Data collection strategy on an elusive population: technique, process design and monitoring indicators

Monica Perez, Linda Porciani, Federico De Cicco – ISTAT|Directorate for Data Collection
Francesca Inglese – ISTAT|Directorate for Methodology and statistical process design
Eugenia De Rosa – ISTAT|Directorate for Social statistics and welfare

ISTAT|Italian National Institute of Statistics

Outline

ISTAT-UNAR project "Labour discrimination against LGBT+ people and diversity policies in enterprises"

❑ **Mixed method** (quantitative-qualitative) and **multiperspective approach** (stakeholders, enterprises, LGBT+ people)

❑ Different surveys and different target groups of LGBT+ people based on respondents' self-identification

Experimentation of the Respondent Driven Sampling Technique (RDS)

❑ **Survey Design**: Sampling Technique and Web-questionnaire

❑ **Process design** (2 step) and privacy concern

❑ Monitoring **indicators**

❑ Lessons learnt

Surveys targeted at LGBT+ people: the challenges

- Provide insights on **labour discrimination against LGBT+ people**

- **Hard to reach an invisible population**: reticence of some people and underreporting of discriminatory phenomena

- **Generalizability of results and sampling challenges**: representative surveys on the LGBT+ population are difficult to carry out mainly due to the **absence of lists** of people whose sexual orientation and/or gender identity are known

Different data collection strategy for different targets within the LGBT+ population

1

Survey on individuals who are/have been in a Civil Union
(same-sex couples, over 21,000 people)

2020-2021

2

Survey on LGB people who have never been in a civil union, through the Web Respondent Driven Sampling and convenience sample

(more than a thousand of LGB respondents)

January-May 2022

3

Survey on trans and non-binary people, through a non-probabilistic sample
in progress

Istat

Respondent Driven Sampling (RDS)| *motivations*

RDS strategy is helpful to reach the hidden population

1. **The sampling strategy has a probabilistic approach**

    It combines the <u>snowball technique</u> with a <u>mathematical model</u>
    (probabilistic) (Salganik e Heckathorn, 2004; Volz e Heckathorn, 2008)

2. **The sample is based on social network of individuals of target population**
    It starts with a *sample of convenience*
    It is *respondent-driven:* at each wave, respondents are used to select or drive the next sampling wave by selecting other individuals from the target population
    Through many sampling waves *the dependence of the final sample on the initial sample is reduced*
    The RDS sample *inclusion probabilities are estimated assuming that the sampling process is a Markov chain*
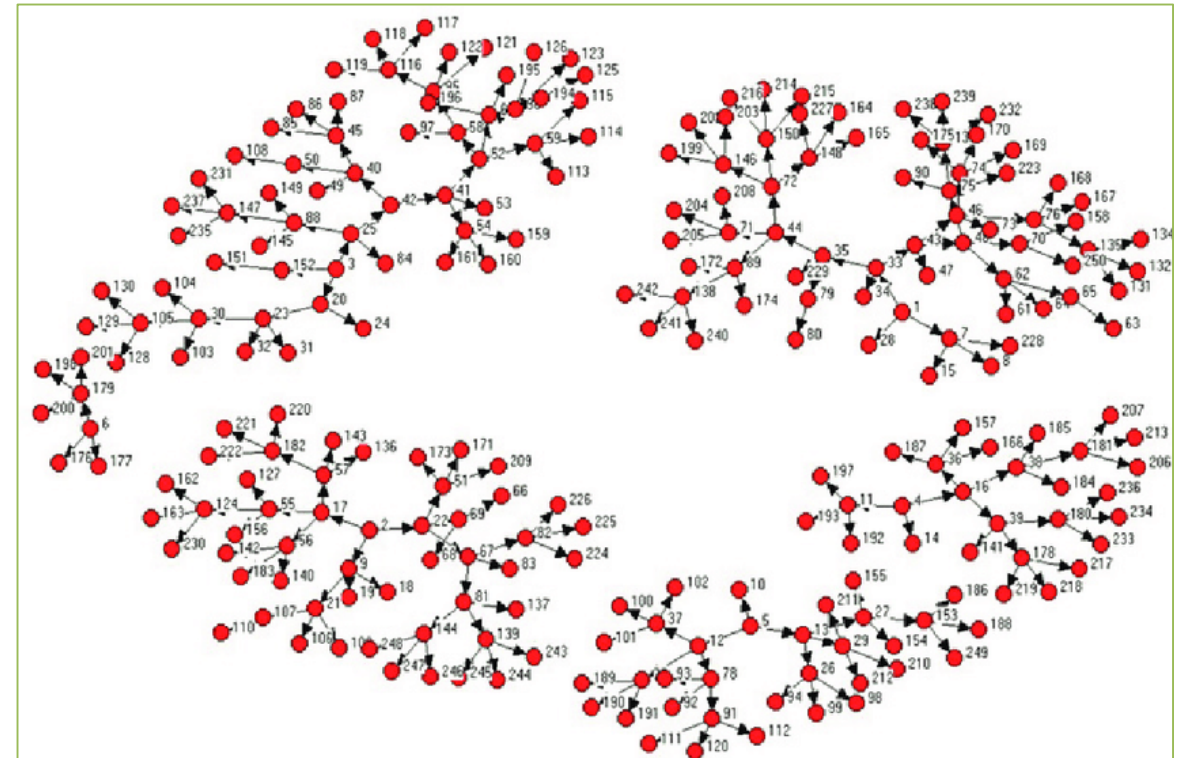
3. **The RDS permits to make inference**

