BADAN PUSAT STATISTIK

Pelopor
Data Statistik
Terpercaya
Untuk Semua

Additional Course Material

CONSTRUCTING COMPOSITE INDEX

Ema Tusianti

Statistics-Indonesia

Tirane, 3-4 September 2018

Related Concepts

1

Variable and Indicator

1. Variable: A characteristic, number, or quantity that increases or decreases over time, or takes different values in different situations.
2. Indicator: a quantitative or a qualitative measure derived from a series of observed facts that can reveal relative positions (e.g. of a country) in a given area.
3. A composite indicator or synthetic index is an aggregate of all dimensions, objectives, individual indicators and variables used. This implies that what formally defines a composite indicator is the set of properties underlying its aggregation convention.

Pros and Cons OF Composite Index

Pros:	Cons:
• Can summarise complex, multi-dimensional realities with a view to supporting decision-makers. • Are easier to interpret than a battery of many separate indicators. • Can assess progress of countries over time. • Reduce the visible size of a set of indicators without dropping the underlying information base. • Thus make it possible to include more information within the existing size limit. • Place issues of country performance and progress at the centre of the policy arena. • Facilitate communication with general public (*i.e.* citizens, media, *etc.*) and promote accountability. • Help to construct/underpin narratives for lay and literate audiences. • Enable users to compare complex dimensions effectively.	• May send misleading policy messages if poorly constructed or misinterpreted. • May invite simplistic policy conclusions. • May be misused, *e.g.* to support a desired policy, if the construction process is not transparent and/or lacks sound statistical or conceptual principles. • The selection of indicators and weights could be the subject of political dispute. • May disguise serious failings in some dimensions and increase the difficulty of identifying proper remedial action, if the construction process is not transparent. • May lead to inappropriate policies if dimensions of performance that are difficult to measure are ignored.

Step of Constructing Composite Index

2

Step of Constructing Composite Index

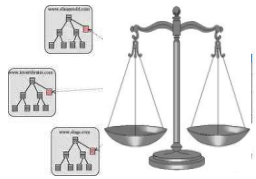
Theoretical framework

Data selection

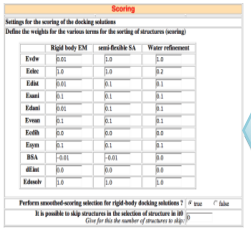
Indicator Structure

Imputation of missing data

Normalisation



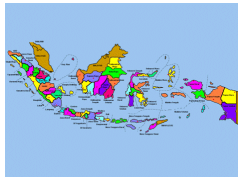
Weighting



Aggregation

Robustness and sensitivity

Back to the real data



Presentation and Visualisation

Step of Constructing Composite Index

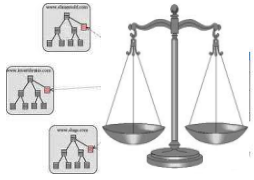
Theoretical framework

Data selection

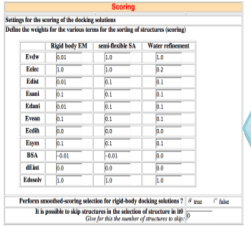
Indicator Structure

Imputation of missing data

Normalisation



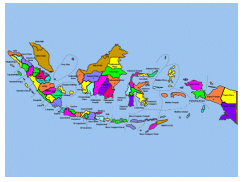
Weighting



Aggregation

Robustness and sensitivity

Back to the real data



Presentation and Visualisation

Step 1. Building Theoretical framework

1. To get a clear understanding and definition of the multidimensional phenomenon to be measured.
2. To structure the various sub-groups of the phenomenon (if needed).
3. To compile a list of selection criteria for the underlying variables, e.g., input, output, process.

Human Development Index

1. Measuring human quality
2. Three basic needs:
   - *A Long and Healthy Life*
   - *Knowledge*
   - *A Decent Standard of Living*

Step of Constructing Composite Index

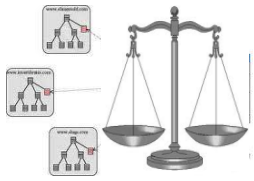
Theoretical framework

Data selection

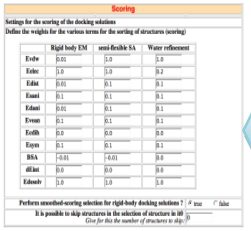
Indicator Structure

Imputation of missing data

Normalisation



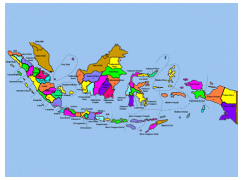
Weighting



Aggregation

Robustness and sensitivity

Back to the real data



Presentation and Visualisation

Step 2. Data selection

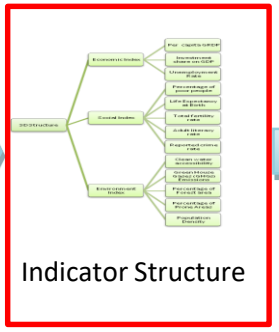
1. To check the quality of the available indicators.
2. To discuss the strengths and weaknesses of each selected indicator.
3. To create a summary table on data characteristics, e.g., availability (across country, time), source, type (hard, soft or input, output, process).

