

Overview of standards for data disaggregation

Read me first:

This document gives an overview of possible and existing standards, thoughts and ideas on data disaggregation, as well as questions arising during the work on this document, in order to facilitate the discussion in the work stream. Please note, that this document only refers to the disaggregation dimensions stated in §74 (g) of the Resolution 70/1, as the further disaggregation dimensions are mainly demanded in just one or two indicators/ targets and thus need no common treatment at this moment.

The first chart is a collection of standards already in use for presenting disaggregated data (this document only contains the European perspective, probably you and your colleagues could complete the list).

The other table includes some thoughts, ideas and questions on how to proceed with the data disaggregation for the SDG-Indicators.

Dimensions	Different categories already in use	Information/ Evaluation	Categories used in the Global Indicator Framework	Existing global standards	Existing regional standards	Comments
Income/economic status/ poor and vulnerable	Income per capita Income quintiles DHS Wealth Index (wealth quintiles) Multidimensional Poverty Index Unsatisfied Basic Needs - Deprivation	No single standard measure available; measured in income, economic status, poverty or wealth and in relative as well as absolute numbers Usage of small area estimates in poverty/ income mapping (e.g. methodology used in the Poverty Atlas by the World Bank) combines disaggregation of income/ poverty and geographical location	Wealth: Low to high socio-economic parity status index Income: Growth rate of income for bottom 40% and total	Rio Group on Poverty Statistics, last meeting in 2006, no standard developed Canberra Group on Household Income Statistics: no definitive set of standards, presentation of all relevant issues Poverty Mapping (Poverty mapping group of the World Bank) UNDP: Multidimensional Poverty Index UN Handbook on Poverty Statistics <i>Headcount measure</i> <i>Poverty gap</i> <i>Watts index</i> <i>Squared poverty gap</i>	EU-Standard: Income: income quintiles (1 st , 2 nd , 3 rd , 4 th , 5 th) Poverty: 3 dimensions in Europe 2020 strategy target on the risk of poverty and social exclusion <ul style="list-style-type: none"> Monetary poverty Severe material deprivation Very low work intensity EU-SILC: <ul style="list-style-type: none"> Net equivalent income (median) At-risk of poverty rate 	
Sex	Gender and Agriculture Research Network (CGIAR): Standards for Collecting Sex Disaggregated Data Demographic and Health Survey (DHS): woman's/ male questionnaire in households	CGIAR provides intern guide with must haves for sex/ gender analysis; might be too comprehensive for the inclusion in household surveys with regard to the SDGs monitoring DHS provides sex disaggregated data mainly for 15-49 year-olds; could be limited by small sample sizes	Female, male, both gender parity indice	Female, male	EU-Standard: Female, male	
Age	Date of Birth Age groups 1-year-age-groups	Use of different age groups in national and international data Differing age groups demanded in indicator or target	Differing age groups: Commonly used categories 15-49, <15, 15-49, >15 15-65 <5	UN definition of age groups: Infants: 0-5 years Children: 0-15 years Youth: 5-24 years, (UN Youth) Adults 15 years and older; Older Persons: 60 years and older (DSPD: Focal Point for Ageing)	EU-Standard: differing age groups Often 10 year intervals are used e.g. in the EU SDI database	Canada: Suggest age grouping rather than single year age groups whenever possible. We suggest that 5 year intervals is the lowest level of disaggregation for age.

