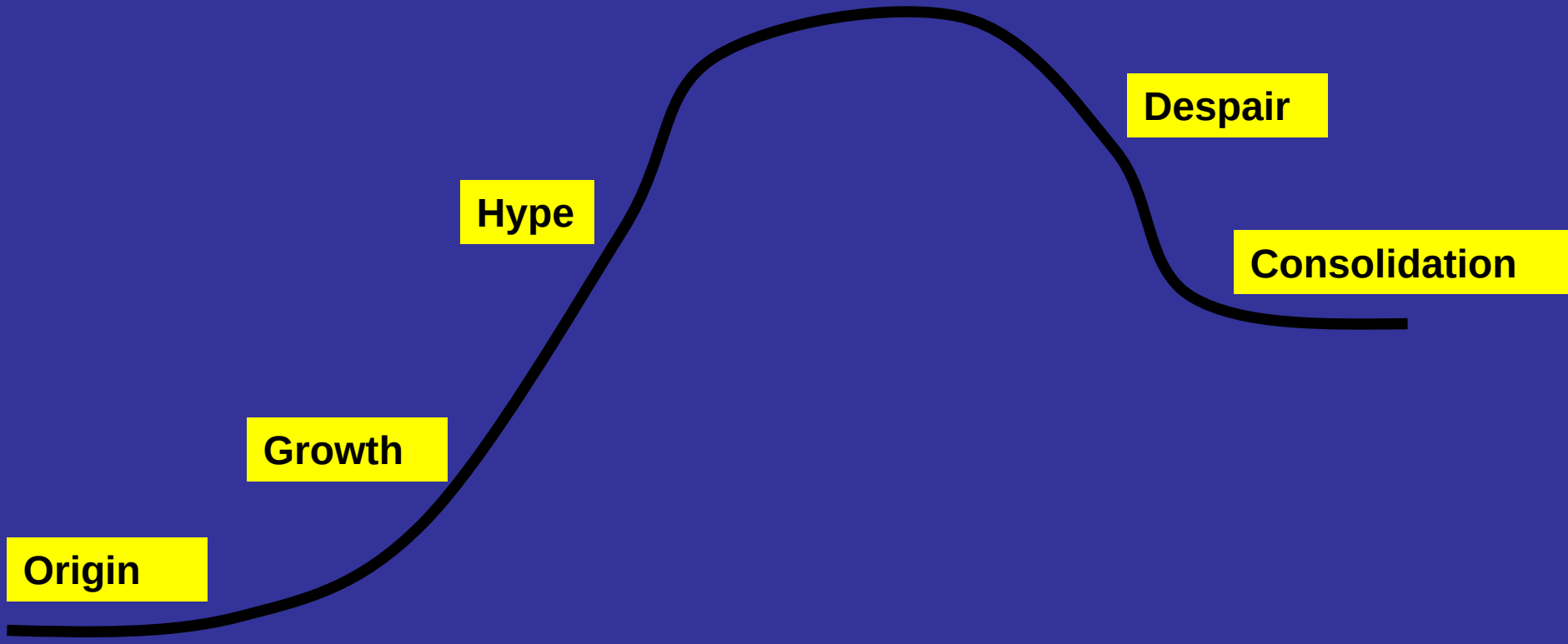


Stewardship

- Almost always under-provided
- Everyone underestimates the speed by which data becomes invisible or unintelligible
- Inter-disciplinary, problem-oriented data especially vulnerable


Reliability and continuity
Discovery
Access
Stewardship
Catalogs
On-line databases
Data
Documentation, visualization
Domain Understanding
(What do these data mean? Are they suitable for my use?)
Standards; georeferencing; time-stamping
Integration
Interoperability
Linkage Understanding
(How do these data relate to other data?)

Conclusions

- We don’t know how to do everything yet, but we know a lot more now than a decade ago
- The investments show positive economies of scale
  - each step forward getting the data questions right generates more research and policy return than previous steps
- But what remains is going to require sustained, focused effort
  - There’s a lot of hard stuff yet to do
- Historically, funders don’t like this kind of work
  - That seems to be changing

Plan for the issue-cycle