Indicator is identify by *brainstorming*

Step of Constructing Composite Index

Theoretical framework

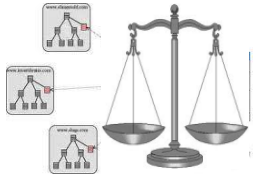
Data selection



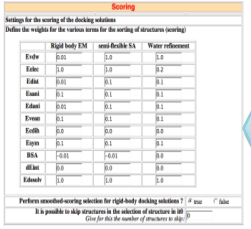
Indicator Structure

Imputation of missing data

Normalisation



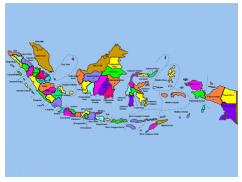
Weighting



Aggregation

Robustness and sensitivity

Back to the real data

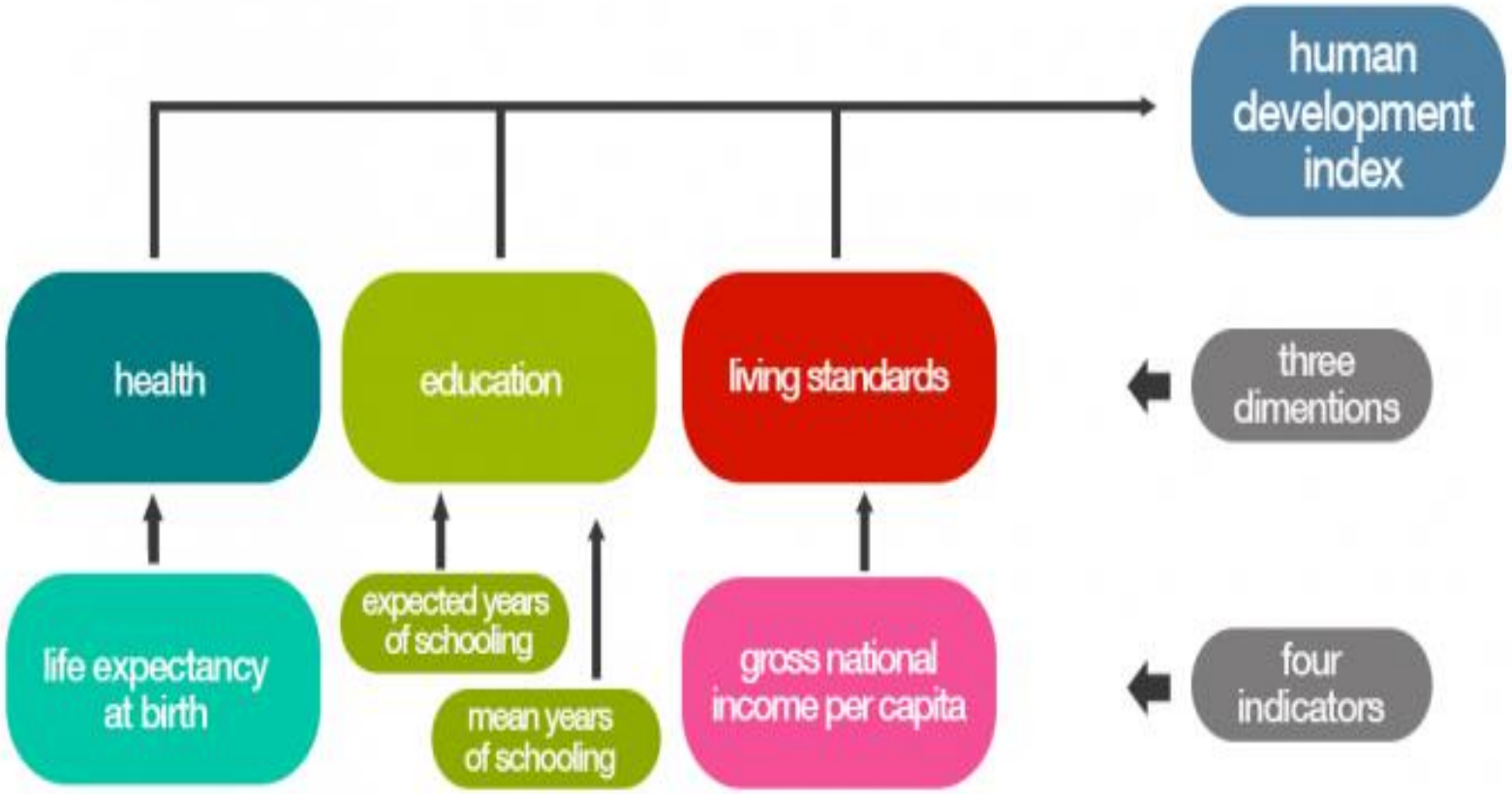


Presentation and Visualisation

Grouping Indicators

- The analyst must first decide whether the nested structure of the composite indicator is well defined and whether the set of available individual indicators is sufficient or appropriate to describe the phenomenon (see Step 2).
- This decision can be based on expert opinion and the statistical structure of the data set.
- Different analytical approaches, such as principal components analysis, can be used to explore whether the dimensions of the phenomenon are statistically well-balanced in the composite indicator. If not, a revision of the individual indicators might be necessary

Human development Index Structure



Step of Constructing Composite Index

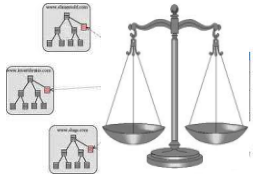
Theoretical framework

Data selection

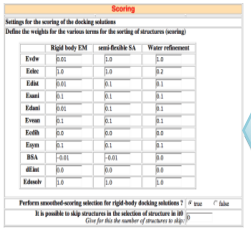
Indicator Structure

Imputation of missing data

Normalisation



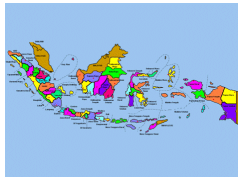
Weighting



Aggregation

Robustness and sensitivity

Back to the real data



Presentation and Visualisation

Step 4: Eliminating Missing Value

1. To estimate missing values
2. To provide a measure of the reliability of each imputed value, so as to assess the impact of the imputation on the composite indicator results.
3. To discuss the presence of outliers in the dataset.

Step of Constructing Composite Index

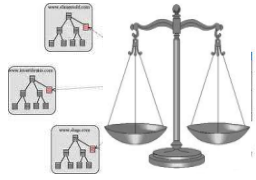
Theoretical framework

Data selection

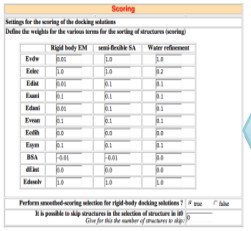
Indicator Structure

Imputation of missing data

Normalisation



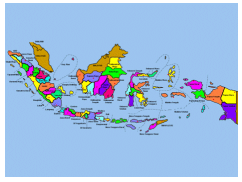
Weighting



Aggregation

Robustness and sensitivity

Back to the real data



Presentation and Visualisation

Step 5: Standardization

1. To select suitable normalization procedure(s) that respect both the theoretical framework and the data properties.
2. To discuss the presence of outliers in the dataset as they may become unintended benchmarks.
3. To make scale adjustments, if necessary.
4. To transform highly skewed indicators, if necessary.

