

# Ways of Reducing coverage and sampling error as part of the Total Survey Error Framework for Establishment Surveys in Europe: Recent Developments

Nikola Jovanovski - Sample Solutions BV  
Carsten Broich - Sample Solutions BV

PREMIER. SAMPLE. PROVIDER

Sample Solutions



# CONTENT OVERVIEW

1

## Introduction

Company Background  
European Establishment Surveys  
Big Data as a crucial part in sampling

2

## Research

Sample challenges and Research Problems  
ESRA 2019 Future considerations

3

## Establishment Sample sources

What are the different sources of  
establishment sample data for the EU27

4

## Enrichment Methodology

How can we use a Big Data approach to  
decrease sampling error in European  
Establishment Surveys

5

## Pros and Cons

Benefits and limitations from using the  
enrichment approach

6

## Conclusion

Research Problem address  
Q&A

# 1. INTRODUCTION

---

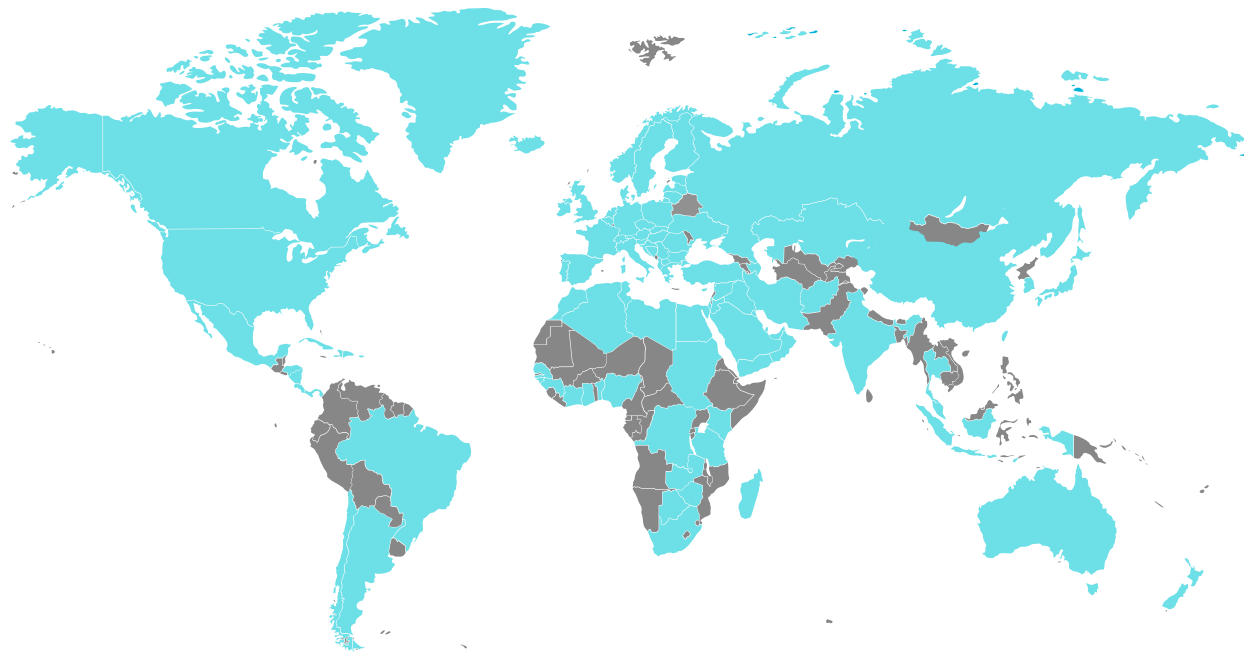
3

PREMIER. SAMPLE. PROVIDER

Sample Solutions



## About Sample Solutions



## Background

Founded in the Netherlands, back in 2009  
with the focus on Business & Consumer  
Telephone Sample

## Sample Survey Platform B2B module

Specialized B2B Database designed Survey  
Research -instant counts and sample  
ordering

## Multi country B2B sample Projects

Eurobarometer, London Economics, PwC,  
ABB, World Bank, American Express

# WHAT IS BIG DATA?

---

PREMIER. SAMPLE. PROVIDER

Sample Solutions





# European Establishments Surveys:



Flash Eurobarometer (business survey part)



European Companies Survey (ECS)



European Survey of Enterprises on New and Emerging Risks  
(ESENER)



# Sampling frame sources per country:

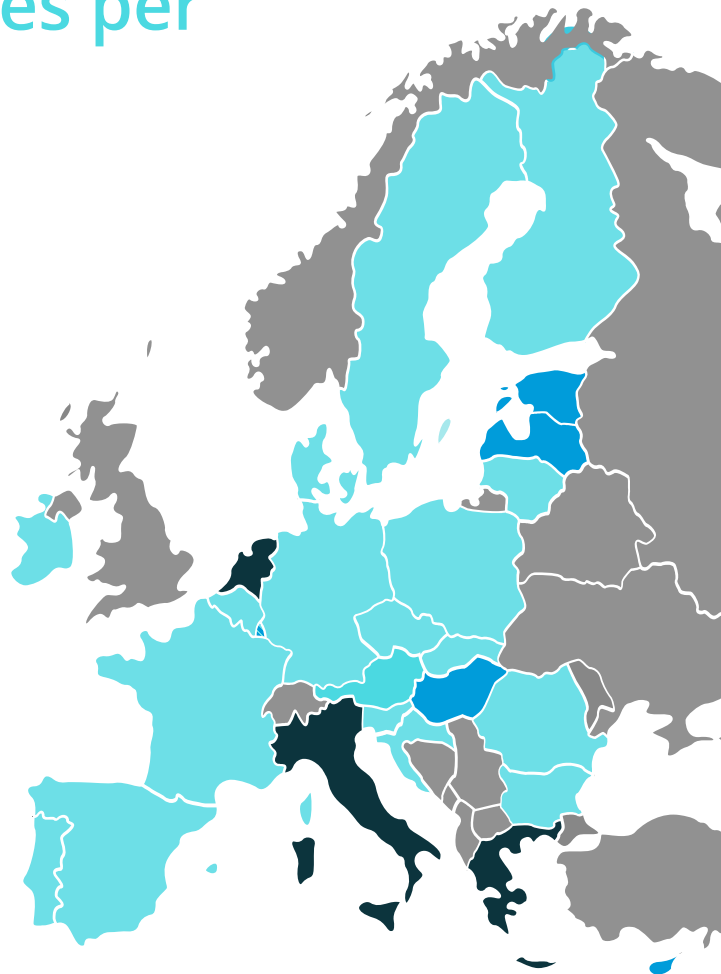
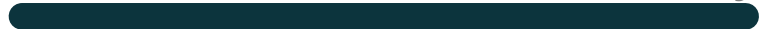
B2B data vendors (finance,  
marketing,sales purposes)



