

UNECE 2024 Expert Meeting on Statistical Data Collection and Sources
22 – 24 May 2023

Citizen Generated Data and machine learning: a way to study violence against women

Gianpiero Bianchi

Alessandra Capobianchi

Maria Giuseppina Muratore

Claudia Villante

OUTLINES

1. The helpline against Violence Against Women and Stalking 1522
2. Adding data-sources on VAW, by using citizen-generated data: classification rules adopted and data processing
3. The methodological approach to improve data quality
4. Analysis of the results on real word data from helpline 1522
5. Next steps

1. HELPLINE 1522

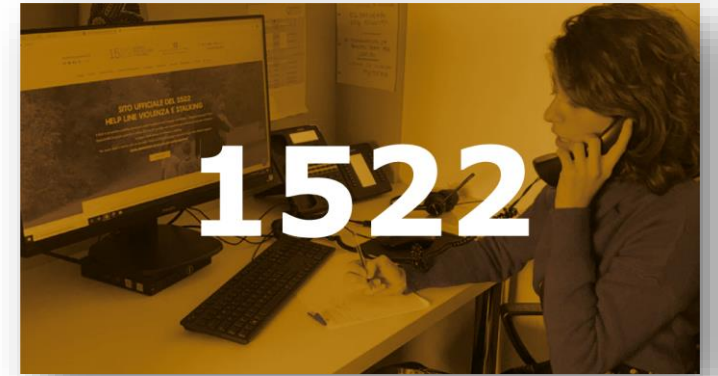
1522 is a public service promoted by the Presidency of the Council of Ministers - Department for Equal Opportunities (DEO). It was activated in 2006 with the **aim of developing a broad systemic action for emerging and combating the phenomenon of intra- and extra-family violence against women.**

In 2009, with the entry into force of Law 38/2009 amended in 2013 on the subject of persecutory acts, it also started an action to **support stalking victims.**



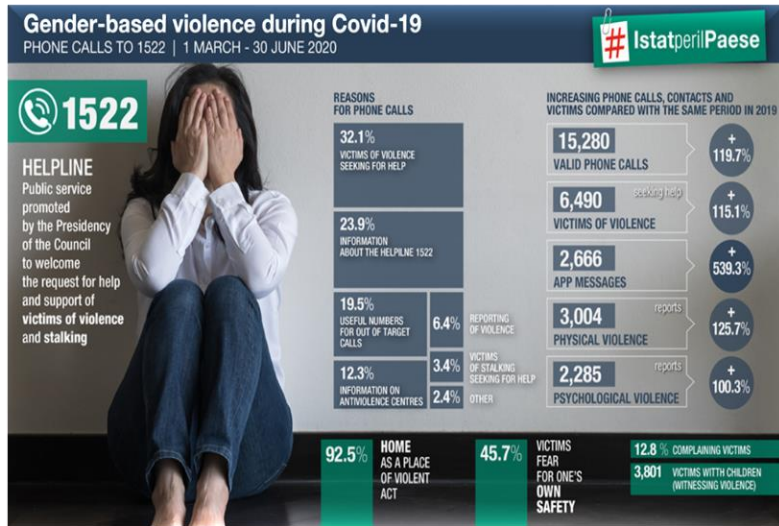
1. HELPLINE 1522 - MAIN GOALS

The helpline number is **free and active 24 hours a day** and the reception is available in **different language** (*11 languages*)



- Provide a **first listening** to the needs of those experiencing violence and stalking
- Offering **useful information and orientation** towards Anti-Violence Centers, social-health services and other organisations on the national territory included in the **official mapping** of the Presidency of the Council – DEO
- Activate **intervention of law enforcement** agencies in cases of violence of an emergency nature.