- Ranking is the simplest normalisation technique. This method is not affected by outliers and allows the performance of countries to be followed over time in terms of relative positions
- Standardization (or z-scores) converts indicators to a common scale with a mean of zero and standard deviation of one. Indicators with extreme values thus have a greater effect on the composite indicator.
- Min-Max normalizes indicators to have an identical range [0, 1] by subtracting the minimum value and dividing by the range of the indicator values. However, extreme values/or outliers could distort the transformed indicator.

- Distance to a reference measures the relative position of a given indicator vis-à-vis a reference point. This could be a target to be reached in a given time frame. The reference could also be an external benchmark country.
- Categorical scale assigns a score for each indicator. Categories can be numerical, such as one, two or three stars, or qualitative, such as ‘fully achieved’, ‘partly achieved’ or ‘not achieved’.
- Indicators above or below the mean are transformed such that values around the mean receive 0, whereas those above/below a certain threshold receive 1 and -1 respectively. This normalization method is simple and is not affected by outliers.

Method	Equation
1. Ranking	$I_{qc}^{t} = Rank(x_{qc}^{t})$
2. Standardisation (or z-scores)	$I_{qc}^{t} = \dfrac{x_{qc}^{t} - x_{qc=\bar{c}}^{t}}{\sigma_{qc=\bar{c}}^{t}}$
3. Min-Max	$I_{qc}^{t} = \dfrac{x_{qc}^{t} - \min_{c}(x_{q}^{t_0})}{\max_{c}(x_{q}^{t_0}) - \min_{c}(x_{q}^{t_0})}.$
4. Distance to a reference country	$I_{qc}^{t} = \dfrac{x_{qc}^{t}}{x_{qc=\bar{c}}^{t_0}}$ or $I_{qc}^{t} = \dfrac{x_{qc}^{t} - x_{qc=\bar{c}}^{t_0}}{x_{qc=\bar{c}}^{t_0}}$
5. Categorical scales	Example: $I_{qc}^{t} = \begin{cases} 0 & \text{if } x_{qc}^{t} < P^{15} \\ 20 & \text{if } P^{15} \le x_{qc}^{t} < P^{25} \\ 40 & \text{if } P^{25} \le x_{qc}^{t} < P^{65} \\ 60 & \text{if } P^{65} \le x_{qc}^{t} < P^{85} \\ 80 & \text{if } P^{85} \le x_{qc}^{t} < P^{95} \\ 100 & \text{if } P^{95} \le x_{qc}^{t} \end{cases}$
6. Indicators above or below the mean	$I_{qc}^{t} = \begin{cases} 1 & \text{if } w > (1+p) \\ 0 & \text{if } (1-p) \le w \le (1+p) \\ -1 & \text{if } w < (1-p) \end{cases}$ where $w = x_{qc}^{t} / x_{qc=\bar{c}}^{t_0}$

Maximum-Minimum Method on Indonesian HDI

❖ $I_{health} = \frac{e0 - e0\ _{min}}{e0\ _{maks} - e0\ _{min}}$

❖ $I_{education} = \frac{I_{EYS} + I_{MYS}}{2}$

- $I_{EYS} = \frac{EYS - EYS_{min}}{EYS_{maks} - EYS_{min}}$
- $I_{MYS} = \frac{MYS - MYS_{min}}{MYS_{maks} - MYS_{min}}$

❖ $I_{expenditure} = \frac{\ln(expenditure) - \ln(expenditure_{min})}{\ln(expenditure_{maks}) - \ln(expenditure_{min})}$

Maximum-Minimum Method on Indonesian HDI

Indicator	unit	Minimum		Maximum	
		UNDP	BPS	UNDP	BPS
Life expectancy (e0)	year	20	20	85	85
Expected year of Schooling (EYS)	year	0	0	18	18
Mean Year of Schooling (MYS)	year	0	0	15	15
Expenditure		100 (PPP U$)	1.007.436* (Rp)	107.721 (PPP U$)	26.572.352 ** (Rp)

Step of Constructing Composite Index

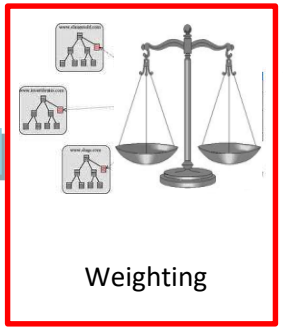
Theoretical framework

Data selection

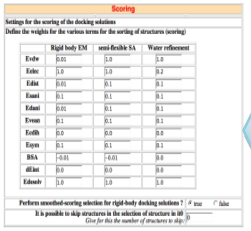
Indicator Structure

Imputation of missing data

Normalisation



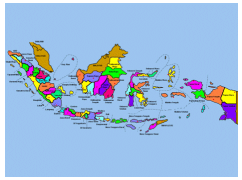
Weighting



Aggregation

Robustness and sensitivity

Back to the real data



Presentation and Visualisation

Step 6: Weighting

1. The relative importance of the indicators is a source of contention
2. Selected the appropriate weighting and aggregation procedure(s) with reference to the theoretical framework.
3. Considered the possibility of using alternative methods (multi-modelling principle).

Equal & Unequal Weighting

- equal weighting (EW), i.e. all variables are given the same weight. This essentially implies that all variables are “worth” the same in the composite
- Unequal; A number of weighting techniques exist. Some are derived from statistical models, such as factor analysis, data envelopment analysis and unobserved components models (UCM), or from participatory methods like budget allocation processes (BAP), analytic hierarchy processes (AHP) and conjoint analysis (CA).

Step of Constructing Composite Index

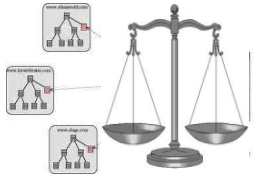
Theoretical framework

Data selection

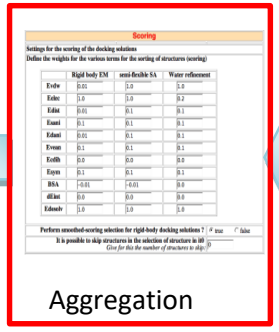
Indicator Structure

Imputation of missing data

Normalisation



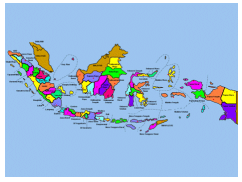
Weighting



Aggregation

Robustness and sensitivity

Back to the real data



Presentation and Visualisation

Step 7: Aggregation

1. linear aggregation method is useful when all individual indicators have the same measurement unit, provided that some mathematical properties are respected.
2. Geometric aggregations are better suited if the modeler wants some degree of non compensability between individual indicators or dimensions.
3. Linear aggregations reward base-indicators proportionally to the weights, while geometric aggregations reward those countries with higher scores.

