SAMPLING AND CIVIL SOCIETY ENGAGEMENT IN SURVEYING HARD TO REACH POPULATIONS

Eugenia De Rosa, Francesca Inglese

Abstract. 'Hard-to-reach' is a term used to describe sub-groups of the population that may be difficult to reach or involve in research and social statistics. Examples include LGBT+ people, irregular migrants, homeless people and more in general those living in vulnerable social and economic situation and people at risk of discrimination. Invisibility and cumulative disadvantages may characterize hard-to-reach populations. These groups are difficult to identify and recruit, and their sampling frames are usually unavailable. Moreover, official statistics need to gather data that can help to design and monitor policies to combat inequalities, discrimination and disadvantages. The National Statistical Institute of Italy has a long tradition on investigating different hard-to-reach groups. This paper critically analyses these different experiences, focusing on two main aspects: a) sampling techniques and b) civil society involvement. Strengths and limitations, as well as future prospects are discussed.

1. Introduction

The term 'hard-to-reach' populations is often used to describe "underground communities whose members may be reluctant to self-identify and for whom no sampling frame is available or can be constructed" (Raifman, *et al.* 2022) such as LGBT+ people, irregular migrants, people experiencing homelessness, and more in general people in vulnerable social and economic situation and at risk of discrimination. They have different levels of visibility and social recognition related to their marginalized status. Invisibility and cumulative disadvantages may characterize hard-to-reach populations, as well as stigma associated with group membership. This can lead to reticence to participate in research, desire for anonymity (choose not to be identified) and mistrust of social research and statistics. In addition, researchers are rarely members of the community under study.

Defining hard-to-reach populations and sampling strategies are common challenges of this kind of research.

The "Human Rights-Based Approach To Data Collection" (2018) calls for the principle of self-identification: respondents must be placed in a position to choose whether or not to identify themselves as members of certain groups. Ideally, specific questions about protected grounds by international legislation such as sex, age, racial or ethnic origin, religion or belief, disability and sexual orientation should be addressed in surveys targeted to the whole population. While some self-

identification questions and their associated conceptual issues and definitions are clear and data well available in social statistics, as for the grounds of sex, age and disability, there is a need to improve the definitions and data availability for other characteristics, namely, gender identity, sexual orientation, racial or ethnic origin, religion or belief. In addition, extensions beyond the social domain should be considered when relevant, considering social class and cross equality and multiple dimensions.

In terms of sampling techniques these can be placed on a *continuum* from probabilistic to non-probabilistic. The approach most often used to recruit hard-to-reach populations is snowball but for official statistical agencies probability sampling is their preferred tool to meet information needs and make statistical inferences. Common methods for facilitating the participation in surveys, include the use of self-completed questionnaire to ensure maximum privacy of the respondents or questionnaire administered by experienced interviewers with knowledge of the specific populations/phenomena. At the same time different strategies to recruiting enough number of study participants and break down barriers are adopted such as developing partnerships with stakeholder organizations and relationships within the target population, community-engagement activities, including as researchers or surveyors' members of the target population, working with a community advisory board.

The National Statistical Institute of Italy (Istat) has a long tradition on investigating hard-to-reach groups such as homeless, LGBT+ people and other groups at risk of inequality and discrimination. Besides conceptual and methodological challenges, a deep knowledge of the groups investigated and the involvement of civil society emerge as common aspects when designing and producing official data on hard-to-reach populations.

The aim of this article is to critically analyses these different experiences focusing in particular on two main aspects: sampling techniques and civil society involvement, and the interplay between these two aspects. What are the most common samples used? At what stages of the data value chain does the civil society involvement take place? What can civil society do to contribute to the construction of the sample design? Main challenges, strengths, and limitations of designs for surveys of hard-to-reach populations, as well as future prospects are discussed.