1. HELPLINE 1522 - DATA COLLECTION FROM CSO

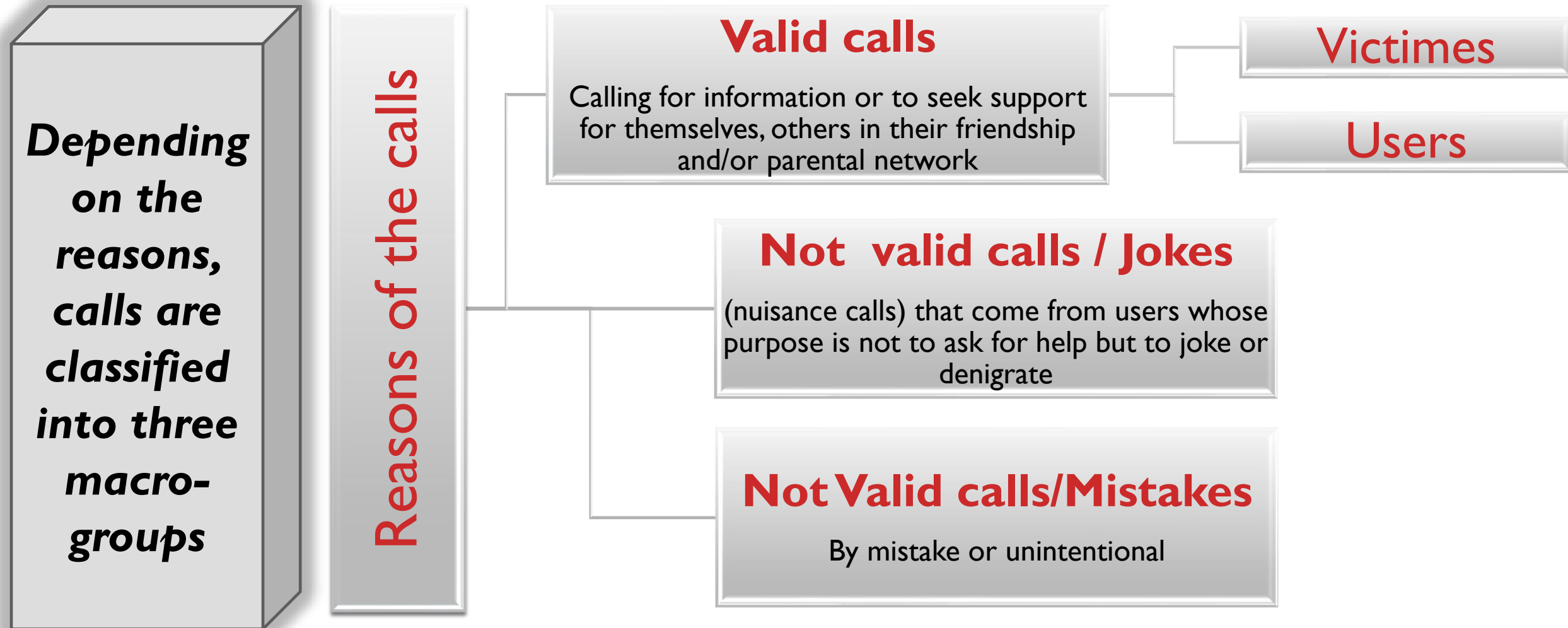


The information provided during the call is recorded on a **computerised platform** whose data has been available since January 2013.

Registration takes place following questions asked by the *NGO PROFESSIONALS* of the public utility number according to **classification rules** whose filter is the **reason for the call**.

According to the cooperation agreement between DEO-ISTAT for the creation of an '**Integrated Data Collection and Processing System**', since 2018, , the data from the computerized system in use at 1522 are sent to ISTAT, which takes care of the analysis and processing