Aggregation Method

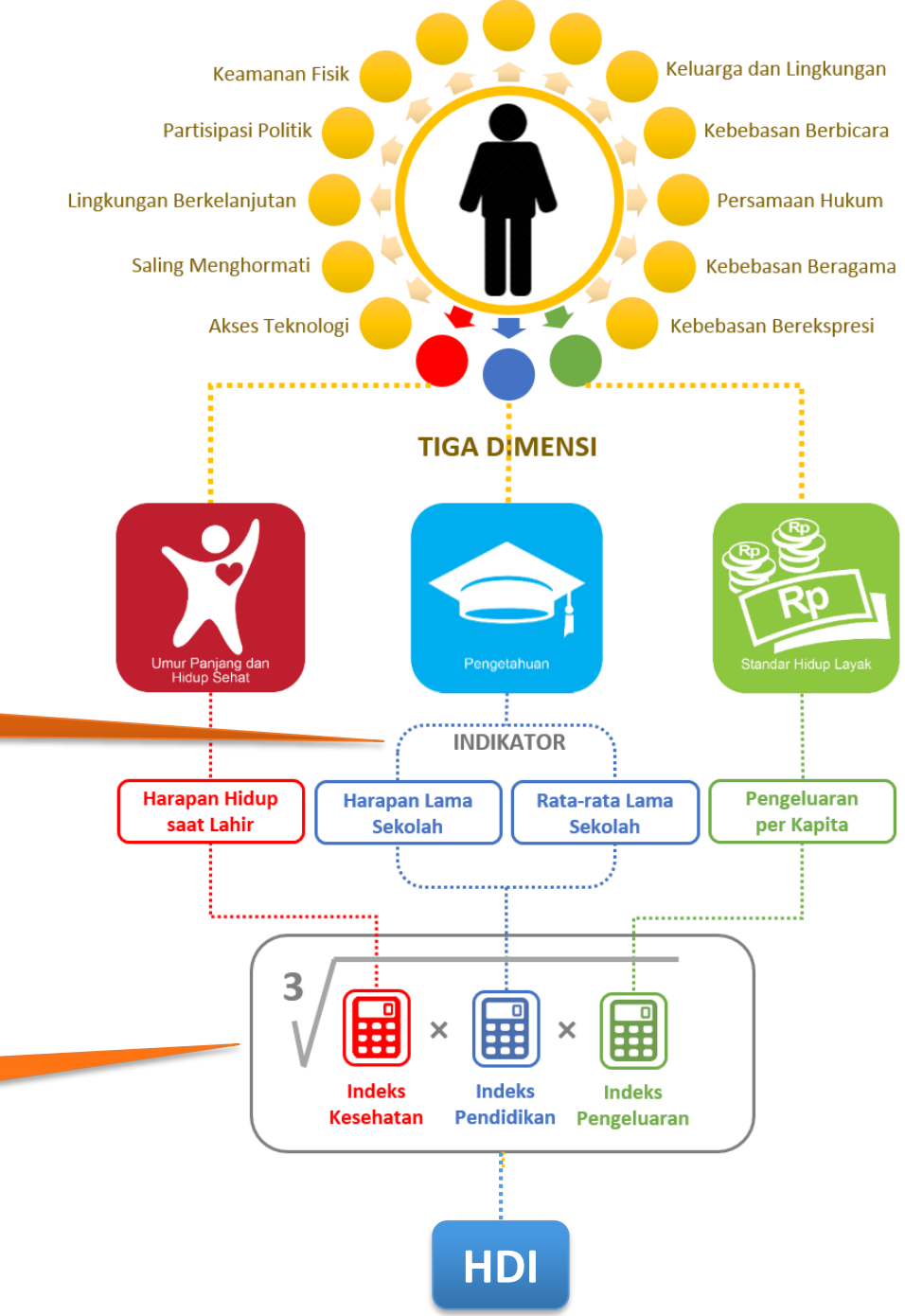
Linear Aggregation

$$CI_c = \sum_{q=1}^{Q} w_q I_{qc}$$

Geometric Aggregation

$$CI_c = \prod_{q=1}^{Q} x_{q,c}^{w_q}$$

with $\sum_q w_q = 1$ and $0 \leq w_q \leq 1$, for all $q=1,..,Q$ and $c=1,...,M$.



Linear Aggregation on education index

Geometric aggregation on final Index

Step of Constructing Composite Index

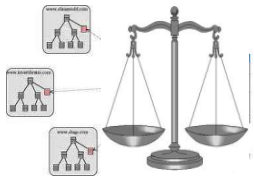
Theoretical framework

Data selection

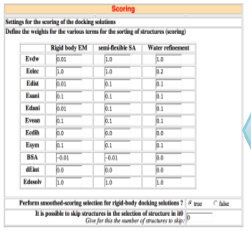
Indicator Structure

Imputation of missing data

Normalisation



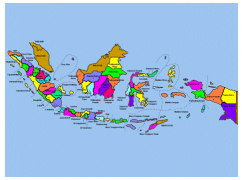
Weighting



Aggregation

Robustness and sensitivity

Back to the real data



Presentation and Visualisation

Step 8: Sensitivity and Robustness

1. To consider a multi-modelling approach to build the composite indicator, and if available, alternative conceptual scenarios for the selection of the underlying indicators.
2. To identify all possible sources of uncertainty in the development of the composite indicator and accompany the composite scores and ranks with uncertainty bounds.
3. To conduct sensitivity analysis of the inference (assumptions) and determine what sources of uncertainty are more influential in the scores and/or ranks

Step of Constructing Composite Index

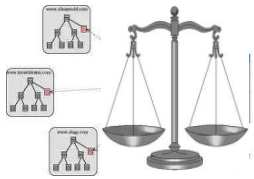
Theoretical framework

Data selection

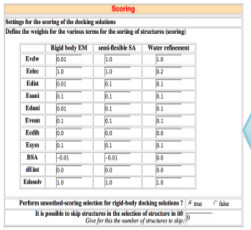
Indicator Structure

Imputation of missing data

Normalisation



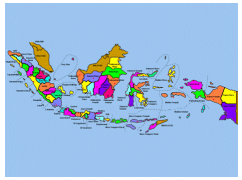
Weighting



Aggregation

Robustness and sensitivity

Back to the real data



Presentation and Visualisation

Step 9: Back to Detail

1. To profile country performance at the indicator level so as to reveal what is driving the composite indicator results.
2. To check for correlation and causality (if possible).
3. To identify if the composite indicator results are overly dominated by few indicators and to explain the relative importance of the sub-components of the composite indicator.