Instead, in order to analyse the role of civile society, we start from the "Generic Statistical Business Process Model" that describes and defines the set of business processes needed to produce official statistics, namely specify needs, design, collect, analyse and evaluate. This model is integrated with the UN framework "Harnessing data by citizens for public policy and SDG monitoring" (Pratesi, 2023) that defines citizen's contribution to data as "the engagement of citizens in multiple processes in the data value chain, from specifying needs to use of the data to inform policy… is

increasingly recognized for its unique ability to help overcome many data challenges of our times" (UN, 2022, p. 2).

2. Sampling technique: representativeness and generalizability of results

For sampling purpose, surveying hard-to-reach populations occurs when (Marpsat and Razafindratsima, 2010): the population of interest has a relatively low number of people, which makes a survey of the entire general population very expensive; the members of the target population are difficult to identify; there is no sampling frame or there is only a very incomplete one that produces biased results; the persons concerned do not wish to reveal that they are members of the population of interest, because their behaviour is illicit, because it is socially stigmatised, and for other causes; the behaviour of the target population is not known, which leads to an inappropriate choice of places to interview them, or more in general, an inappropriate choice of method for recruitment. The sampling of hard-to-reach populations partly overlaps with the sampling of rare populations for which the frame is generally unavailable, incomplete or consists of general population lists.

In the latter case, an efficient sampling design that includes specific/rare populations must take into account that in order to have representative samples of these aggregates, a large number of people need to be surveyed. Representativeness may be compromised as the units to be selected is not controlled a priori, making it difficult to define the optimal sample size for the particular domain. Also, the estimates may be biased for differences in the socio-demographic distribution of the target population compared to the general population. One solution to the representativeness issue, as suggested by Kalton (2009) in defining sampling designs for rare populations, might be to start from a large sample to conduct a screening that can generate an adequate sample size for the domain of interest.

Observing hard-to-reach populations often involves a range of solutions based on quantitative and qualitative methods, with non-standard probability sampling techniques based on the locations frequented by the population (location/venue based sampling, time location sampling, centre sampling, capture-recapture) or their social network (respondent driven sampling - RDS) and non-probability sampling techniques based on social network or other (snowball, network sampling, convenience or random sampling, voluntary sampling, quota sampling, web panel, etc.).

The absence of 'lists' of people raises some important questions, primarily on how to construct a frame of reference population for the design of probability samples when possible or how to reach the population of interest through different techniques in such a way that statistical representativeness is guaranteed.

Sampling techniques based on the locations require the construction of lists: depending on the type of information available on the frame it is possible to refer to

different techniques. In the case of the homeless population, for example, an indirect sampling design (Lavallée, 2007; Deville and Lavallée, 2006) that uses the list of services provided in centres for the homeless can be used as a sampling frame. Indirect sampling strategy exploits the link that exists between the target population and services for obtaining the final weight of homeless person based on the probability of inclusion of the selected services and the total number of service links with the user. Other approaches that count homeless people in the places they frequent, such as the point-in-time census (Boeri *et al*., 2009), can indeed provide a framework to define sampling techniques for a subsequent survey with interviews. The indirect sampling produces correct estimates, only the coverage of the part of the homeless population that does not use services may be affected. The point-in-time census does not ensure a complete and accurate count of the population for areas not covered because inaccessible or unknown (under-coverage) and the inclusion of people who are not homeless or are counted several times (over-coverage): errors in the count, and hence in the frame, will affect the final sample estimates. From an inferential point of view, the effectiveness of such approaches depends on the completeness of the locations list and the possibility of using probabilistic selection methods.

Sampling techniques, which exploit the social network of the population of interest, through a snowball type recruitment, allow to build a list from scratch or to build a larger one. These techniques move between probability and non-probability sampling depending on the type of population, the research question and objectives; also include a variety of procedures different for a series of elements, such as the methods of choosing the people who are part of the initial group/sample, the methods of recruitment and selection of the people who gradually enter the sample. RDS is an example of probability sampling that combines the snowball technique with a mathematical model that formalises the recruitment process as a Markov chain (Heckathorn, 1997, 2002; Crawford *et al*., 2018). The recruitment process evolves in waves generated by the initial recruiters, selected non-randomly on the basis of the criterion of differentiation and ability to recruit. The data collected during the sampling process are used to make inferences about the structure of the social network and to obtain unbiased estimates. The applicability of the RDS may be limited by the strict assumptions required to achieve the desired inference results that are not easily verifiable a priori. Some conditions can be verified by information obtained from respondents, such as the dimension of their social network or the reciprocity of links between recruits and recruited.