National Business Register



Chamber of Commerce Directory





## 2. RESEARCH

---

PREMIER. SAMPLE. PROVIDER

Sample Solutions

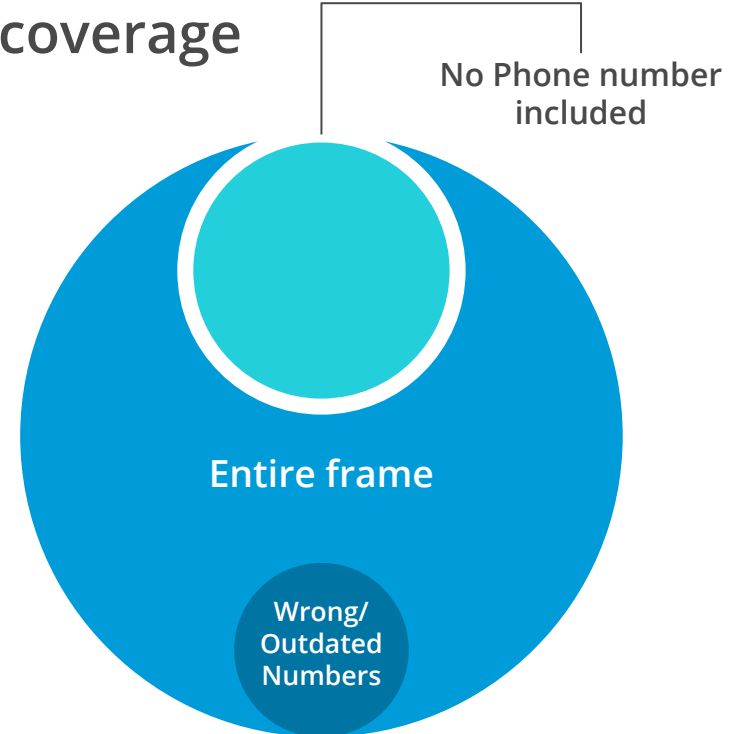


# Sampling Challenge

~60% telephone coverage

## Issues among the Sampling sources

- Decentralized Sample data vending (1 vendor per country)
  - ◆ Under representativeness among certain:
    - Industries
    - Company Sizes
    - Area typologies
    - Subsidiaries(Branches)
  - ◆ Variable sample costs among countries
  - ◆ Limited coverage of Contact details of decision makers and officers
  - ◆ No or limited coverage of email addresses



Source: Business Count that includes entire frame with and without phone number records

# Sample comparison: Phone v Email

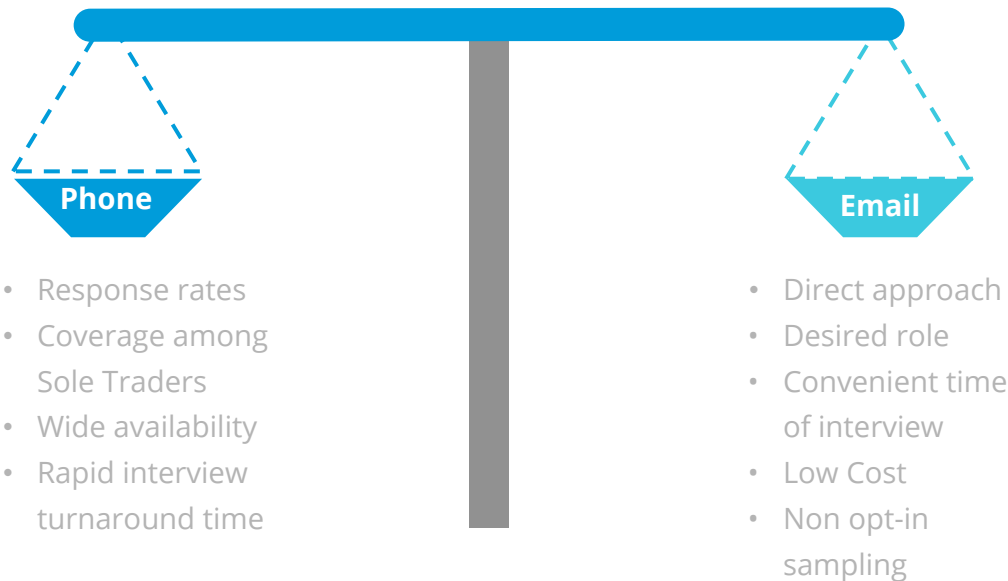
## Establishment sample

### → Offline - phone

- ◆ Key breakdown values (Industry code, Size)
- ◆ No direct contact point with decision maker
- ◆ Overall better response rates

### → Online - email

- ◆ Key breakdown values (Industry code, Size)
- ◆ Direct email of Decision Maker
- ◆ Targeting specific job roles based on research



# Research questions

---

What are the ways of reducing coverage and sampling error for Establishment

## Surveys in EU27:

1. How can Big Data be used for increasing the telephone coverage in the Sampling Frame
2. How can the Sampling frame go online, for a mixed mode sampling approach?
3. How can we ensure sample data accuracy and validity in a international and longitudinal study?

# Future considerations:

---

**From ESRA 2019 we have posted these future action steps:**

- Use Big Data to Identify Company Size(no. of Employees) on businesses with unknown size
- Include a validation step in the enrichment process to lower data inaccuracy
- Include a Area Typology stratification criteria in order to make a more representable sample to frame
- Design sample of all countries from a centralized source
- Source Contact person data and email

### 3. ESTABLISHMENT SAMPLE SOURCES

---

14

PREMIER. SAMPLE. PROVIDER

Sample Solutions



# ESTABLISHMENT SAMPLE ELEMENTS:

15

01	LOCATION	<ul style="list-style-type: none"><li>Country, City, Postcode, Region, HQ Location, Branches Location</li></ul>
02	INDUSTRY	<ul style="list-style-type: none"><li>SIC code, Keywords (from LinkedIn), NAICS code, NACE code</li></ul>
03	COMPANY SIZE	<ul style="list-style-type: none"><li>Employee Range</li><li>Revenue Range</li></ul>
04	COMPANY STATUS	<ul style="list-style-type: none"><li>Founded date</li><li>Operating status, Entity Type</li></ul>
05	CONTACT OPTIONS	<ul style="list-style-type: none"><li>Email available, Phone available, Website URL, Decision maker (C-level)</li></ul>
06	TECHNOLOGY	<ul style="list-style-type: none"><li>Technologies divided into categories</li></ul>
07	CONTACTS INFO	<ul style="list-style-type: none"><li>Departments, Job title</li><li>C-level, VP level, Director, Manager, Non - Manager</li></ul>

# Three sources for Establishment sample sources:

---

## TRADITIONAL



**Available lists**

**Basic Business  
information**

## ENRICHMENT



**Online lookups**

**Big Data**

## GENERATED



**Email patterns**

**Verification of email  
activity**



# Traditional sources

## National Registers

State or Privately run  
Categorized with NACE and SME  
sizes



## Chamber of Commerce

Rich in business information  
Can be Member only



## Phone book directories

Open and free access for  
general public



## Commercial Data

### vendors

Paid access  
For Marketing, Credit report  
and Sales purposes  
Rich in detail



# CHARACTERISTICS OF TRADITIONAL SOURCES:

BRAND	MANDATORY	FREE ACCESS	EMAIL ADDRESSES	CONTACTS
National Business Register	✓ *	✓ *	✗	✗ *
Phone book	✗	✓	✗ *	✗
Commercial Data Vendor	✗	✗	✓	✗ *
Chamber of Commerce	✗ *	✗ *	✓ *	✓ *