Step of Constructing Composite Index

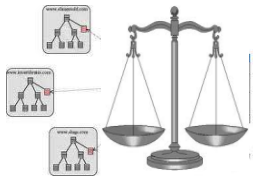
Theoretical framework

Data selection

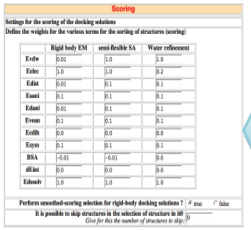
Indicator Structure

Imputation of missing data

Normalisation



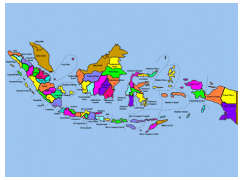
Weighting



Aggregation

Robustness and sensitivity

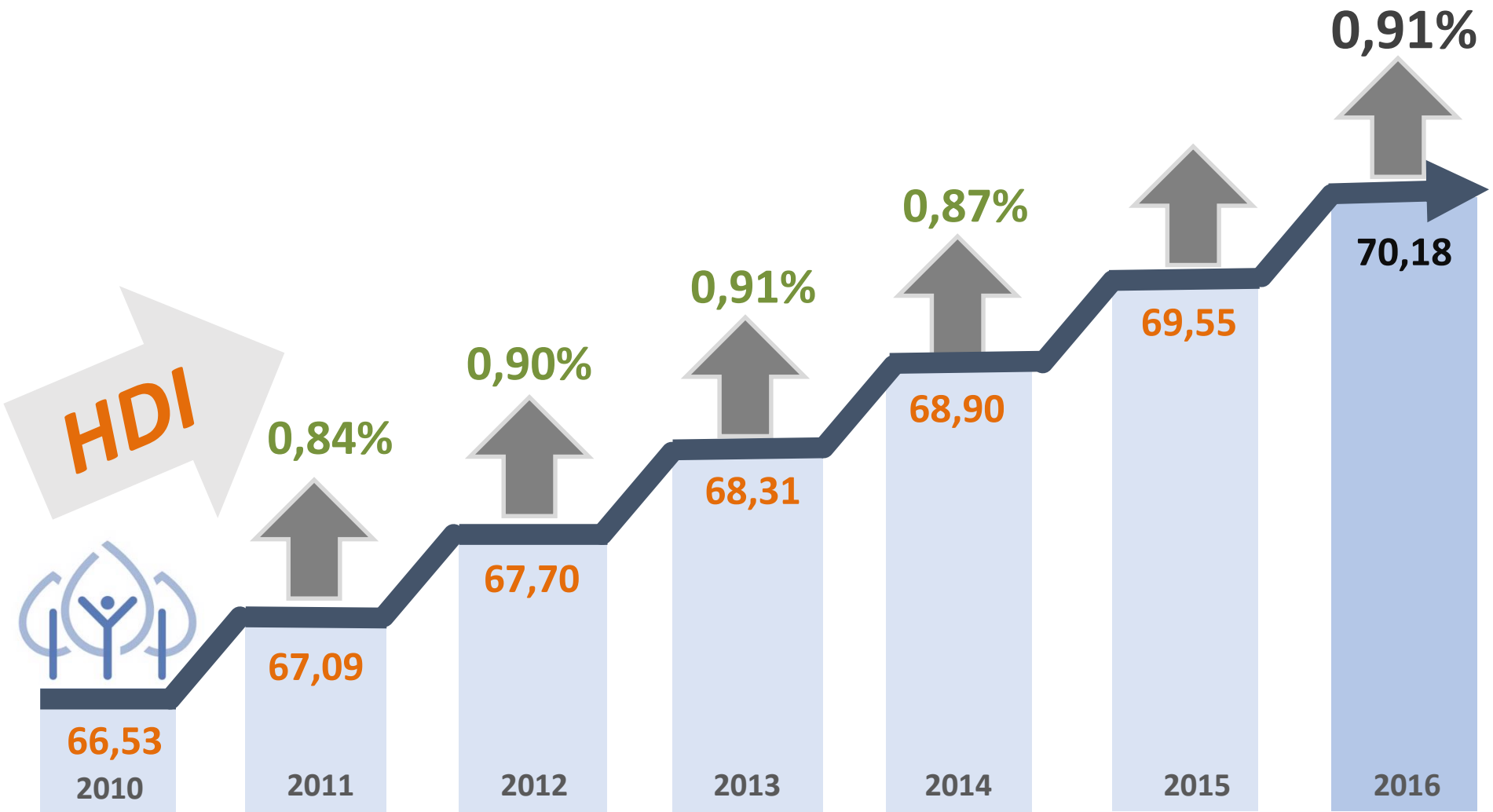
Back to the real data



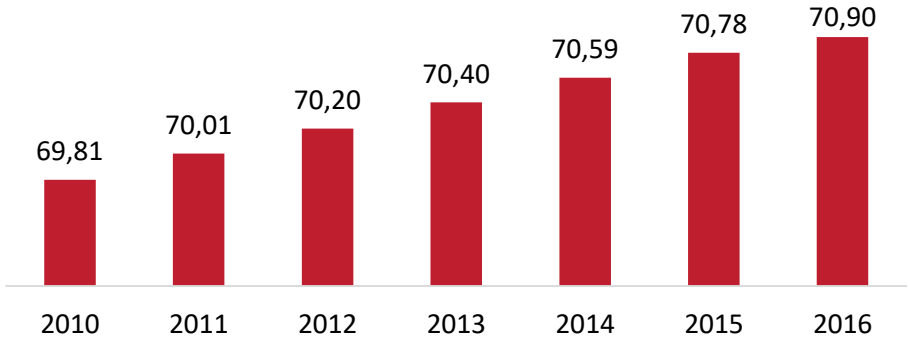
Presentation and Visualisation

Step 10: Dissemination

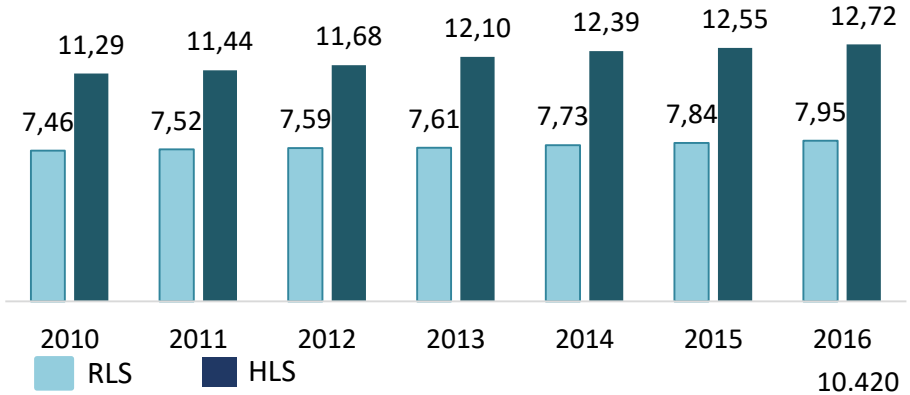
1. To identify a coherent set of presentational tools for the targeted audience.
2. To select the visualisation technique which communicates the most information.