In general, these techniques are characterized by complexities linked both to the definition of the probability of inclusion of the sampling units and to the estimation phase in which the risk of multiple counting of the units must be faced, the same individuals can be indicated by different networks or frequent more places.

In order to derive conclusions about the population from the sample and to obtain generalizable results, it is necessary to assume that the sample is representative of the population of interest. In non-probability sampling - that constitutes a quick, easy and inexpensive way of obtaining data - the assumption of representativeness for the generalisation of results is a risky, as it is difficult to assess its validity. There is no way of estimating the inclusion probability of a unit and is no certainty that unit has a chance of being included, making it impossible to estimate sampling variability or identify possible bias. Statistical representativeness is an unattainable goal when using non-probability techniques, but it can also be a problem when using non-standard techniques based on places attended or social network. The representativeness can be compromised by partial coverage, e.g. when people who do not attend venues intended for the population are excluded, when using a web-based survey technique or survey techniques that are too invasive of people's privacy. Also in techniques based on social network, "the representativeness of the sample is conditioned by the way in which individuals are recruited" (Caputo 2013).

3. Non-probability vs probability sampling techniques

Probability sampling is the preferred tool adopted by statistical institutes to produce official statistics. In recent years, research on the use of data generated by non-probability sampling techniques for the production of official statistics has been increasingly explored (Kalton 2023, mainly thanks to the diffusion of big data. The use of alternative sources in official statistics has led to a paradigm shift and a change in the traditional approach to statistics. However, data from non-probability sources pose some challenges in terms of data quality, including the potential presence of participation and selection biases: data collected should be used with greater caution.

In non-probability samples, the degree to which certain forms of bias occur can greatly affect the accuracy of the results; the efficiency of inference is not guaranteed by a theoretical-mathematical basis but depends exclusively on the correct specification of models defined on untestable assumptions on the phenomenon and on the similarity between the sample and the population; the presence of selection bias makes these hypotheses dangerous. To generalise the results to the whole population, it is necessary to use probability sampling or integration techniques.

In adopting non-probability sampling techniques, specific strategies can be implemented at the sampling stage to reduce potential bias, e.g. by drawing the non-random sample from a large panel of volunteers, taking care to ensure that there is a match of characteristics between sample units and those belonging to a probability sample selected from a frame covering the same reference population.

In the data processing phase, various methods can be used to correct or reduce the effects of bias in the results, thereby making the inferential process more reliable and increasing the external validity of a survey. In general, estimation methods from

non-probability samples refer to integration techniques that use weighting or predictive approaches. Weighting approaches (calibration, propensity score adjustment, statistical matching) attempt to reduce selection bias and coverage problems by manipulating the data so that the sample comes to resemble the population in the distribution of covariates. Predictive approaches fall within the framework of super-population models. They treat the analysis variable as if it had been generated by a model specified on the basis of the data collected from the non-probability sample. The values of the auxiliary variables must be known not only for the sample units but also for those of the target population, or at least the totals. Common to the different approaches is the use of auxiliary information observed in the non-random sample, including at least the totals (or averages) in the target population, the values for each unit in the target population, and the values for each unit in a reference probability sample. Data may be obtained from official statistics or high-quality probabilistic surveys. It may be necessary to conduct parallel probabilistic surveys in order to detect more relevant variables for the integration.