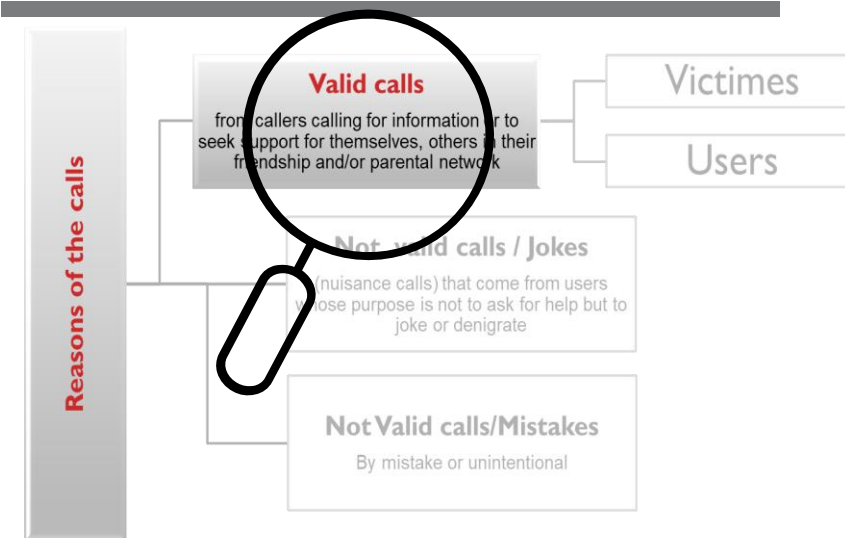
2. HELPLINE 1522 - DATASET STRUCTURE



2. HELPLINE 1522 - DATASET STRUCTURE

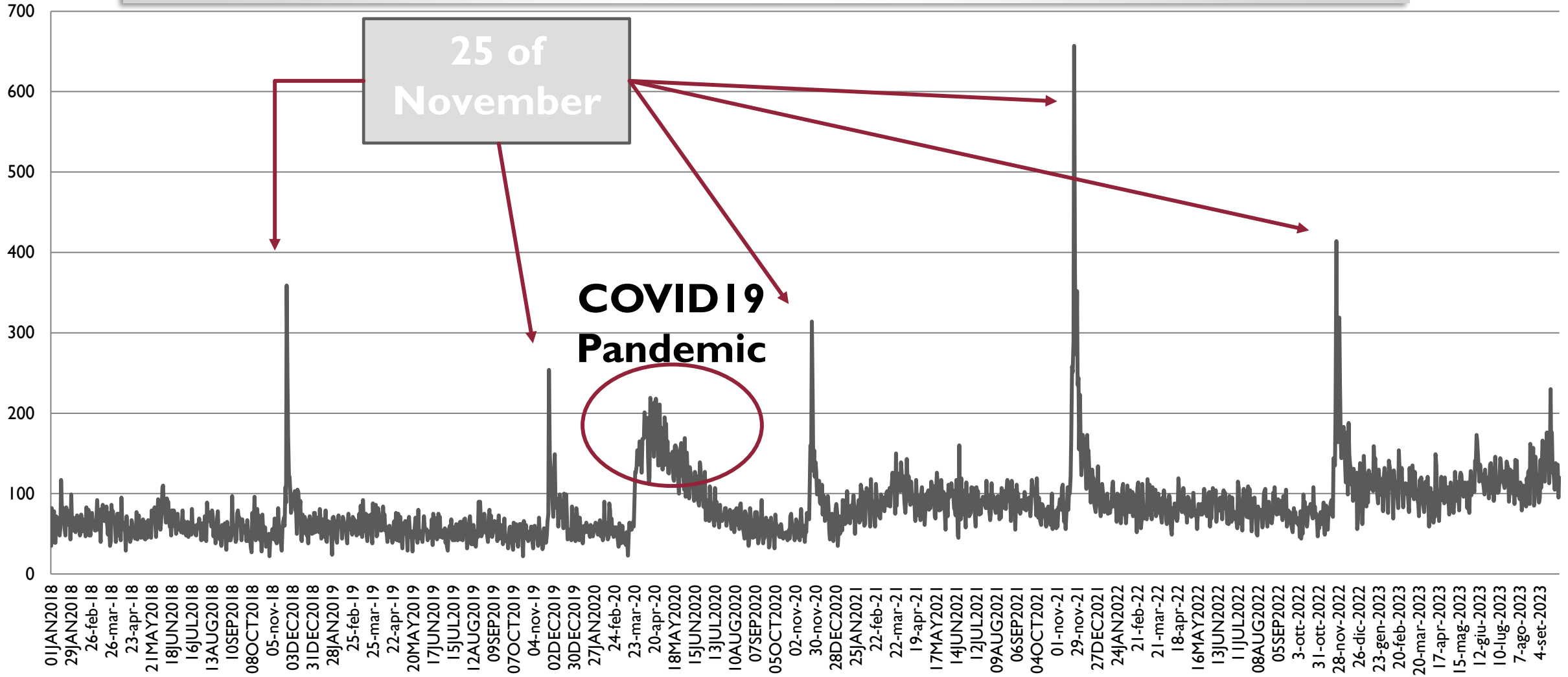
CATEGORIZATION OF **VALID CALLS**

- Victim of violence seeking for help
- Information about the helpline 1522
- Information about national shelters for victims of violence
- Reporting of violence
- Useful phone numbers for out of target calls
- Victim of stalking seeking for help
- Legal information
- Emergency
- Information for professionals on the procedures to be followed in the event of violence
- Reporting of public services malfunctions
- Reporting of media misinformation
- Information on legal responsibility of the public services workers
- International after hours calls
- Victim of discrimination seeking for help