3. To present the composite indicator results in a clear and accurate manner.


HDI
66,53
2010
0,84%
67,09
2011
0,90%
67,70
2012
0,91%
68,31
2013
0,87%
68,90
2014
69,55
2015
0,91%
70,18
2016

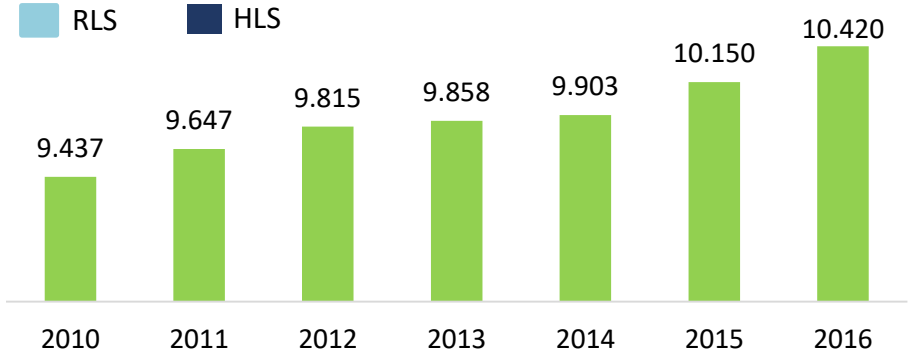
Life expectancy



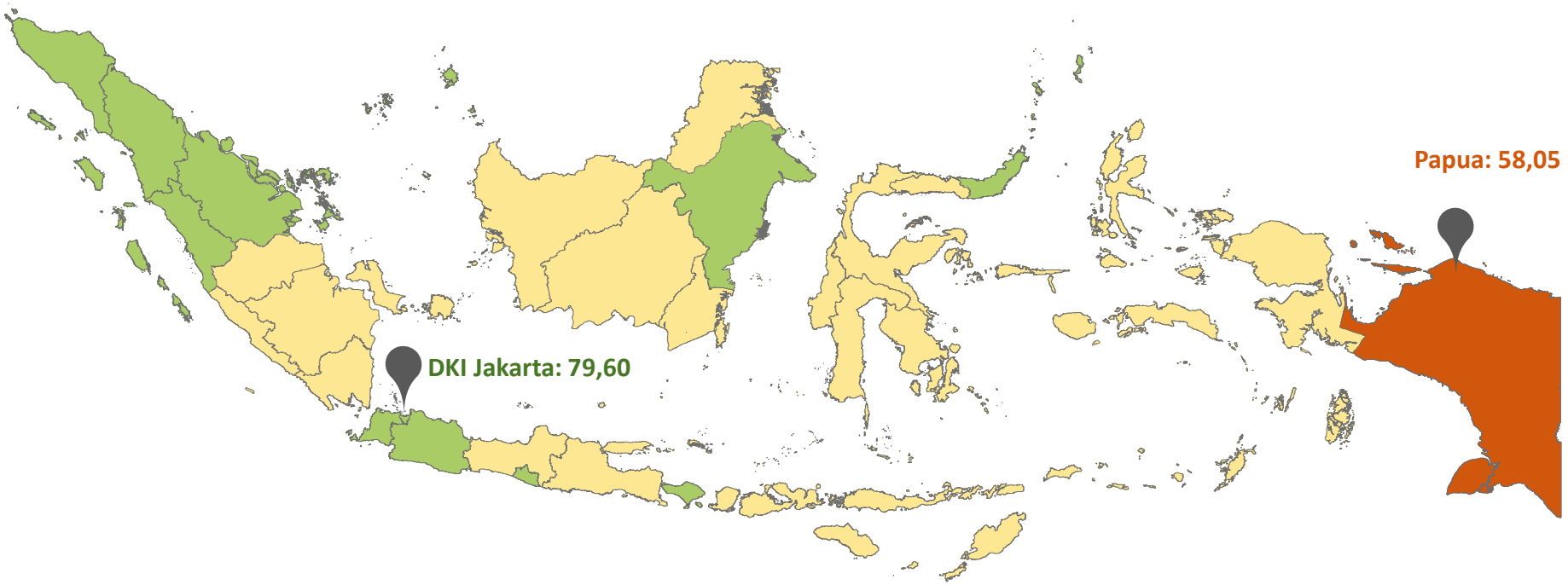
- Expected year of Schooling
- Mean Year of Schooling



Expenditure



Top Mover Province



Very High (HDI ≥ 80)

- PROV

No province in this category

Moderate (60 ≤ HDI < 70)

21 PROV

Excluding Aceh, Sumut, Sumbar, Riau, Kep. Riau, DKI Jakarta, Jabar, DI Yogyakarta, Banten, Bali, Kaltim, Sulut,

High (70 ≤ HDI < 80)

12 PROV

Aceh, Sumut, Sumbar, Riau, Kep. Riau, Jakarta, Jabar, DI Yogyakarta, Banten, Bali, Kaltim, Sulawesi Utara