*\*Applies to majority of cases in the EU27*

# BIG DATA - ENRICHMENT SOURCES



# Enrichment sources and tools:

## Search

### Engine

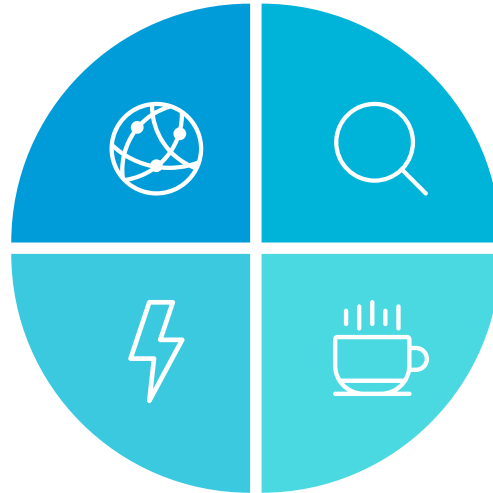
- Google Snippet
- Company name and address search
- Email address search
- Geographic lookup

## Social Media

- Facebook
- Phone lookup

### Linkedin

- Contact and Emp. size lookup



## Directory lookups

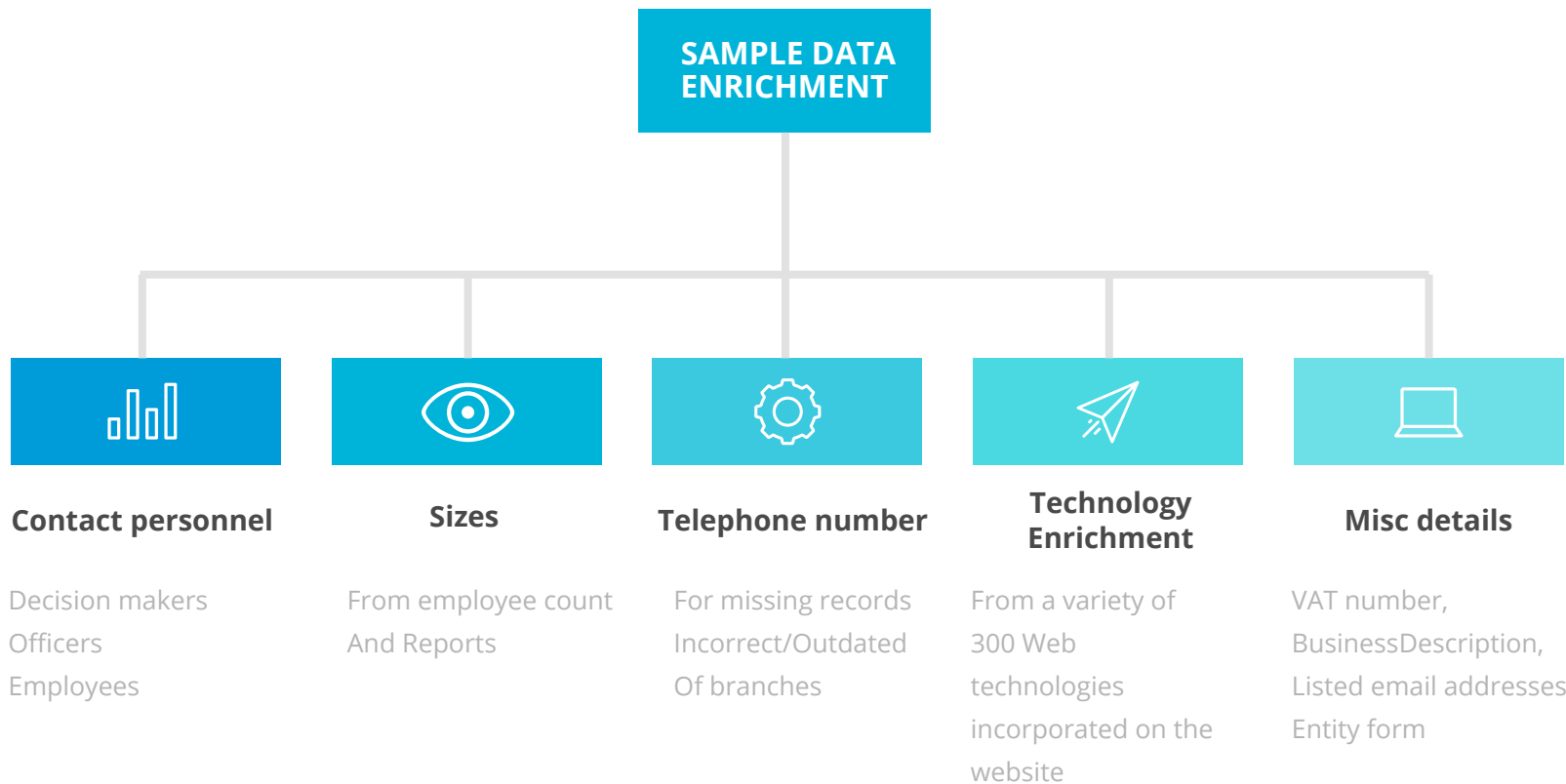
- Reverse phone lookups
- Company Lookups( Company Name + Address)
- Contact person lookup

## Review Sites

- Company lookup
- Geographic lookup

# SAMPLE DATA ENRICHMENT

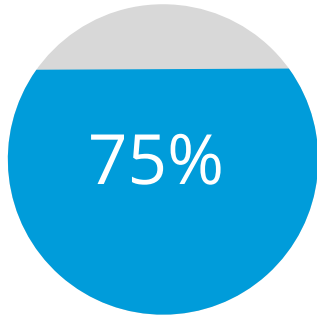
21



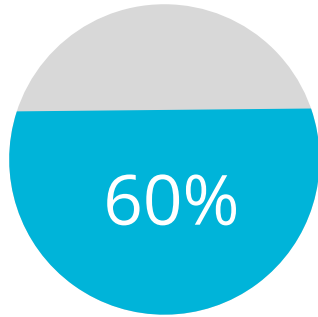
## Which source can produce which data

FEATURES	Employee size	Phone number	Contact person	Email Address
LinkedIn	✓	✗	✓	✗
Google	✗	✓	✓	✓
Facebook	✗	✓	✓	✓
Directory lookup	✓	✓	✓	✓
Yelp & Tripadvisor	✗	✓	✗	✓

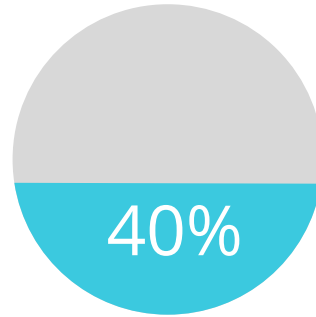
## Overall Enrichment rates among elements