Dimensions	Different categories already in use	Information/ Evaluation	Categories used in the Global Indicator Framework	Existing global standards	Existing regional standards	Comments
				UN Principles and Recommendations for a Vital Statistics System (Rev.3): Infants: <1 year Pre-school age: 1-4 years School age: 5-14 years Childbearing age: 15-49 years Working ages: 15-64 years Elderly persons: 65 years and older		
Race	<ul style="list-style-type: none"> • Colour 	Caution: different connotation of race Disaggregation categories could offend certain population groups	Data is not disaggregated by race		EU-Standard: SDG data is not disaggregated by race	Canada: Not available in Canada and other countries may not allow the collection of data based upon race.
Ethnicity	<ul style="list-style-type: none"> • Ethnic ancestry or origin • Ethnic identity • Cultural origins • Race • Minority status • Tribe • Language • Religion • Ethnic Self-identification • Recognised (national) minorities 	UN Concepts and definitions: “[...] By the nature of this topic, these categories and their definitions will vary widely from country to country; therefore, no internationally accepted criteria are possible.” UN Standards and Methods: „Ethnicity is multidimensional and is more a process than a static concept, and so ethnic classification should be treated with movable boundaries“ Caution: different connotation of origin and tribe Disaggregation categories could offend certain population groups	Data is not disaggregated by ethnicity	No international standard possible due to varying national circumstances	EU-Standard: SDG data is not disaggregated by ethnicity Country/type of citizenship	
Migration status	Country of Birth Country of Citizenship (Legal Status?)	UN recommendation: Country of Birth (native or foreign-born), Country of Citizenship(foreign citizen), Year of arrival in country of enumeration (to measure length of stay), also relevant if national	Data is not disaggregated by migration status	SDG data is not disaggregated by migration status Migration: Country of Birth Country of Citizenship Year of arrival in country of enumeration	EU-Standard: SDG data is not disaggregated by migration status Immigrant measurement by <ul style="list-style-type: none"> • Country of citizenship • Country of birth 	

Dimensions	Different categories already in use	Information/ Evaluation	Categories used in the Global Indicator Framework	Existing global standards	Existing regional standards	Comments
		<p>boundaries change over time</p> <p>Proposed coding of country of birth: Numerical coding system of Standard Country or Area Codes for Statistical Use</p>		<p>Refugees: UNHCRR standard</p> <ul style="list-style-type: none"> • Refugees (incl. refugee-like situations) • Asylum-seekers (pending cases) • Returned refugees • Internally displaced persons (IDPs) • Returned IDPs • Stateless persons • Others of concern <p>The Expert Group on Refugee and IDP Statistics is developing a set of international recommendations for refugee statistics and a refugee statistics compiler manual with operational instructions. Guidelines on refugee statistics will be presented at the 49th UNSC session in 2018</p>	<ul style="list-style-type: none"> • Country of previous residence <p>Emigrant measurement by</p> <ul style="list-style-type: none"> • Country of citizenship • Country of birth • Country of next residence 	
Disability	<p>Washington Group (WG) short set of questions on disability</p> <p>UNICEF/Washington Group module on Child Functioning</p> <p>International Classification of Functioning, Disability and Health (ICF)</p> <p>International Classification of Diseases (ICD)</p>	<p>Washington Group's sets of questions are proposed as standard for the monitoring of the SDGs by the United Nations Expert Group Meeting on Disability Data and Statistics, Monitoring and Evaluation</p> <p>ICF and ICD are rather classifications than standards</p>	<p>Disability: Severe disabilities collecting disability social protection benefits</p>	<p>International Classification of Functioning, Disability and Health, (ICF) Custodian: WHO</p> <p>Washington Group on Disability Statistics</p>	<p>EU-Standard: In SDG data: Type of disability measured by level of activity limitation</p> <ul style="list-style-type: none"> - None - Some or severe <p>EU Labour Force Survey: Type of disability:</p> <ul style="list-style-type: none"> - Difficulty in basic activity - No difficulty in basic activity - Limitation in work caused by a health condition or difficulty in basic activity - No limitation in work caused by a health 	