Low (HDI < 60)

1 PROV

Papua

Application in Indonesia

3

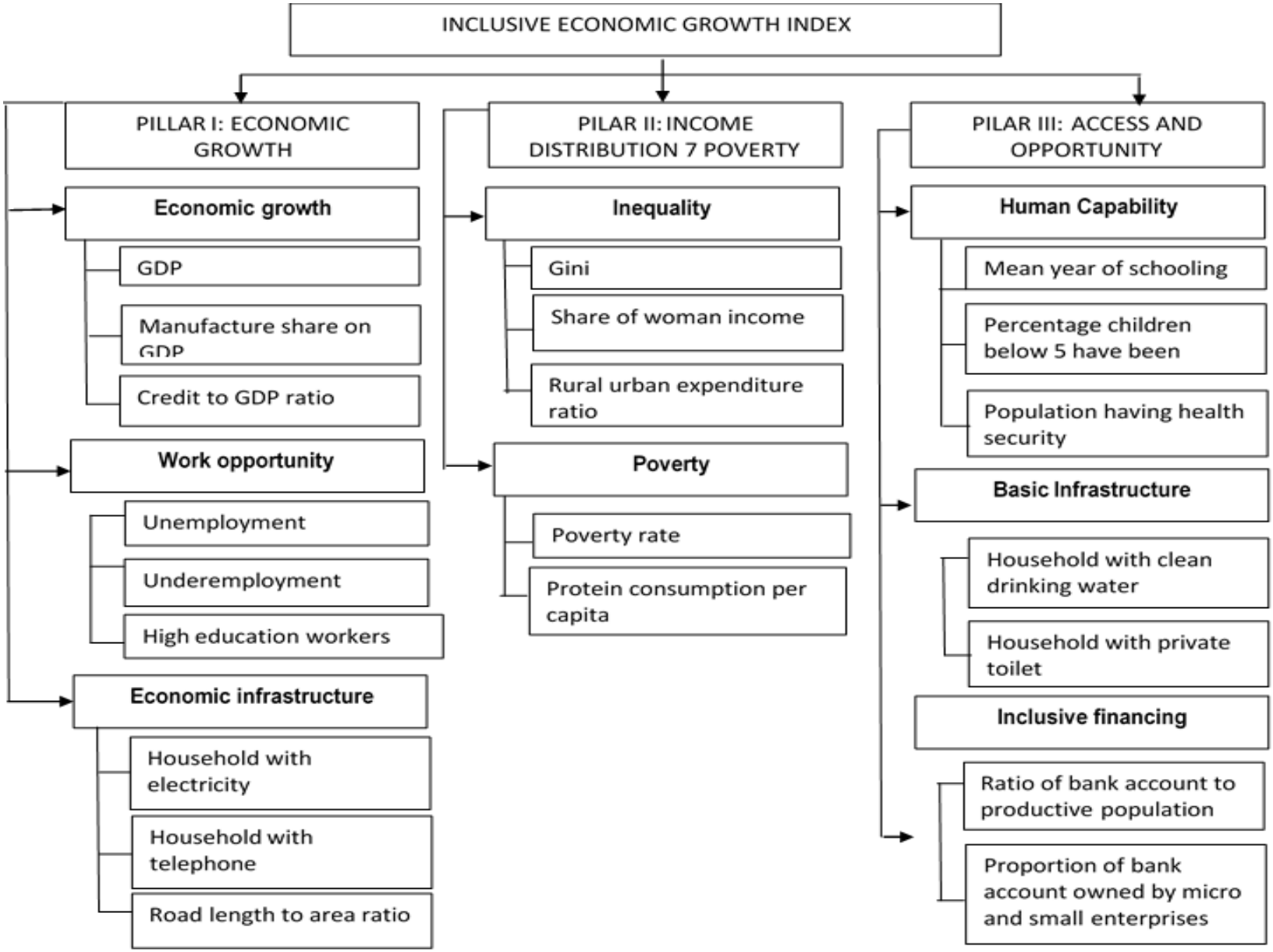
Welfare and Sustainable Development Measures

- Multidimensional poverty
- Inclusive growth
- Gender Inequality Measures
- Sustainable Development Index (thesis project)

Multidimensional Poverty Index

Dimension	Indicator	Poor definition	Weight
1. Living standard	Asset ownership (TV, car, motorbike, telephone, refrigerator, boat)	Poor: owning at least one of the mentioned assets but not car or boat Non Poor: if owning car or boat or having 2 or more other assets	1/18
	Housing floor	Poor: ground	1/18
	Electricity	Poor: no electricity	1/18
	Cooking energy sources	Poor: Firewood, coal, charcoal	1/18
	Toilet	Poor: no private toilet	1/18
	Drinking water	Poor: no clean water source	1/18
2. Education	Education attainment	Poor: no one educated (at least primary school)	1/6
	School participation	Poor: if family members below 15 years are not participated in the formal/informal school	1/6
3. Health	Children below 5y mortality case	Poor: If the household ever had a children died aged below 5	1/6
	Nutrition	Poor: if family member (at least 1) experienced malnutrition	1/6