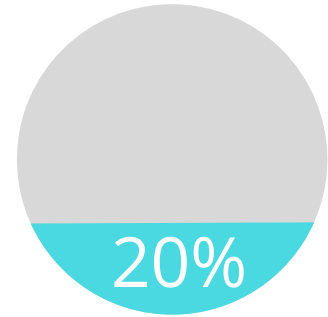
**Phone  
number**



**Employee size**

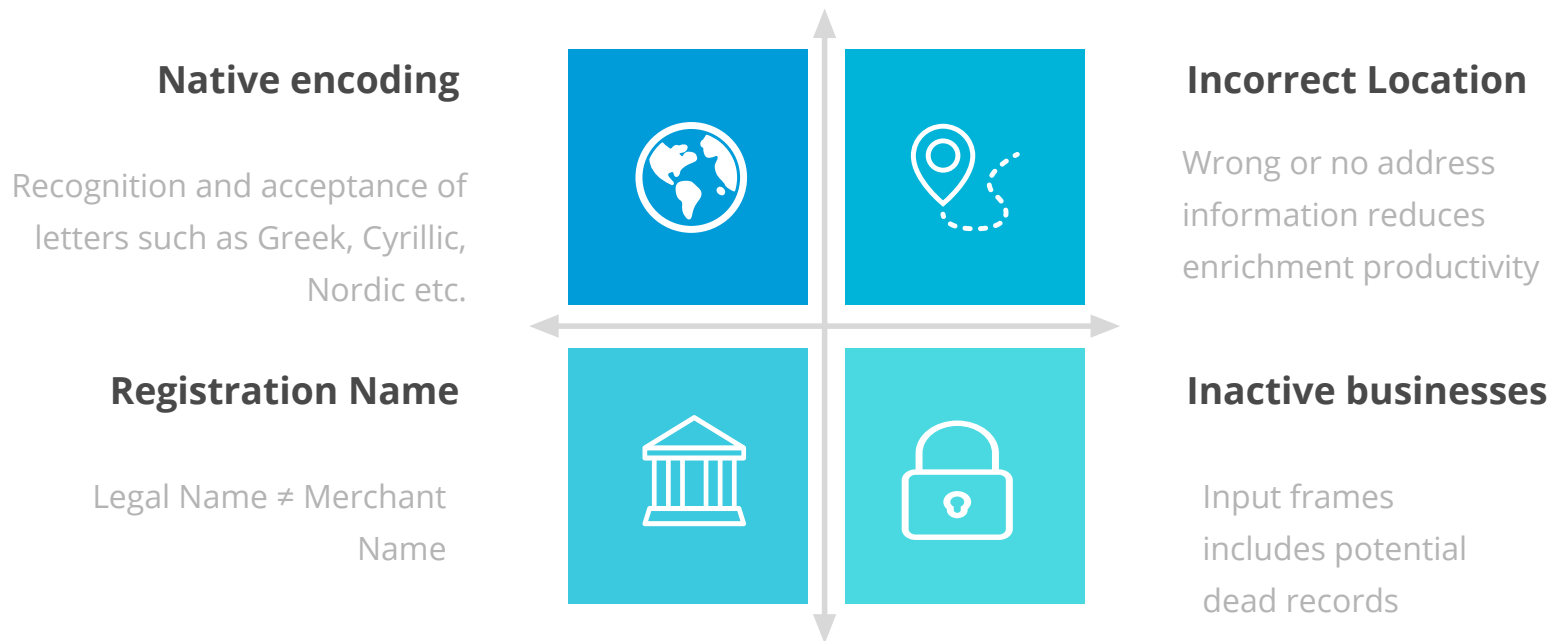


**Contacts**



**Email address**

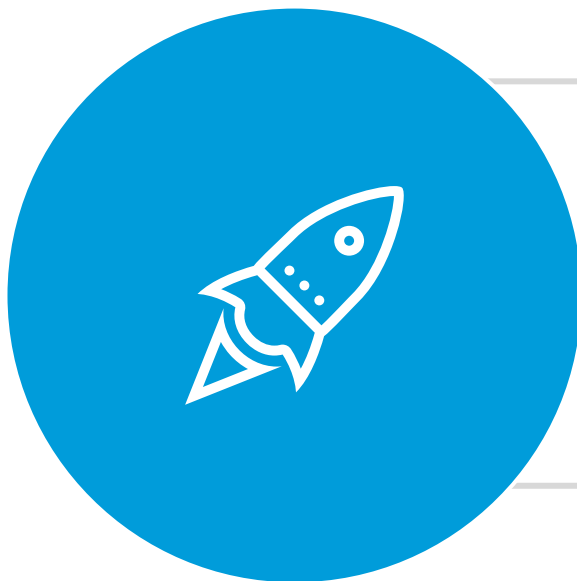
# Challenges in Big Data enrichment





# GENERATED SAMPLE SOURCE: EMAIL ADDRESS SAMPLING METHODOLOGY

25



## Input Data

Sourced Company URL to create email domain  
Contact person Name



## Generating Patterns

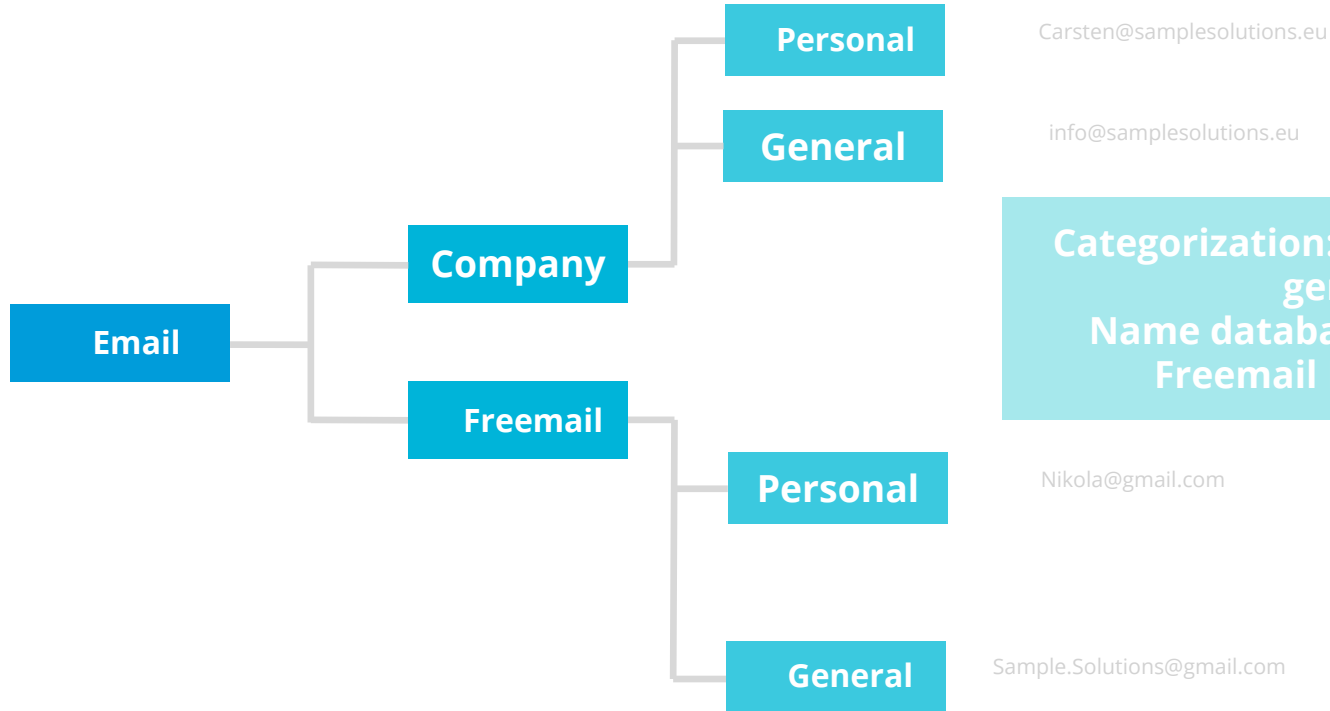
Apply input data to the pattern  
Combinations eg. [firstname@company.com](#)  
[f.lastname@company.com](#)  
[firstname.lastname@company.com](#)