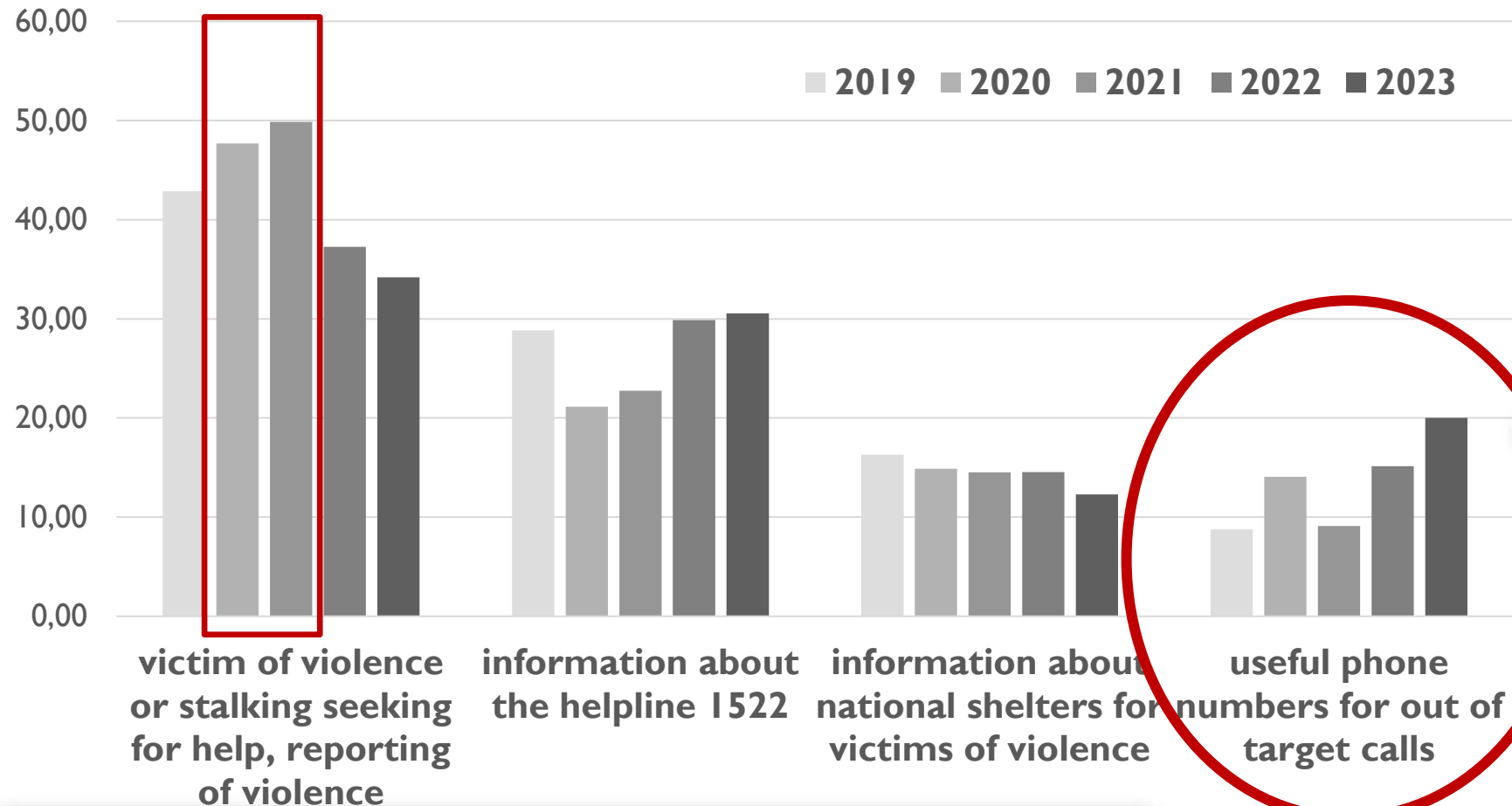


Daily calls January 1, 2018 - September 30, 2023

Data highlight the supportive role played during the pandemic



2. Reasons to call (users and victims) *first Q3th data*



What about?