<p>General remarks and questions</p>	<p>In general there are two possibilities for data collection</p> <ul style="list-style-type: none"> • Combination of different survey and register data In order to improve data disaggregation and to allow for cross-analysis of different disaggregation dimensions, register-based data is necessary. Identifying a person who is for example female, poor, with migration status and in a certain geographical location would require using a unique identifier across the registers and possibly surveys. This is not always available and the access to registers might imply legal problems and obstacles. • Household surveys Current discussions on standards for data disaggregation include the idea to implement question sets for specific disaggregation dimensions, e.g. the Washington Group's question set on disability. The question set would have to be implemented in the national data collection process. This could result in a large amount of question sets resulting in a duplication of work among the processes of data collection and analysis. 	<ul style="list-style-type: none"> • Q1: How should we deal with register data? • Q2: How to deal with specific question sets for disaggregation dimensions? Should there be several individual question sets for disaggregation dimensions? • Q3: These are questions concerning the technical implementation of the data disaggregation, referring to Workflow C of the Data Disaggregation Plan. It could be a good idea to set up a taskforce to deal with technical and methodological questions. Who is interested in initiating and/ or participating in this task force?
	<p style="text-align: center;">Comments</p> <p style="text-align: center;">...</p>	<p style="text-align: center;">Comments</p> <ul style="list-style-type: none"> • Sweden: Q1: the disaggregations will need to be done differently in different regions and for different indicators as the possibilities for accessing disaggregated data are vastly different. Research studies that make analyses of particular questions are probably necessary before pilot statistics can be set up.
<p>Income/economic status/ poor and vulnerable</p>	<ul style="list-style-type: none"> • The concept of purchasing power parity could be one option to disaggregate by income. However, the question remains how income should be measured in detail. • A further idea is to disaggregate by the poor and vulnerable by means of income measurement. The poverty line could be determined by regional/ national standards or national poverty lines. A further option is the use of the World Bank's international poverty line (1.90 US \$ per day). In the next step, the persons identified as poor according to income measurement could be further disaggregated by relevant dimensions relating to vulnerability. 	<ul style="list-style-type: none"> • Q4: Should income be measured in absolute values, in quintiles, in steps monetary units etc.?
	<p style="text-align: center;">Comments</p> <ul style="list-style-type: none"> • Canada: For global - It seems better to examine income in quintiles within the country. It would provide a relative sense of vulnerable 	<p style="text-align: center;">Comments</p> <p style="text-align: center;">...</p>
<p>Sex</p>	<ul style="list-style-type: none"> • We propose to focus on the dimension sex. Consequently gender would not be considered in the disaggregation. • We suggest the use of the categories female and male for the dimension „sex“. • It could be considered that indicators and targets that specifically refer to women only (e.g. 5.1 -5.5) are partly collected for men as well, to enable comparisons by sex. (e.g. 5.5.2 Proportion of women in managerial 	<p style="text-align: center;">/</p>

	positions)	
	Comments	Comments
	<ul style="list-style-type: none"> • Canada: Agree that at this time it is not possible to disaggregate beyond the sex dimension. • Sweden: To start with, a division into sex is a good first step, and to also collect data for men for some of the inequality indicators is also in line with good statistical practice. Some issues concern the LGBT community that often is a group that is among those 'left behind' and so merit some representation in the follow up. Possibly this follow up can be largely focussed on policy or legal systems and thus avoiding registration of a vulnerable group. Other special studies might be possible outside of the indicator system conducted by NGOs or by statistics from the health care system.... 	...
Age	<ul style="list-style-type: none"> • As the SDGs indicators and targets refer to specific and context based different age groups, like e.g. newborns, children or older people, different classifications of age groups are required. • If available, data disaggregation by age could be implemented in subject related contexts. For certain aspects (e.g. elections, tobacco or alcohol consume) age groups could be determined on national level. • It is recommended not to truncate age reporting over a certain age (e.g. 55 or 65 years), due to increasing longevity and heterogeneity among elderly population. 	/
	Comments	Comments
	<ul style="list-style-type: none"> • Canada: • While different classifications of age may be required across the framework as noted – whenever possible standardized age classifications should be used and single year ages should be avoided whenever possible. Agree that age should not be truncated (i.e. – under 65), except in obvious cases. 	...
Race	<ul style="list-style-type: none"> • With regard to the fact that the dimension race is characterised by similar problems as the dimension ethnicity, we also suggest that there should not be one international standard on data disaggregation by race. • The decision, whether data is disaggregated by race should also be made at individual country level. • With regard to disaggregation by ethnicity and race, the principle of self-identification could be applied in the process of data collection. Furthermore data privacy and the principles of confidentiality and discretion need to be fulfilled. 	/
	Comments	Comments