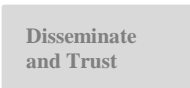### 4. Istat experiences: Surveying Homeless People

In Italy, research on people living in extreme poverty was carried out in 2011 and 2014 by Istat - in agreement with the Ministry of Labour and Social Policies, the Italian Federation of Organisations Working with the Homeless (fio.PSD) and Italian Caritas - and led to a national estimate of the homeless population (De Vitiis *et al*., 2014). In 2014, an experimental survey was carried out only in the city of Turin to estimate also the part of the population that does not use the services.

The surveys were based on the system of services used by homeless people. This made it possible to adopt a location sampling design referring to the theory of indirect sampling: in the national surveys the services were represented by soup kitchens, lunch and dinner places, and night shelters, while in the experimental survey conducted in Turin in 2014, they were services provided in public spaces by street units that carry out rescue and support activities for people in a state of social marginality (Inglese and Masi, 2019). The information on services were collected through censuses carried out prior to the implementation of the surveys, which provided the necessary information base for the development of the indirect sampling methodology. Starting from existing archives, the complete mapping of services was carried out using the snowball method, through a survey network selected by fio.PSD and composed mainly of operators in the sector. The list of street unit services was updated in 2014 from the previous complete mapping of services through interviews conducted with street unit managers. The censuses made it possible to build up a picture of the system of services provided by private and public organisations to people living in a state of social marginality, including the structures that operate "on the street". The definitions of "street unit" and "home service"

(services/support action distributed by street units in public spaces at the same time) provided the basis for applying the same methodology used in the national survey. For the city of Turin it was possible to obtain an overall estimate of homelessness.

In 2023, Istat provided support for the implementation of a survey of the homeless people in the city of Rome, in collaboration with Roma Capitale. The "Notte della solidarietà survey" (a point in time survey) was carried out on the evening of 20 April 2024, and counted people sleeping in the overnight shelters and on the streets, by means of a short web questionnaire that captured some of their basic characteristics. Difficulties in measuring homelessness lay in how to reach them but also in the lack of a definition on which living conditions may lead to consider a person as a homeless. More than 1700 volunteer surveyors (from associations, universities and citizens) participated in the night survey. Volunteers, organized in teams, were in charge of conducting the survey on the street, while the survey in the night shelters was carried out by the people in charge of running these facilities. The pilot survey (2023) showed how in point-in-time approach to investigate homelessness, training of non-professional surveyors is fundamental in approaching homeless people but also a critical issue, to guarantee the quality of data. Therefore, for the finale survey, a more targeted training on the questionnaire was designed for a more professional approach, and an increased care in the team composition by providing for each team at least one student, always keeping a gender balance. The involvement of civil society was crucial, also in terms of socio-cultural impact on the representation and knowledge of homeless people (Table 1).

Table 1 – *Sampling strategies and civil society involvement in the Istat and Istat-Roma Capitale surveys on homeless people (2011, 2014, 2024).*

Target	Sampling	Civil Society Engagement
Homeless people who frequent services (soup kitchens, night shelters) (2011, 2014)	Non standard probabilistic **Design**	▪ Fio.PSD, the Italian Caritas and surveyors from services **Collect**
Homeless people in Rome (2024) sleeping rough and in overnight shelters	Homeless Census – Point in Time **Collect**	▪100 associations and entities, 6 Universities and citizens ▪ 1,700 volunteers and state workers spend the night walking the streets of Rome to make a count/survey **Disseminate and Trust**

In 2024, Istat launched a project for the continuous monitoring of extreme poverty, which involves numerous research activities and types of surveys. The survey on homelessness in the 14 metropolitan areas of Italy is currently being carried out. The approach used is the point-in-time census (shelter and street night)

followed by a sample survey with in-depth thematic interviews. Istat investigated also the severe housing exclusion of Roma people within a project carried out in collaboration with the National Antidiscrimination Office (Unar), during the period 2018-2023 (Di Leo and Nur, 2023).