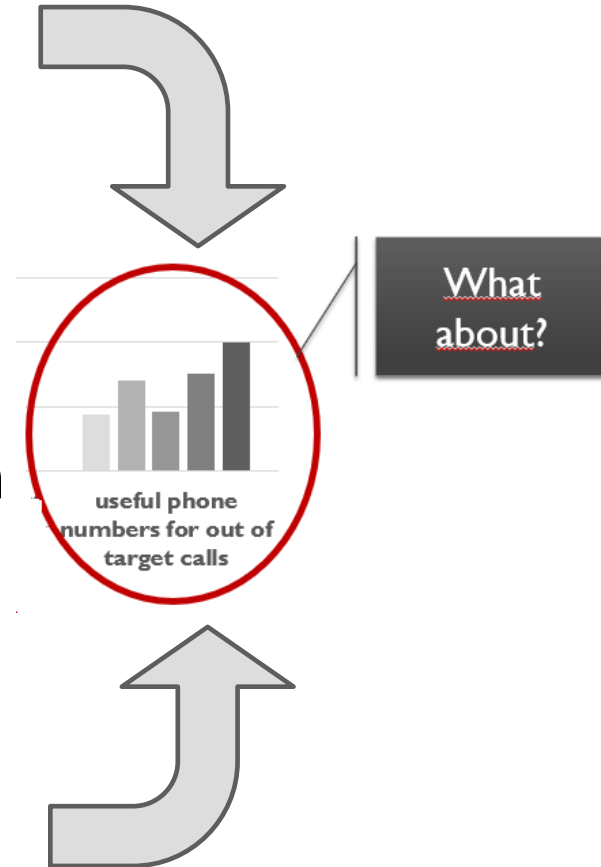
"Pulling" effect of awareness campaigns

2. Adding data-sources on VAW, by using citizen-generated data

- Collecting and using this citizen-generated data on VAW **complements** data collected through administrative sources and dedicated population surveys.
- Timely and accurate because gathered by NGOs skilled professionals of the service, according with “*Improving the collection and use of administrative data on violence against women*” (UN Women and WHO)

HOW TO IMPROVE?

NSO-driven collaboration
Initiated by NSOs and implemented in collaboration with NGO



2. Main aims of the NSO driven-collaboration

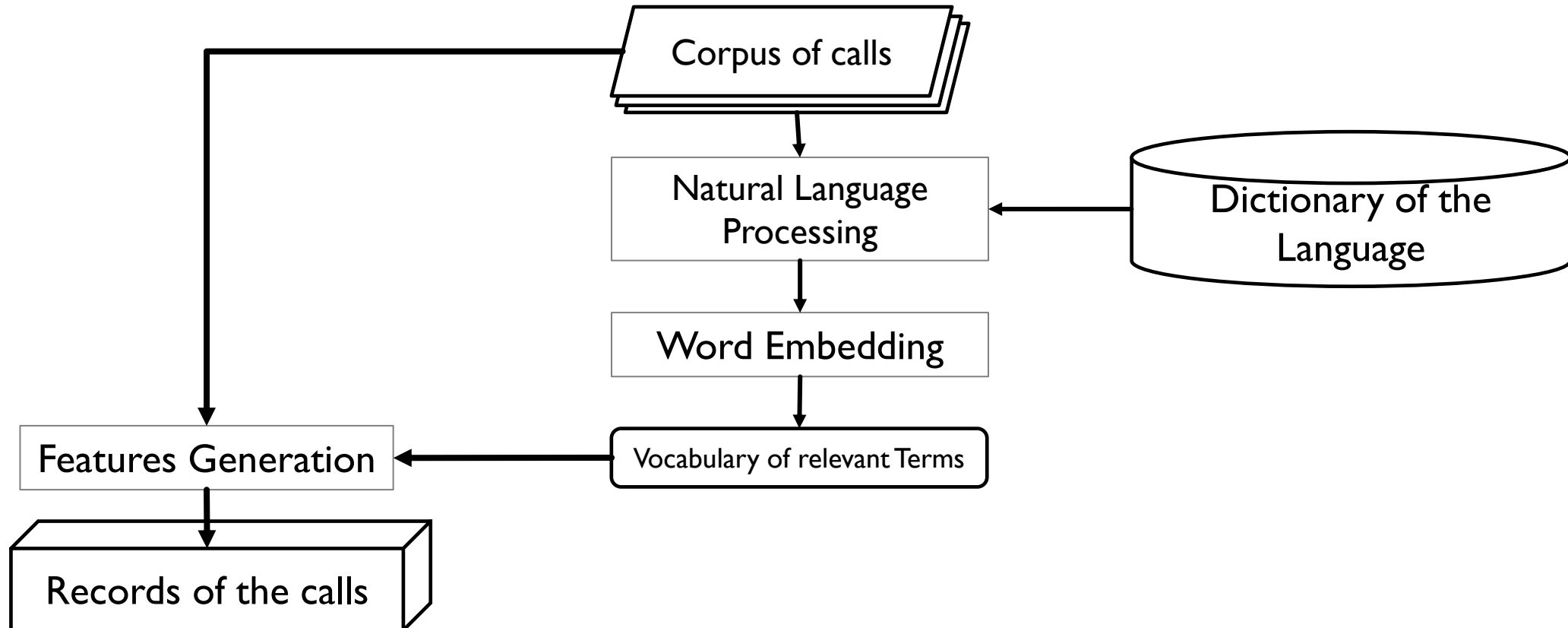
- Monitoring the call's categorization (classification rules) provided by the 1522 NGOs professionals according the domain rules
- Detecting insight and new information target in order to define new classification rules
- Improving the quality of gathering and classification rules process of data from calls.
- Producing a final classification avoiding disambiguation and more representative.
- Supporting data gathering process by using automatic categorization procedures.