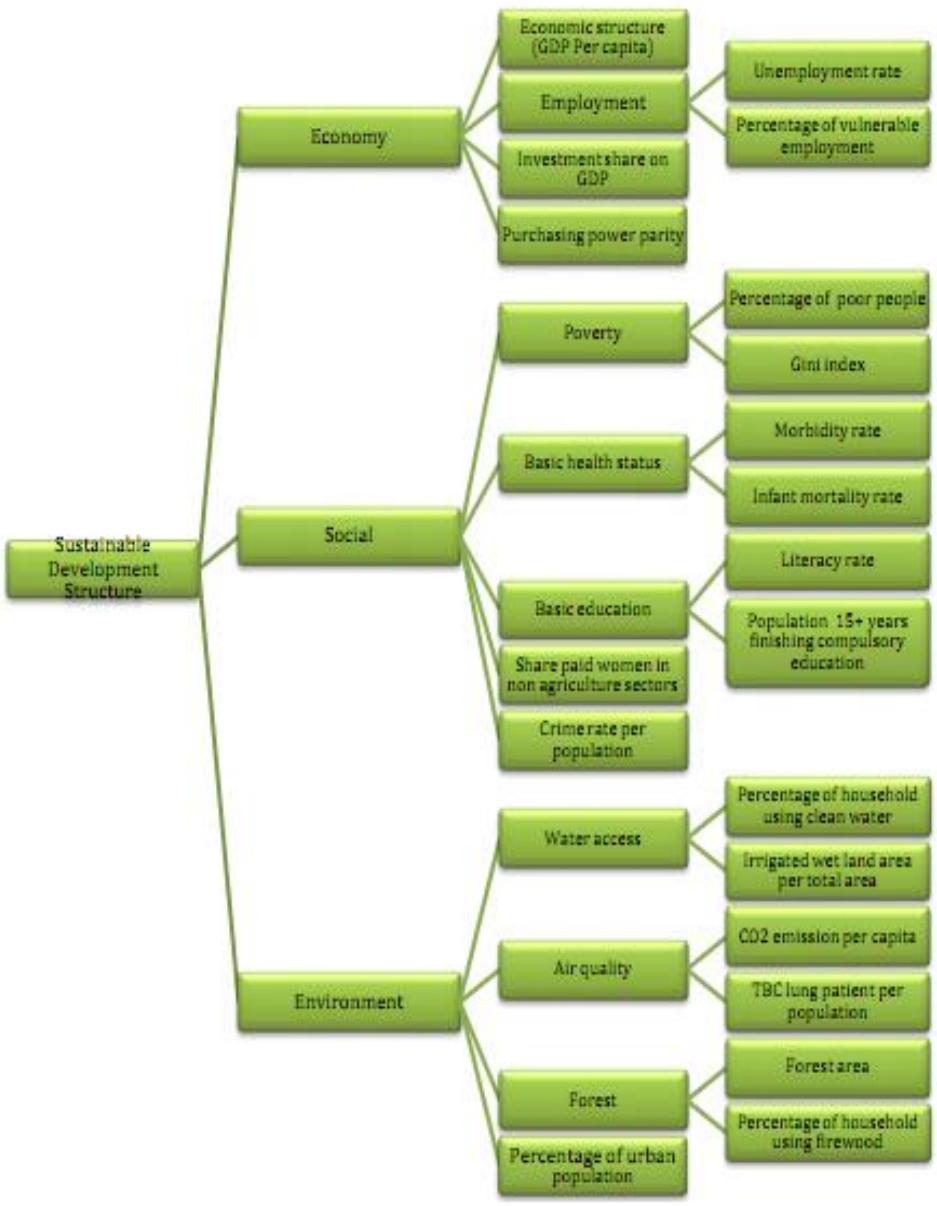
Inclusive Growth Index

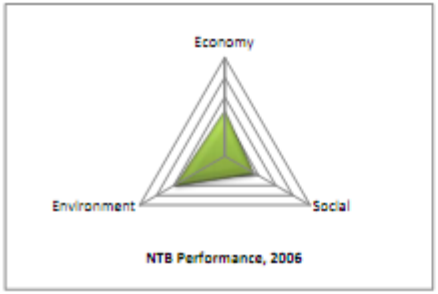


Gender Inequality Measures

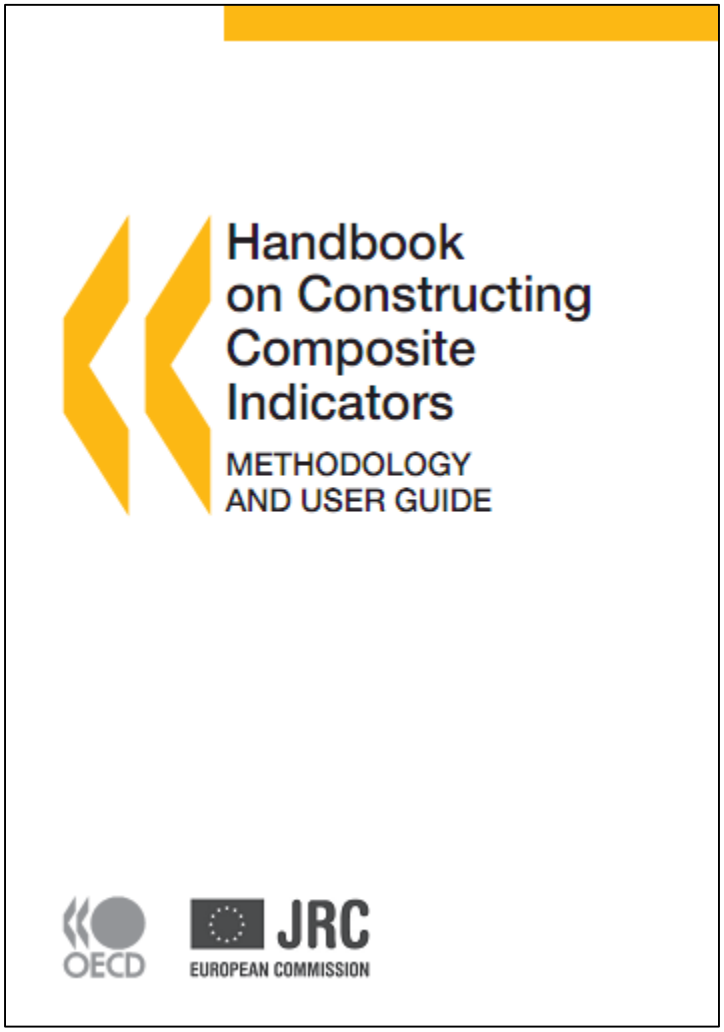
GEM	Empowerment	Economic participation and decision-making power	Female legislators, senior officials and managers
			Female professional and technical workers
		Political participation	Women's shares of parliamentary seats
		Power over economic resources	Ratio of female to male estimated earned income (PPP$)
GII	Empowerment	Reproductive health	Maternal mortality ratio
			Adolescent birth rate
		Empowerment	Female and male shares of parliamentary seats
			Female and male shares of population with at least secondary education
		Labour market	Female and male labour force participation rates
nGDI	Human Development	Health	Life expectancy at birth
		Knowledge	Mean years of schooling
			Expected years of schooling
		Standard of living	Estimated GNI per capita (PPP$)

Sustainable Development Index





Main Reference



BADAN PUSAT STATISTIK

Pelopor
Data Statistik
Terpercaya
Untuk Semua

THANK YOU