...		...	
Ethnicity	<ul style="list-style-type: none"> In some countries data disaggregation by ethnicity, race or colour is a common procedure, while in other countries it is prohibited by national law and/or data provision is not possible due to questions of confidentiality. These aspects show that there is not a one-size-fits-all solution which is why we recommend that there should not be one international standard on disaggregation by ethnicity. The aggregation of data at international level, which is disaggregated by ethnicity in a national context, could be associated with conceptual problems. Discriminated minorities in one country can be as well majorities in a further country. In conclusion we suggest that countries should be free to decide at country level, whether data is disaggregated by ethnicity or not. If data is disaggregated by ethnicity, the respective countries should document and publish their definitions and criteria for disaggregation, so that they are readily available. With regard to disaggregation by ethnicity and race, the principle of self-identification could be applied in the process of data collection. Furthermore data privacy and the principles of confidentiality and discretion need to be fulfilled. 	<ul style="list-style-type: none"> Q5: How to deal with small sample sizes? 	
Comments		Comments	
	<ul style="list-style-type: none"> Canada: Small sample sizes will continue to be an issue particularly for the most vulnerable groups. I believe this is something that statistical offices need to better explain to civil society etc. We cannot compromise quality or risk respondent disclosure. While working towards being able to release more, it is important to explain that we often are not able to release at the desirable level because of sample constraints. 		...
Migration status	<ul style="list-style-type: none"> Due to the existence of numerous different definitions of „migrant“ and „migration status“ between countries, a harmonised definition of migration status is required to enable comparability. As a first step we suggest the use of the UN concept of country of birth (native or foreign-born) and country of citizenship (native or foreign citizen) In the further course, data disaggregation by migration status could be extended to further population groups mentioned in the Agenda 2030. 	<ul style="list-style-type: none"> Q6: Should data be disaggregated by migration status or migration background? 	
Comments		Comments	
	<ul style="list-style-type: none"> Canada: Note that by examining country of birth it does not come close to representing migration status. Would it not be better to examine immigrant status (i.e. recent immigrants, or through the examination of the type of immigrant – i.e. refugee etc.?) 		...

<p>Disability</p>	<ul style="list-style-type: none"> The Washington Group set of questions on disability seems like a solid and widely accepted standard that is proposed as standard for the monitoring of the SDGs by the United Nations Expert Group Meeting on Disability Data and Statistics, Monitoring and Evaluation. Possible limitations could arise due to small sample sizes. 	<ul style="list-style-type: none"> Similar to Q2
<p>Comments</p>		<p>Comments</p>
		<p>...</p>
<p>Geographical location</p>	<ul style="list-style-type: none"> There are numerous different definitions of urban and rural. Therefore a harmonisation of the definition of urban and rural, respective non-urban and non-rural is necessary for the comparability of data. Specifically differences between rural areas and suburbs as well as cities, towns and mega cities should be defined clearly, with regard to varying meanings in different countries. 	<ul style="list-style-type: none"> Q7: The Working Group on Geospatial Information is working on a harmonised approach of geographical location. The identification of suitable data sources and calculations, as well as work on harmonising the definition of urban and rural is in progress. It could be one option to cooperate with the the Working Group on Geospatial Information regarding the disaggregation by geographical location.
<p>Comments</p>		<p>Comments</p>
<p>...</p>		
		<ul style="list-style-type: none"> Sweden: Cooperation is good. The urban and rural definition is known to vary between countries and be hard to use even for regions of the world, so global definitions will not be easy to settle. It will most likely be a learning experience to try and identify some central indicators and choose definitions that are suited to the questions at hand.