3. Methodological approach

- ❑ We propose an **unsupervised text clustering** and **topic extraction framework** capable of achieving high-quality clustering result and extracting cluster-specific topics.
- ❑ The framework includes **three main components**:
 - **Feature Extraction**: to **convert the target texts into vector representations** (text embeddings) by capturing the semantic information, so that it can be processed by the clustering algorithm. It also includes dimensionality reduction techniques to improve data processing.
 - **Call clustering**: to perform a **clusterization** by using **the k -means algorithm**. This algorithm aims at partitioning the calls into k clusters in which each observation belongs to the cluster within nearest mean, being the centroid of the cluster. The “Elbow” method is used to determine the value of k .
 - **Topic extraction**: to find **main topics from each cluster** using the keywords extraction methods TF-IDF (Term Frequency-Inverse Document Frequency).

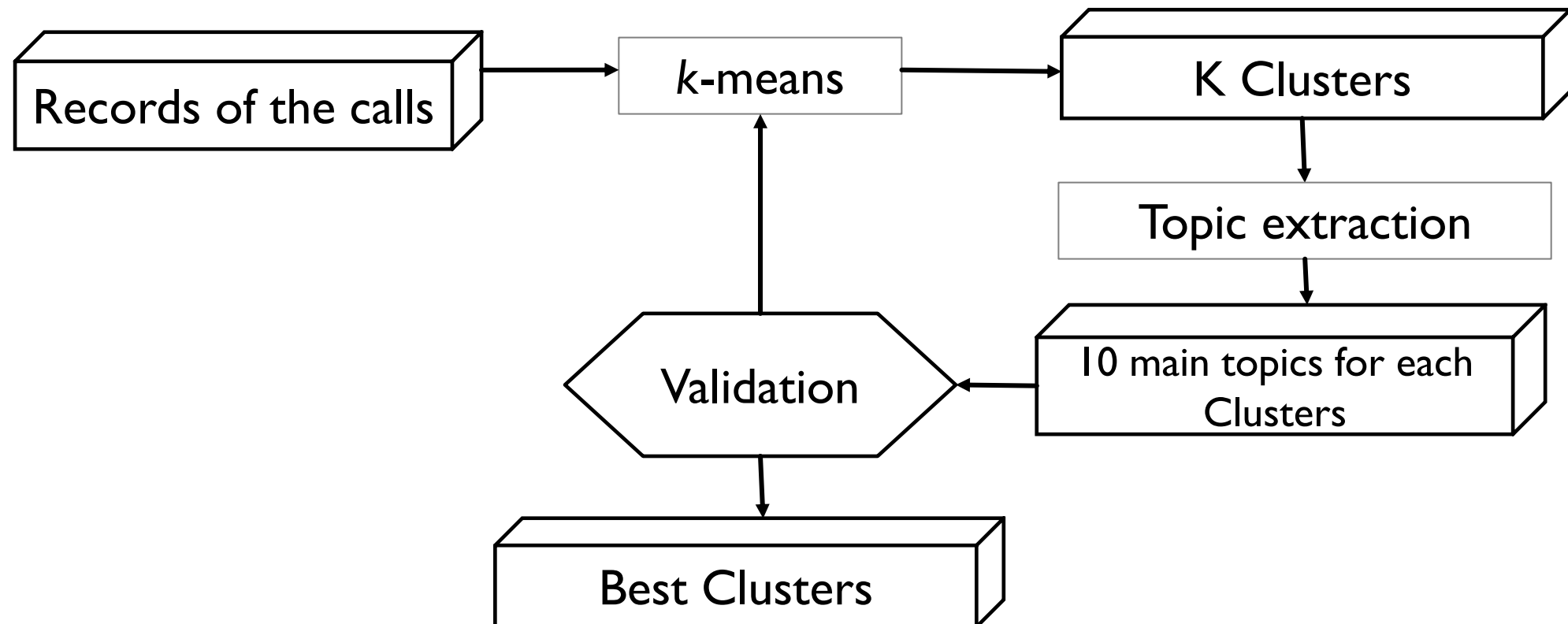
3. Overall scheme of the proposed approach: text mining