    It allows to make inference about the network structure and estimation on the target population

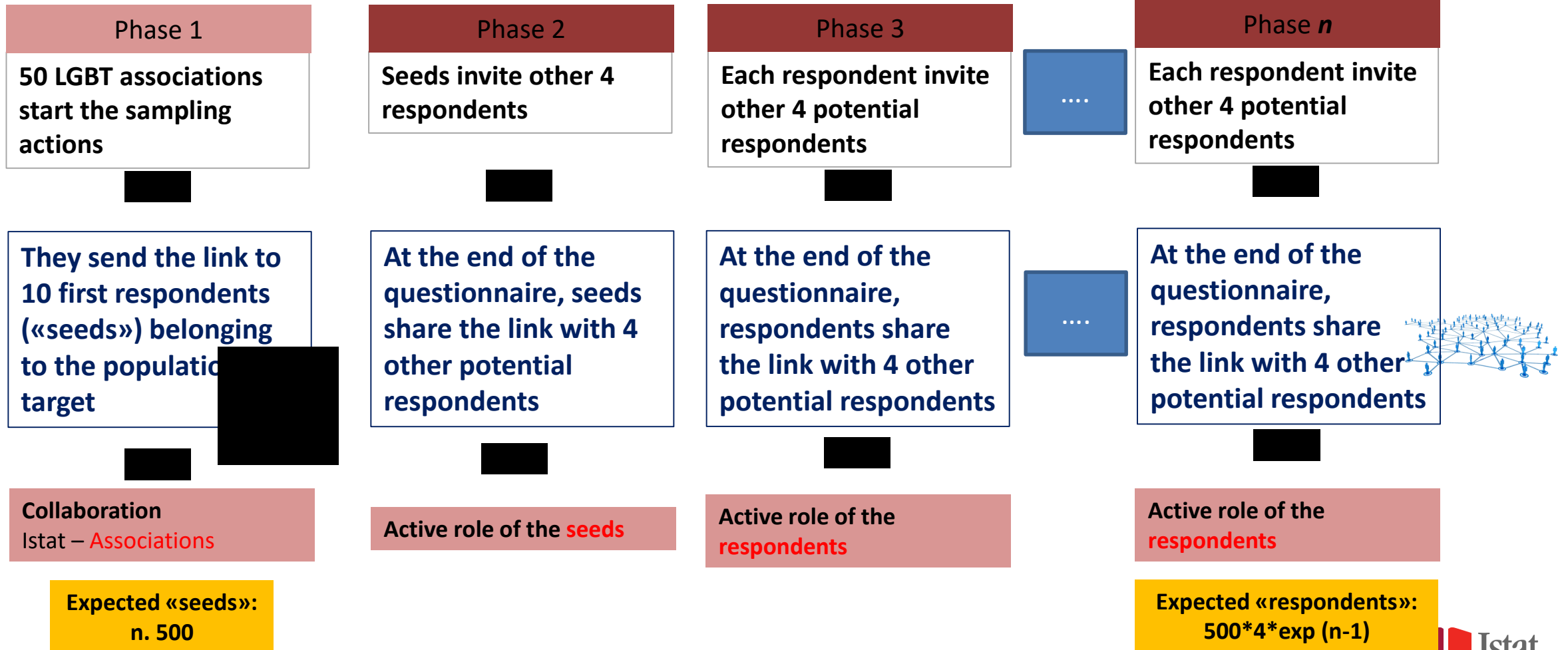# Respondent Driven Sampling (RDS)| *survey design*