## Validation of email activity

Screen all combinations using a SMTP service to see whether the email server recognizes the email addresses

# Email Categorization

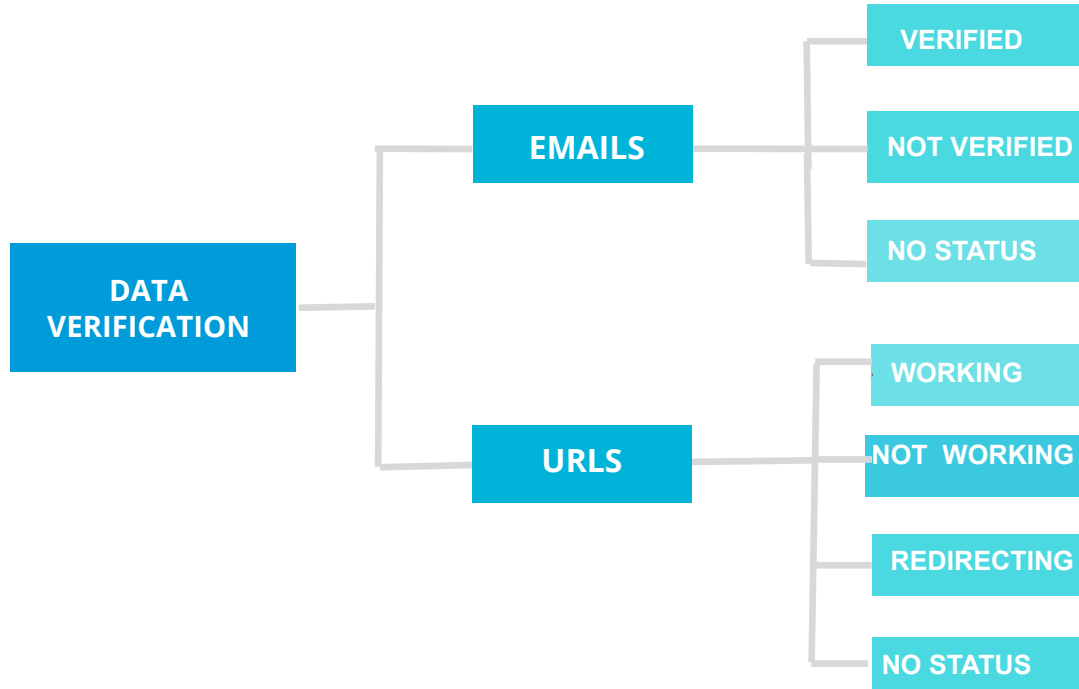


**Categorization: Lookup table for  
general  
Name database for personal  
Freemail Provider list**

# SAMPLE DATA VERIFICATION

---

27



## 5. ENRICHMENT METHODOLOGY

---

PREMIER. SAMPLE. PROVIDER

Sample Solutions



# Outline

European enterprises and establishments sampling featured in both online and offline sampling modes

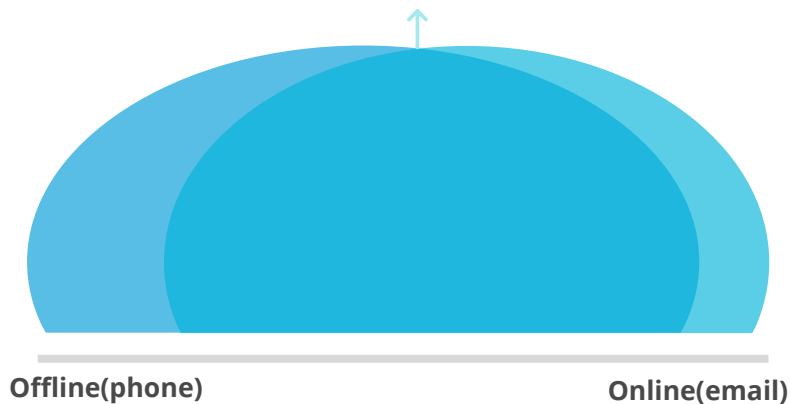
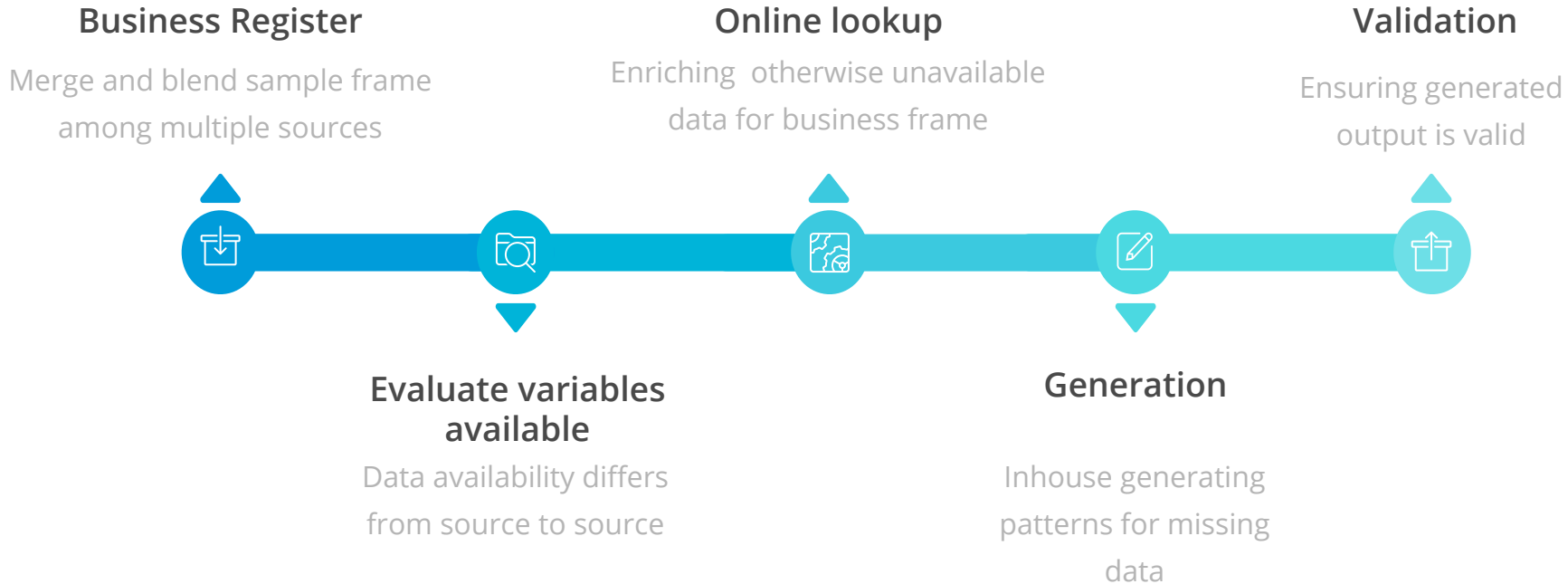