5. Istat experiences: Surveying Group at risk of discrimination

To date, Istat adopted two main strategies to investigate the condition of groups at risk of discrimination, by introducing self-identification questions into (a) Discrimination surveys targeted to the whole population and (b) Surveys targeted to specific population such as LGBT+ people.

In 2011, Istat addressed for the first time issues related to diversity in terms of sexual orientation and gender identity with the "Survey on Discriminations by Gender, Sexual Orientation and Ethnic Origin" (Istat, 2013). The survey, CAPI (mandatory) and with an optional paper self-administered questionnaire, included questions on sexual orientation which allowed to provide a first estimate of the homosexual and bisexual population in Italy. However the limited sample size (7,725 individuals aged between 18 and 74) did not allow for estimates on some specific groups at risk of discrimination as for transgender people or bisexual man and women. In view of the new edition of the Discrimination Survey (2025), the mode technique was redesigned as well as the sample size of the whole population - about 50 thousand people. At the same time updated and more self-identification questions have been introduced. This opens up an interesting scenario in terms of estimation possibilities as well as corrections of results from non-probabilistic sample surveys.

In general, the involvement of civil society was on the questionnaire design of the first edition of the survey. In 2018, Istat dealt with these issues again in the framework of a collaboration agreement signed with Unar to fill an information gap on LGBT+ populations. The Istat-Unar project "Labour discrimination against LGBT+ people and diversity policies" (2018-2023) is characterized by a mixed method approach (probabilistic and non-probability sampling technique; qualitative and quantitative tools), multiple perspective approach (LGBT+ people, employers, stakeholders) as well as a participatory approach with the interaction of experts, academics and LGBT+ associations (De Rosa and Inglese, 2018). It included the direct collection of information from LGBT+ people and from employers, particularly enterprises, and the main stakeholders. Three CAWI surveys based on a web self-completed questionnaire were carried out on three different target groups of LGBT+ people (people in civil union, LGB people not in civil union, trans and non-binary people) using different sampling strategies. The first two surveys mainly focused on aspects related to sexual orientation; the third focuses on gender identity issues. Self-identification of respondents as LGBT+ was a key principle adopted.

SOGIESC (sexual orientation, gender identity and expression and intersexuality) indicators were gradually tested and introduced into the three surveys.

The project included surveys both with standard and non-standard sampling techniques. The most robust survey was on People in Civil Union or formerly in union because a list was available. A second survey, addressed to LGB people who have never been in Civil Union, was carried out in 2022 with a snowball technique, the web-RDS. Fifty LGBT+ associations throughout the national territory identified first respondents ("seeds") belonging to the population target, and respondents played an active role in recruiting (at least 4 people). After an established time passed from the beginning of the survey, there was evidence that the RDS snowball technique was not working properly. In order to go on with the work, the option of a convenience sample was considered since it could anyway provide interesting and qualitative information on the target population of homosexual and bisexual persons. Finally, a survey on "Labour Discrimination against Trans and Non-binary people", based on a convenience sampling, was carried out in 2023. Questions on gender identity allow to identify the target population. Table 2 shows LGB participant distributions in Istat-Unar surveys and EU FRA LGBT II survey 2019 for Italy.

Table 2 – *Main LGB target profiles.*

Survey	Lesbian	Gay	Bisexual Women	Bisexual Men
Istat-Unar 2020-2021	5,828	13,162	847	352
LGB in Civil Union	(28.9%)	(65.2%)	(4.2%)	(1.7%)
Istat-Unar 2022	282	640	166	71
LGB not in Civil Union	(24.3%)	(55.2%)	(14.3%)	(6.1%)
FRA 2019	1,853	4,789	1,481	627
LGB Sample size (unweighted)	(21.2%)	(54.7%)	(16.9%)	(7.2%)

Definitions and indicators provided in the questionnaire where discussed and shared with LGBT+ associations and LGBT+ people but also with experts, academics and other stakeholders (e.g. enterprises' associations, trade unions, networks of LGBT+ workers etc….) were also involved (design phase). Self-identification questions were developed consultation with civil society. Civil society was involved in different stages of the surveys and its role was crucial in defining-recruiting-data collection operations (Table 3).