Web-based-Respondent Driven Sampling Survey

- **Formative** study

- **Fifty LGBT+ association** throughout the national territory identified **first respondents** ("seeds") belonging to the population target

- Seeds must have some characteristics defined by Istat' researchers (sex, age, sexual orientation, geographical area)

- **Respondents play an active role in recruiting** new respondents who belong to the target population and to their network of relationships

- A **convenience sample as an exit strategy**

Respondent Driven Sampling (RDS)| *on the field*
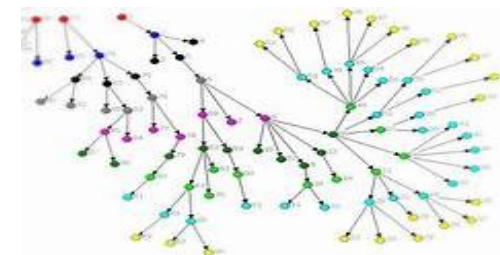
First OS national survey adopting RDS

Phase 1	Phase 2	Phase 3		Phase *n*
50 LGBT associations start the sampling actions	Seeds invite other 4 respondents	Each respondent invite other 4 potential respondents	....	Each respondent invite other 4 potential respondents
They send the link to 10 first respondents («seeds») belonging to the population target	At the end of the questionnaire, seeds share the link with 4 other potential respondents	At the end of the questionnaire, respondents share the link with 4 other potential respondents	....	At the end of the questionnaire, respondents share the link with 4 other potential respondents
Collaboration Istat – Associations	Active role of the **seeds**	Active role of the **respondents**		Active role of the **respondents**
Expected «seeds»: n. 500				Expected «respondents»: 500*4*exp (n-1)

Istat

The effectiveness of the method and the duration of the survey depend on the **propagation capacity of the network**

If for any reason a participant decides not to "propagate" because he/she becomes discouraged , loses confidence, loses referrals, etc. that node does not produce offspring and the network reduces its propagation effectiveness by limiting the achievement of a satisfactory sample
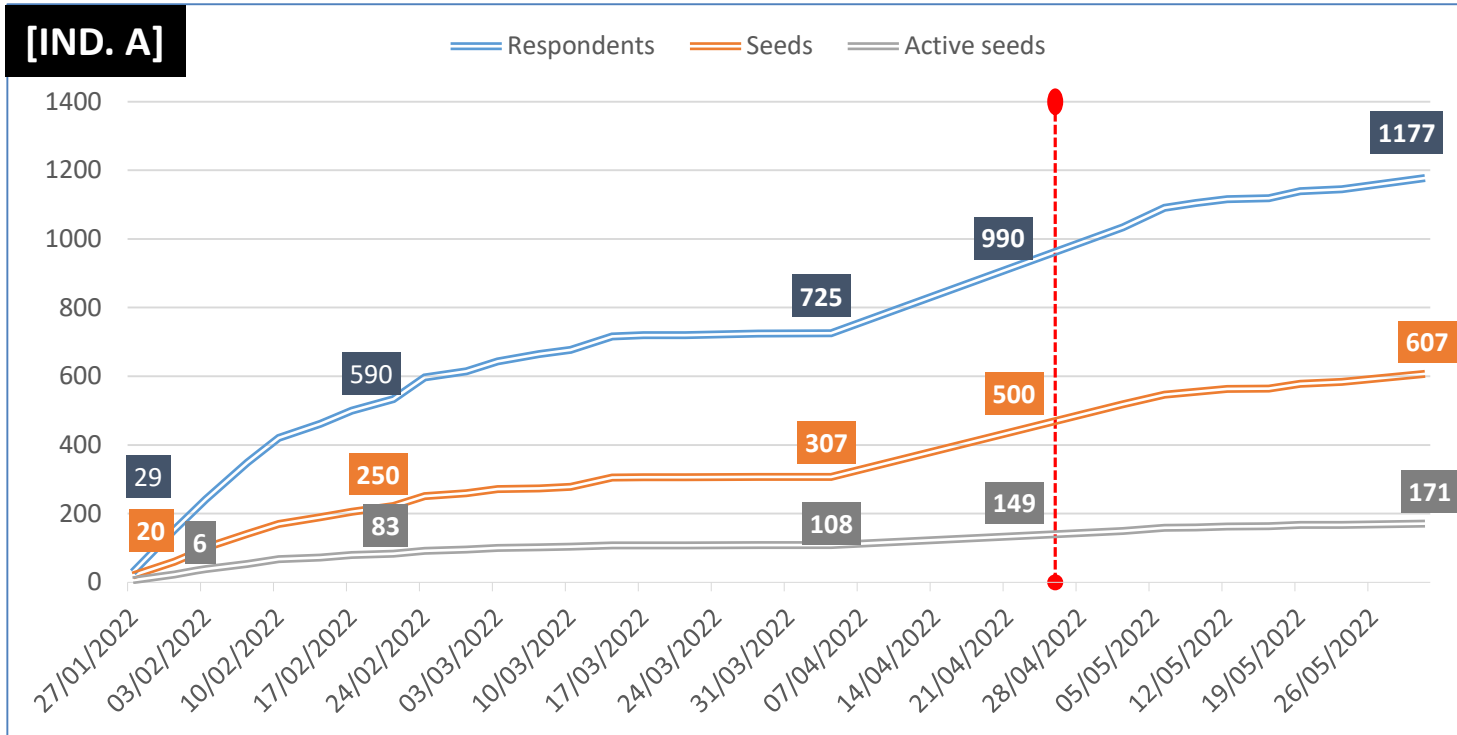
A **system of indicators** ables to monitor