Figure 2. Estimation of Coverage between offline and online modes EU27

## Full Sampling method

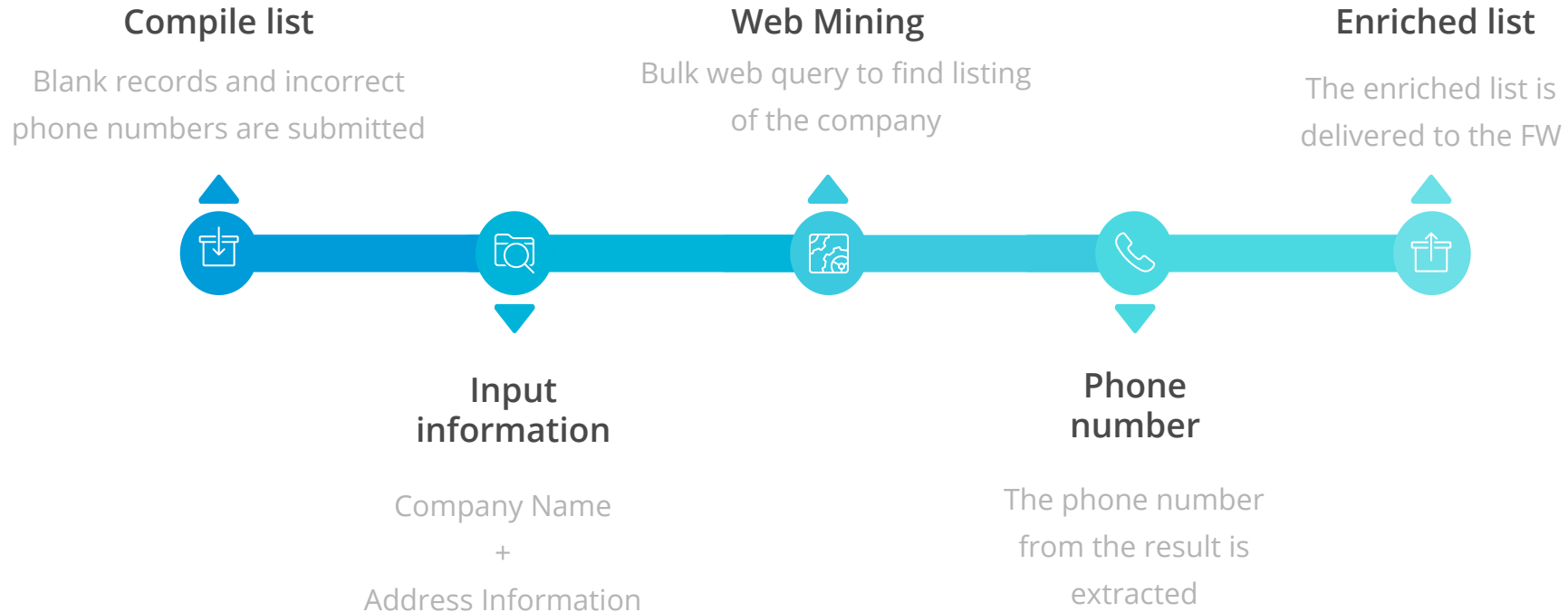
- Mixed Mode Sample (Online and Offline): defined by:
  - ◆ Direct and/or Company Email
  - ◆ Company phone
- Merging and blending multiple business sample sources for a country into one frame ensuring maximum coverage among the EU27
- Collected email address, contact person data and phone number information from publicly available sources(URLs, Social Media etc)
- Generate Email addresses where not available
- Ensure Generated vs. Collected email data (compliance)