- ❑ The **raw text** extracted from **calls** is converted into a **standardized form** (data record) by using **Natural Language Processing** techniques



3. Overall scheme of the proposed approach: machine learning

- ❑ Data records are categorized using the *k*-means clustering algorithm.
- ❑ A *topic extraction method* is applied to extract the best 5 main topics from each cluster.
- ❑ The clusters are validated by domain experts analysing the topics extracted from clusters.



4. Experimental results: real-world data from the Helpline

- ❑ We analyzed the calls received by DEO's 1522 helpline:
 - 12.315 “out of the target” calls;
 - 5.761 “out of service” calls.
- ❑ 38 clusters were produced:
 - 37 clusters validated from domain experts:
 - 1 cluster with unclassifiable calls: i) 13% of “off target” calls; ii) 11% of “out of service” calls.
- ❑ Clusters have been classified in **3 different categories**:
 1. Cluster coherent with the existing target ➡ because reporting **same contents**
 2. Cluster out of the target (not coherent with the helpline services)
 3. Cluster with new needs from the target ➡ identifying **new classification rules**

4. Experimental results: classification rules that improve data storage

Clusters out of the target

- **Cluster 2 topics:** friendly phone, helpline violence against children
- **Cluster 4 topics:** social service, social worker
- **Cluster 6 topics:** extortion victim, suffer extortion money

Clusters coherent with the existing target

Cluster 1 topics: victim violence, rape;
Cluster 3 topics: antiviolence center, cav;
Cluster 5 topics: stalking, stalker;
Cluster 7 topics: sexual harassment, sex maniac;
Cluster 8 topics: free legal aid, free lawyer;
Cluster 14: beating, abuse;
Cluster 20: police, law enforcement;
Cluster 34: report, killing;
Cluster 37: 1522, info.

Clusters with new needs from the target

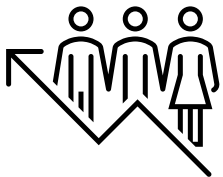
- **Cluster 0 topics:** psychiatric subject, schizophrenic
- **Cluster 19:** mental disorder, bipolarism

Enhancing classifications rules for the helpline, by reducing the “out of the target” calls and adding “mental disease”

5. Next Steps



- Develop a supervised classification model that uses the validated clusters as a training set to learn the classification criteria and predict the class of unlabeled calls.
 - The predicted class corresponds to the association of the call with the reference cluster.
 - This approach would allow all unlabeled calls received by the 1522 helpline to be classified



- Improving the data gathering process of the helpline by adopting a classification model
- Generating timely and quality data from the citizen and NGOs

Thank you!

<https://www.istat.it/en/violence-against-women>

claudia.villante@istat.it