The engagement was also in the build phase: associations and other civil society actors tested and gave feedback on the questionnaires developed for the different project surveys. Another contribution came from respondents' remarks in open questions within the surveys. In this sense a participatory approach increased the validity of data. Regarding the collection phase, in all three surveys civil society foster the participation of target populations and create a climate of trust in official

statistics. In addition, for the survey with RDS technique respondents played a crucial role in recruitment.

Table 3 – *Sampling strategies and civil society in the Istat-Unar project "Labour discrimination against LGBT+ people and diversity policies" (2018-2023).*

Target	Sampling	Civil Society Engagement
People in Civil Union or formerly in Union	Census over 21,000 respondents aged 18 and over ▪Self-identification questions on sexual orientation	▪Working groups with stakeholder, LGBT+ associations, experts and consultation with LGBT+ people
LGB People who have never been in Civil Union	▪ 1,159 respondents aged 18 and over ▪ Self-identification questions on sexual orientation and (testing) gender identity	▪Working groups ▪Associations and respondents: recruiting and dissemination of the survey
Trans e no-binary people	▪Self-identification questions on gender identity (2 step approach)	▪Working groups ▪Associations: dissemination of the survey

Specify needs	Design	Built	Collect and Trust

However, this experience showed the importance to qualitative approaches in order to deepen the contextual knowledge about the network of the population (more formative study). At the same time the involvement of association and civil society was not enough to ensure the participation of well-connected members of the target population to serve as initial participants or aid rapport-building to increase the likelihood of people participating in the study.

6. Future prospects and conclusions

At European level, the EC calls for a national strategy for equality data and, at the same time, Eurostat is working on equality and non-discrimination statistics toward a harmonization of concepts and classifications. The introduction of self-identification questions about sex/gender, racial or ethnic origin, religion or belief, disability, age and sexual orientation in surveys with large samples is the ideal for investigating groups at risk of discrimination. This is not always possible, and in addition, this is not appropriate for specific groups such as Roma and homeless People that require ad hoc sampling strategies and surveys.

The Istat' tradition in surveying hard-to-reach groups underlines that integration techniques and other ad hoc solutions need to be found to obtain in-depth information on the target group of the population. Quantitative and qualitative methods, which can respond to different research needs, should be able to go together, and probability sampling can help correct various forms of bias (potential presence of participation and selection biases) that affect non-probability sampling. Data from non-probability sampling pose some challenges in terms of data quality, the accuracy and generalisation of the results to the whole target population, need to adopt integration

techniques based on the use of auxiliary information observed in the non-random sample and in high-quality probabilistic surveys, or know from censuses, administrative or demographic sources and others official statistics. Parallel surveys to detect variables relevant for integration may need to be conducted.

More generally, the design of surveys on groups at risk of discrimination could benefit from programming discrimination survey design for correcting estimates of subsequent surveys on specific groups (such as LGBT+) based on non-probability sample (integration techniques); using longitudinal big datasets to follow the same group of people over time or asking the same questions to representative groups of different people over time (i.e., panel studies - non-representative sample, but generalizable results tank to the possibility of identifying the temporal order of the variables); programming dedicated surveys to reach small groups; defining a good recruitment strategy and variables/indicators and paradata useful for the explanation of the specific group' survey participation.

The participation of civil society often serves as a bridge to define sampling strategies, but its role is not limited to this phase. Istat' experiences show how the involvement of civil society can cover different steps such as specify needs and questionnaire design, organizational aspects and research infrastructure, data collection and dissemination. Even citizens can be surveyors as in the case of homelessness surveys. This leads to pushing the methodological debate more systematically in analysing wow does a participatory approach affect the quality of data, considering the different step of the process needed to produce official statistics. The exchange of experiences between countries and the provision of tools to evaluate and monitoring such aspects would be very useful, keeping in mind also how civil society can use these data to foster social change, disseminate socio-cultural meanings supporting alternative narratives of the social reality (Milan and van der Velden, 2016, p. 67).