# Full Sample Enrichment

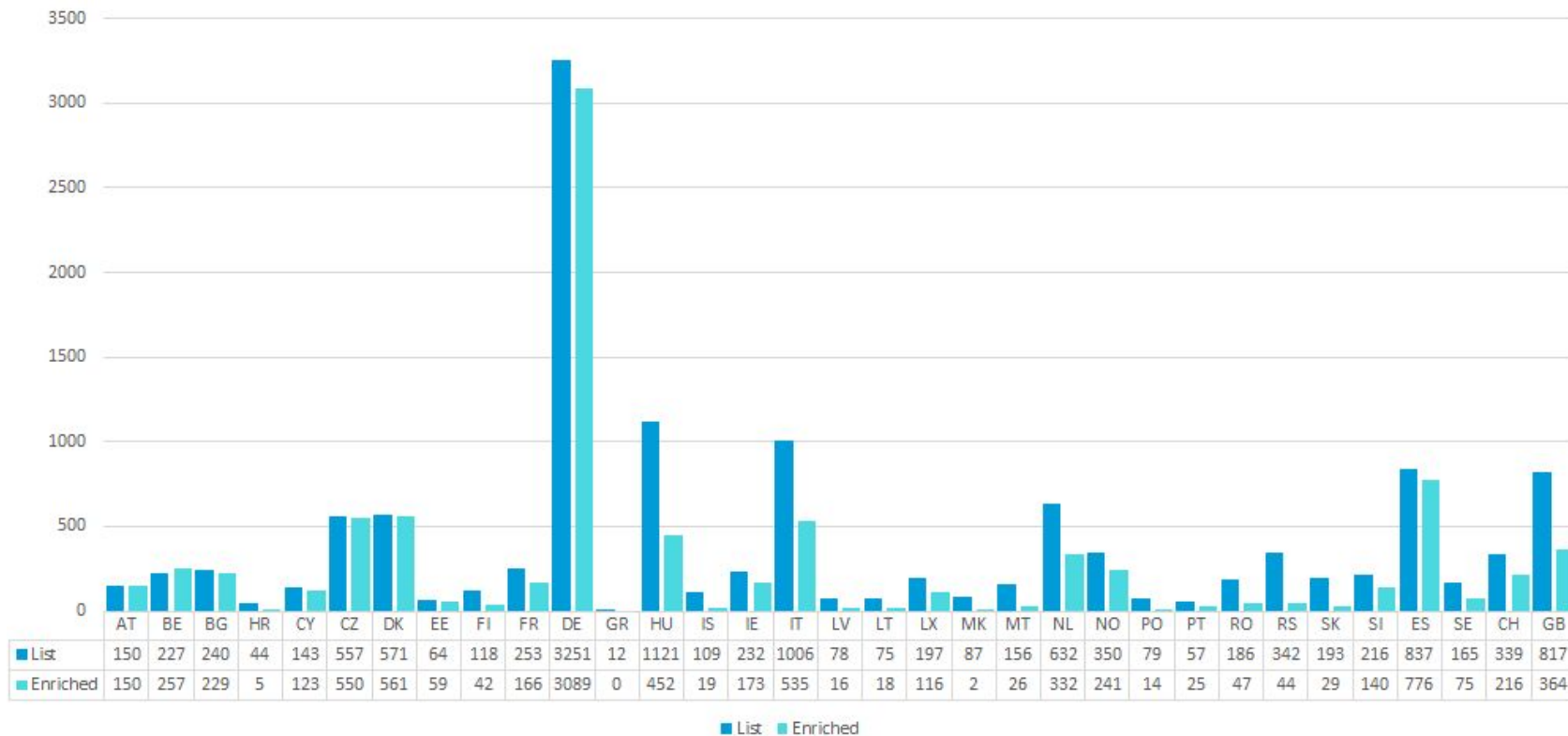


# Enriching B2B records with phone numbers

31

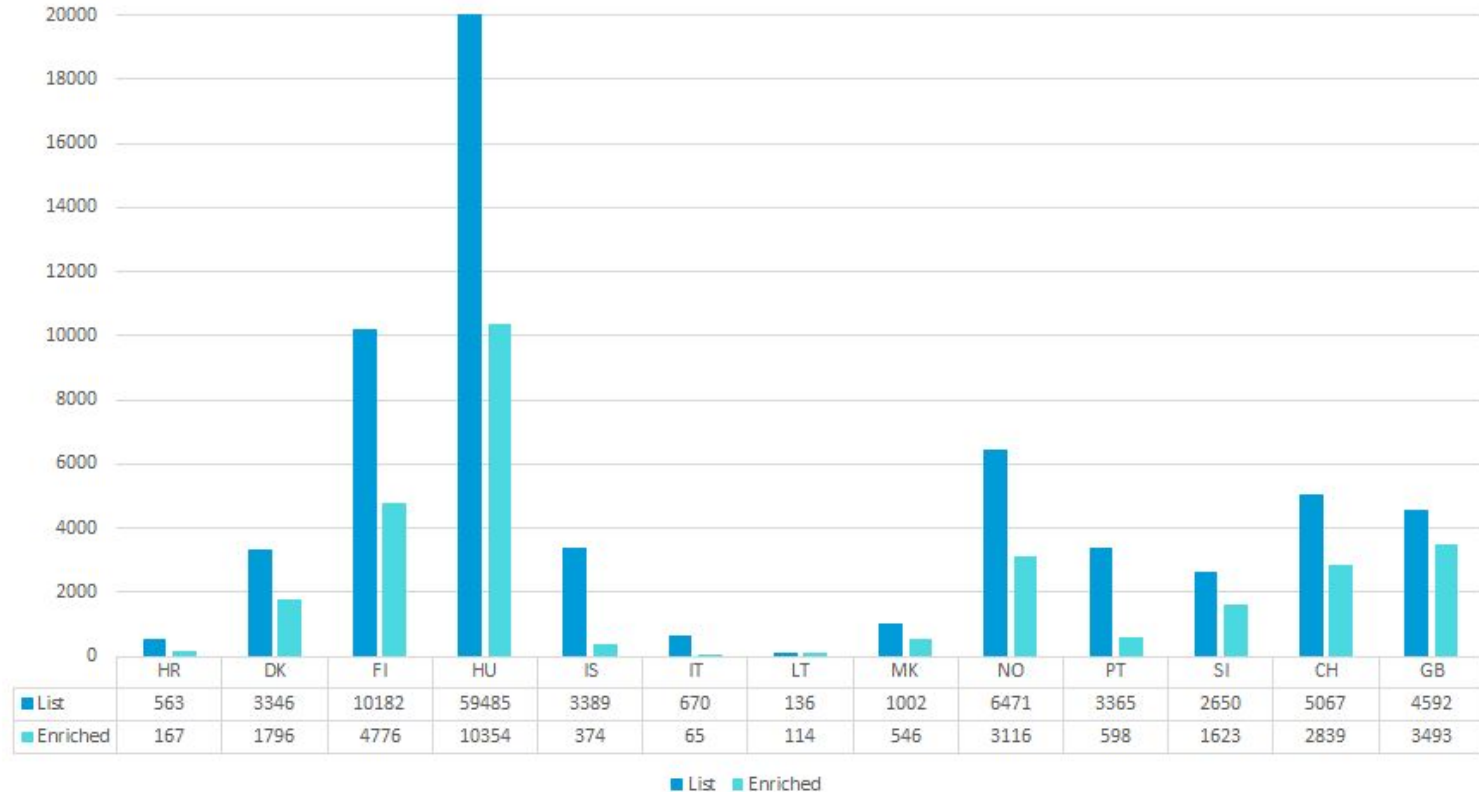


# ESENER-3 Results: enrichment of wrong telephone data

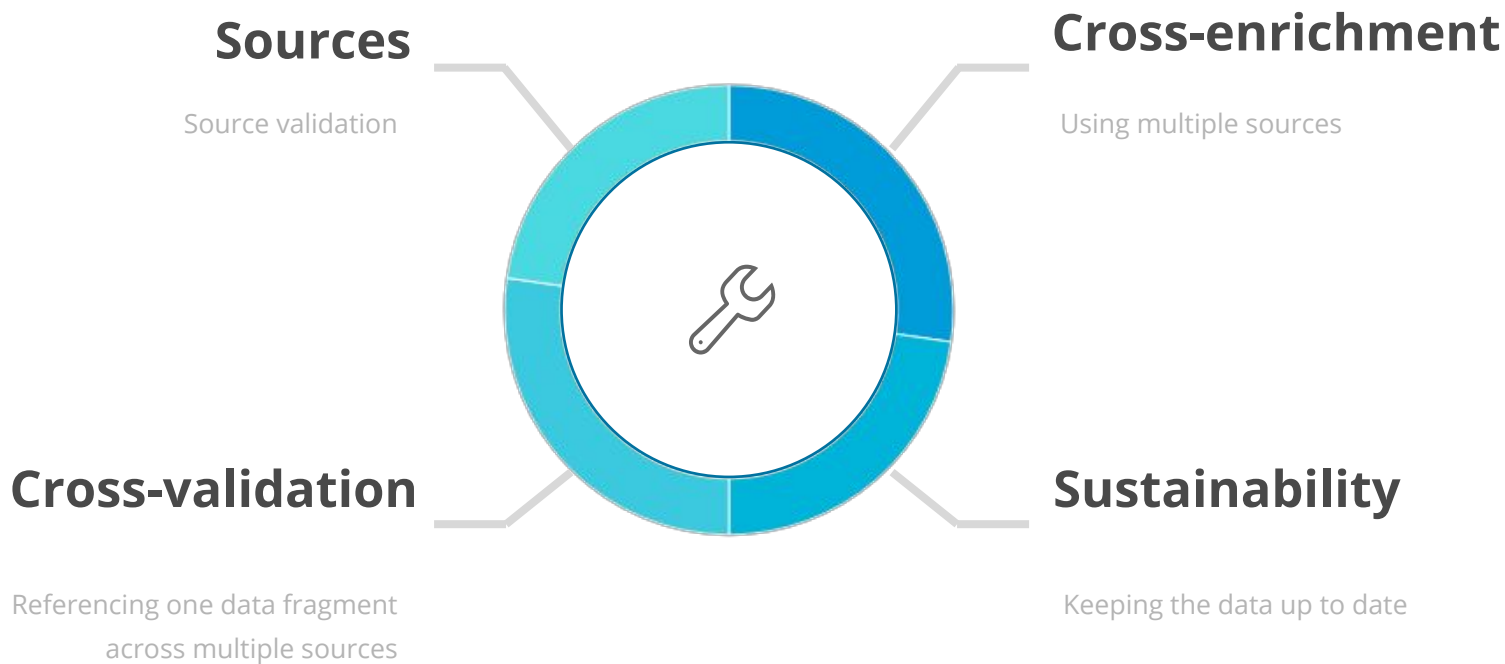




# ESENER-3 Results: Enrichment of without phone



# Strategies for limiting sample Coverage & Bias



\*Weighting is possible but might get too statistical (adapting selection probability)

## 5. PROS AND CONS:

PREMIER. SAMPLE. PROVIDER

Sample Solutions



# BENEFITS

---

## DETAILED RICH SAMPLE DATA

Data contains more data elements than any other individual source

## Verification

Data accuracy is validated by multiple sources

## Sample going Online

Combining phone and online in survey fieldwork

## Centralized source for EU27 data

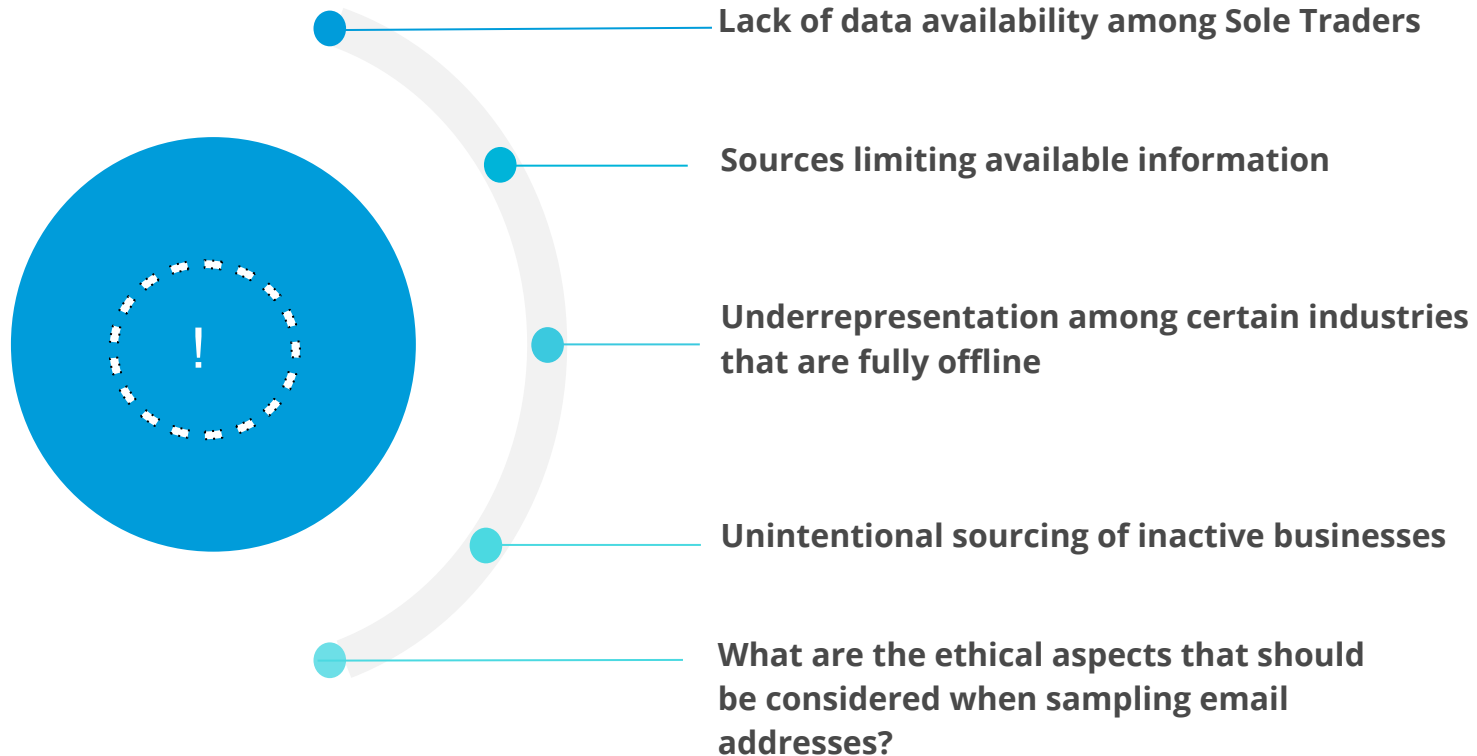
One stop shop for comprehensive data in all 27 countries

## Increased coverage

More sample data can be accessed via phone or email



## B. Limitations/Possible Issues



# CONCLUSION

PREMIER. SAMPLE. PROVIDER

Sample Solutions



# Research questions

What are the ways of reducing coverage and sampling error for Establishment

## Surveys in EU27:

1. How can Big Data be used for increasing the telephone coverage in the Sampling Frame
2. How can the Sampling frame go online, for a mixed mode sampling approach?
3. How can we ensure sample data accuracy and validity in a international and longitudinal study?

# Future considerations:

---

From ESRA 2019 we have posted these future action steps:

- Use Big Data to Identify Company Size(no. of Employees) on businesses with unknown size **(COMPLETED)**
- Include a validation step in the enrichment process to lower data inaccuracy **(COMPLETED)**
- Include a Area Typology stratification criteria in order to make a more representable sample to frame **(PENDING)**
- Design sample of all countries from a centralized source **(PENDING)**
- Source Contact person data and email **(COMPLETED)**



# SAMPLE SOLUTIONS

WWW.SAMPLE.SOLUTIONS



## ADDRESS

Stationsplein 45  
3013 AK Rotterdam  
The Netherlands



## E-MAIL

[info@sample.solutions](mailto:info@sample.solutions)



## TELEPHONE

Direct: 123-4567-9898  
Fax: 123-4567-9898



## WEBSITE

[www.sample.solutions](http://www.sample.solutions)

PREMIER. SAMPLE. PROVIDER

Sample Solutions