- the **strength of «seed» and network propagation [IND. A]**

  - No. Seeds active/ non active
  - No. Respondents active/ non active
  - No. Created chain

- the **typology of the respondents [IND. B]**

  - Sex
  - Sexual orientation
  - Age group
  - Participation in Association

Process Indicators | *the findings*

[IND. A]

Legend: — Respondents — Seeds — Active seeds

Y-axis values: 0, 200, 400, 600, 800, 1000, 1200, 1400

X-axis dates: 27/01/2022, 03/02/2022, 10/02/2022, 17/02/2022, 24/02/2022, 03/03/2022, 10/03/2022, 17/03/2022, 24/03/2022, 31/03/2022, 07/04/2022, 14/04/2022, 21/04/2022, 28/04/2022, 05/05/2022, 12/05/2022, 19/05/2022, 26/05/2022

Data labels:
- Respondents: 29, 590, 725, 990, 1177
- Seeds: 20, 250, 307, 500, 607
- Active seeds: 6, 83, 108, 149, 171

- **Low activity** of LGBT associations: **38 over 50**

- **Low participation** of the «seeds»: **64% over the expected**

- **Low activity of seeds**
  - **6,4 active seeds** over the expected 10
  - **62%** seeds **without propagation**

- **2,4 respondents** by seed (versus the expected 4)

Istat

Process Indicators | *the findings*

Start DC
26 Jan

Respondents profile   [IND. B]

- 61,0% males
- 78,7% homosexual
- 55,6% age group 18-34
- 38,5% participation to association life

- Seeds and other **respondents** are similar
- **Homogeneity** of the respondents
- **Low distance** from the first association

Change strategy
26 April

[IND. A + IND. B]

RDS doesn't allow to make inference!

End DC
31 May

From RDS to SNOWBALL

Istat – LGBT Associations meeting

- For sharing in progress results
- For supporting the DC strategy
- For sharing the DC change
- For opening link in LGBT associations webpage

9

Istat

LESSON LEARNED

- **Experimental procedures to manage privacy issues**: two-step model; request of the respondent's email address

----------------------------- improving procedures and data collection tools to engage distrustful people

- **Low knowledge of initial respondents (seeds)** no chance of training the seeds; no incentives for the respondents

----------------------------- improving communication and training, incentives, choice of seeds by researchers

- The **recruitment of possible LGB respondents: is an indelicate operation**?

----------------------------- improving a better communication on privacy by design approach

- **The LGB networks are too limited or fragmented, even at a territorial level**?

----------------------------- improving the monitor indicators regarding network propagation

- **Other survey based on RDS**

----------------------------- studying the possibilities to apply RDS to other population (foreigners?)

Istat

Thank you!

MONICA PEREZ | perez@istat.it

LINDA PORCIANI | porciani@istat.it

FEDERICO DE CICCO | fedecicc@istat.it

FRANCESCA INGLESE | fringles@istat.it

EUGENIA DE ROSA | eugenia.derose@istat.it

Istat | Istituto Nazionale di Statistica