References

CRAWFORD F. W., WU J., HEIMER R. 2018. Hidden population size estimation from respondent-driven sampling: a network approach, *J Am Stat Assoc*. Vol. 113, No. 522, pp. 755–766. doi:10.1080/01621459.2017.1285775

BOERI T., BRAGA M., CORNO L. 2009. L'economia invisibile dei senza casa. in GNOCCHI R. (a cura di), *Homelessness e dialogo interdisciplinare. Analisi a confronto tra modelli diversi*, Roma: Carocci.

DEVILLE J.C., LAVALLÉE P. 2006. Indirect Sampling: The Foundation of the Generalized Weight Share method, *Survey Methodology*, Vol. 32, No. 2

DE ROSA E., INGLESE F. 2018. Diseguaglianze e discriminazioni nei confronti delle persone LGBT: quale contributo della statistica ufficiale?, *RIEDS,* Vol. LXXII, No. 4, pp.77-88.

EU LGBTI Survey II. 2019. https://fra.europa.eu/en/project/2018/eu-lgbti-survey-ii

HECKATHORN D. D. 1997. Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations, *Social Problems*, Vol. 44, No. 2, 1 pp. 174–199.

HECKATHORN D. D. 2002. Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations, *Social Problems,* Vol. 49, No. 1, pp. 11-34. Oxford University Press.

INGLESE F., MASI A. 2019. Una metodologia per la stima della homelessness attraverso le Unità di strada: il caso della città di Torino, *IWP*, N. 5.

ISTAT. 2014. Indagine sulle discriminazioni. Anno 2011.

ISTAT. 2021. Abitare in transizione.

ISTAT-UNAR. 2022. Survey on Labour Discrimination toward LGBT+ People (in Civil Union or formerly in union). Year 2020- 2021.

ISTAT. 2022. Pilot Survey on Discrimination (2022-2023).

ISTAT-UNAR. 2023. L'indagine sulle discriminazioni lavorative nei confronti delle persone lgbt+ (non in unione civile o già in unione). Anno 2022.

ISTAT, COMUNE DI ROMA. 2024. I senza tetto a Roma, press release.

KALTON G. 2009. Methods for oversampling rare subpopulations in social surveys. *Survey methodology*, Vol. 35, No.2, pp. 125-141.

KALTON G. 2023, Probability vs. Nonprobability Sampling: From the Birth of Survey Sampling to the Present Day, Statistics in transition new series, Vol. 24, No. 3, pp. 1–22, https://doi.org/10.59170/stattrans-2023-029

LAVALLÉE P. 2007. *Indirect Sampling*. New York: Springer.

MARPSAT M., RAZAFINDRATSIMA N. 2010. Survey methods for hard-to-reach populations: introduction to the special issue, *Meth. Innovations Online*, 5(2).

MILAN S., VAN DER VELDEN L. 2016. The Alternative Epistemologies of Data Activism, *Digital Culture & Society*, Vol. 2, No. 1. pp. 57-74.

NUR N., DI LEO F. 2024. Reaching the unreachable. Innovative strategies to approach hard-to-reach populations, *RIEDS*, Vol. 78, No. 2, pp. 41-51.

PRATESI M. 2023. Towards A Framework To Harness Data By Citizens: The Istat Experience, presentation 64th ISI World Statistics Congress - Ottawa, Canada.

RAIFMAN S., DEVOST M.A., DIGITALE J.C. 2022. Respondent-Driven Sampling: a Sampling Method for Hard-to-Reach Populations and Beyond, *Curr Epidemiol Rep*, Vol. 9, pp. 38–47.

UN Office of the High Commissioner for Human Rights. 2018. A human rights-based approach to data

_____

Eugenia DE ROSA, Istat, eugenia.derosa@istat.it
Francesca INGLESE, Istat, fringles@istat.it