A TYPOLOGICAL FRAMEWORK FOR DATA MARGINALIZATION

Identifying forms of marginalization and exclusion in data-intensive ecosystems

UNKNOWN | SILENT | MUTED | UNHEARD | IGNORED
The ‘leave no one behind’ principle of the Sustainable Development Goals (SDGs) seeks to ensure that all people participate in and enjoy the benefits of sustainable development i.e., an end to extreme poverty in all its forms, reduction of inequalities among both individuals and groups, and addressing of discriminatory barriers.

While significant advances have been made across the eight Millenium Development Goals (MDGs), the outcomes have been unequally distributed with various individuals and population groups across the world not enjoying the benefit of development achievements. These individuals and groups are generally clustered as the ‘marginalized’ or the ‘excluded’.

With the centrality of the ‘leave no one behind’ principle within the Global Agenda, developing frameworks and instruments that operationalize this principle for a broad range of stakeholders is imperative to enable its practical implementation, hence giving visibility and voice to marginalized and excluded individuals and population groups.
The call for 'data revolution' has renewed impetus to data collection efforts to break the vicious cycle of invisibility and neglect through focusing monitoring efforts on marginalized groups. However, marginalization could occur throughout the full length of data processes — from data collection to impact — and across data-intensive ecosystems, including social indicators monitoring, citizen science, and urban governance, resulting in invisibility to policymakers, misrepresentation in the datasets, and discrimination.

Framed through the 'marginalized voices' metaphor, the typological framework for data marginalization presented here would facilitate the recognition of the myriad of ways through which individuals and population groups are marginalized in the data processes. It is, therefore, a useful tool to inform the formulation of mitigation strategies against specific forms of data marginalization across fields of practice.
THE UNKNOWN VOICES

DEFINING FEATURES
Being invisible in the mainstream societies, hence are unknown to the data collecting actors

EFFECTS
- Lack of reliable data on the population
- Exclusion from participation in the mainstream society

ASSOCIATED PHENOMENA
Self-exclusion/voluntary isolation

MITIGATION
Multiple development trajectories; sound legal definitions

EXAMPLES

Isolated and untouched communities
These groups usually have a history of contact with the outside world but remain unintegrated into the mainstream societies.

Modern-day slaves
Modern-day slaves include forced labours, trafficked individuals, children working in sweatshop, and sex slaves. Definitional disagreements about what constitutes slavery have created difficulties in tracking them while the fear of contact with government officials has led to self-concealment and disengagement from public life.

Concealed individuals
These groups include those who voluntarily conceal themselves from the surveillance and contact with the authorities. They include undocumented migrants and individuals with illicit income.
THE SILENT VOICES

DEFINING FEATURES
Due to personal and restricted agency, these groups lack the capacity for vocalizing.

EFFECTS
Inability for direct participation without specialized interventions

ASSOCIATED PHENOMENA
Silent to conventional methods of engagement

MITIGATION
Assistive technologies; brokered participation/intermediation

EXAMPLES
Developmentally disabled people
People with intellectual, mental, and physical disabilities face cognitive and functional limitations to articulate their wellbeing, especially through conventional methods and tools of data collection.

Children
Children are often perceived as vulnerable, dependent on others, and not legitimate meaning-makers of their world. This limits their visibility in data. In 2018, the UNICEF reported that over a half billion children are effectively uncounted.

Weak and vulnerable
These groups include homeless people and the aged living in institutions who are at risk of being left behind when they are limited from contributing to data.
THE MUTED VOICES

DEFINING FEATURES
Due to socially constructed systems of classification that devalue and discredit them in particular contexts, they are further marginalized in data processes.

EFFECTS
- Exclusion from data collection
- Subjected to discriminatory practices during the data use phase

ASSOCIATED PHENOMENA
Entrenched marginalization and exclusion; specific to socio-cultural context

MITIGATION
Social inclusion; participation; advocacy

EXAMPLES

Gender non-conforming people
The levels of acceptance for gender non-conforming people and the level of inclusiveness of a country’s laws are different around the world. Discriminatory laws and policies can influence the formal data processes and reflect in the data outcomes.

Groups at the bottom of the social ladder
In some South Asian countries, the Dalits are situated at the very bottom of the caste system. This discrimination is reflected in social policies.

Stigmatized groups
These groups are subjects to stereotypes and social rejection because they deviate from the societal expectations. They include people living with or affected by HIV and sex workers.
THE UNHEARD VOICES

DEFINING FEATURES
Being excluded from the sampling or data collection due to issues such as low capacity of data collecting actors (incl. constrained resources) and the limitations of sampling procedure, data collecting methods, and data infrastructures

EFFECTS
• Unrepresented in data
• Inability to inform the decision making

ASSOCIATED PHENOMENA
Unrepresentative sampling; uncounted

MITIGATION
Representative data collection

EXAMPLES

Non-household groups
Household surveys exclude populations living in other settings. Thus, those who do not live in households, such as institutionalized populations, homeless populations, and highly mobile populations are often not represented in data.

Digitally unconnected people
The increasing use of electronic data collecting technologies and online data collection methods means that the digitally unconnected people are more likely to be excluded from data collection.

Linguistic minorities
These groups have limited opportunity to participate in data collection which is usually conducted in official languages.
THE IGNORED VOICES

DEFINING FEATURES
Being marginalized and excluded during the analysis of the collected data either through traditional statistical analyses, such as aggregate bias and ecological fallacy, or new data approaches, such as Big Data analytics.

EFFECTS
- Underrepresented in data
- Marginal and incremental disempowerment

ASSOCIATED PHENOMENA
Aggregation bias; ecological fallacy; biased and opaque algorithms

MITIGATION
FAT algorithms; small data; participatory indicators; multilevel analysis

EXAMPLES

Minority groups
Treating minority groups, based on their ethnicity, race, or religion as a single group is often the practical option because detailed consideration of group salience can pose greater challenges for cross-national data efforts.

Underrepresented groups in general
Big Data relies on digital technologies for accumulating data. This means that population groups with less engagement with digital technologies, such as children, the elderly, the poorest, and women, may be hidden from data.

Protected classes
Bias in the data collection can result in a discriminating classifier which presents itself as an underrepresentation or overrepresentation of specific groups in the data sets.
The above data value chain visualizes the life cycle of data, from identifying a need for data to using them for impact. Data gain value as they go through each step of the value chain, transforming from low-value inputs into high-value outputs.

Data marginalization cascades through the data value chain to the extent that population groups that are marginalized at the early stages of the value chain continue being marginalized at later stages.
"Giving visibility and voice to marginalized and excluded individuals and groups."

For more information contact: mamello@unu.edu