

Using Big Data to sample minorities in Western European Countries: Muslims

2018 CSDI Workshop, Limerick, Ireland

Premier. Sample. Provider.
Sample Solutions



Overview

1

Introduction

Background topic & purpose

2

Literature Review

Exploration and analysis of previous studies

3

Methodology

Sampling setup

4

Fieldwork & Sampling Results

Lessons learned from sample generation & fieldwork results

5

Conclusions

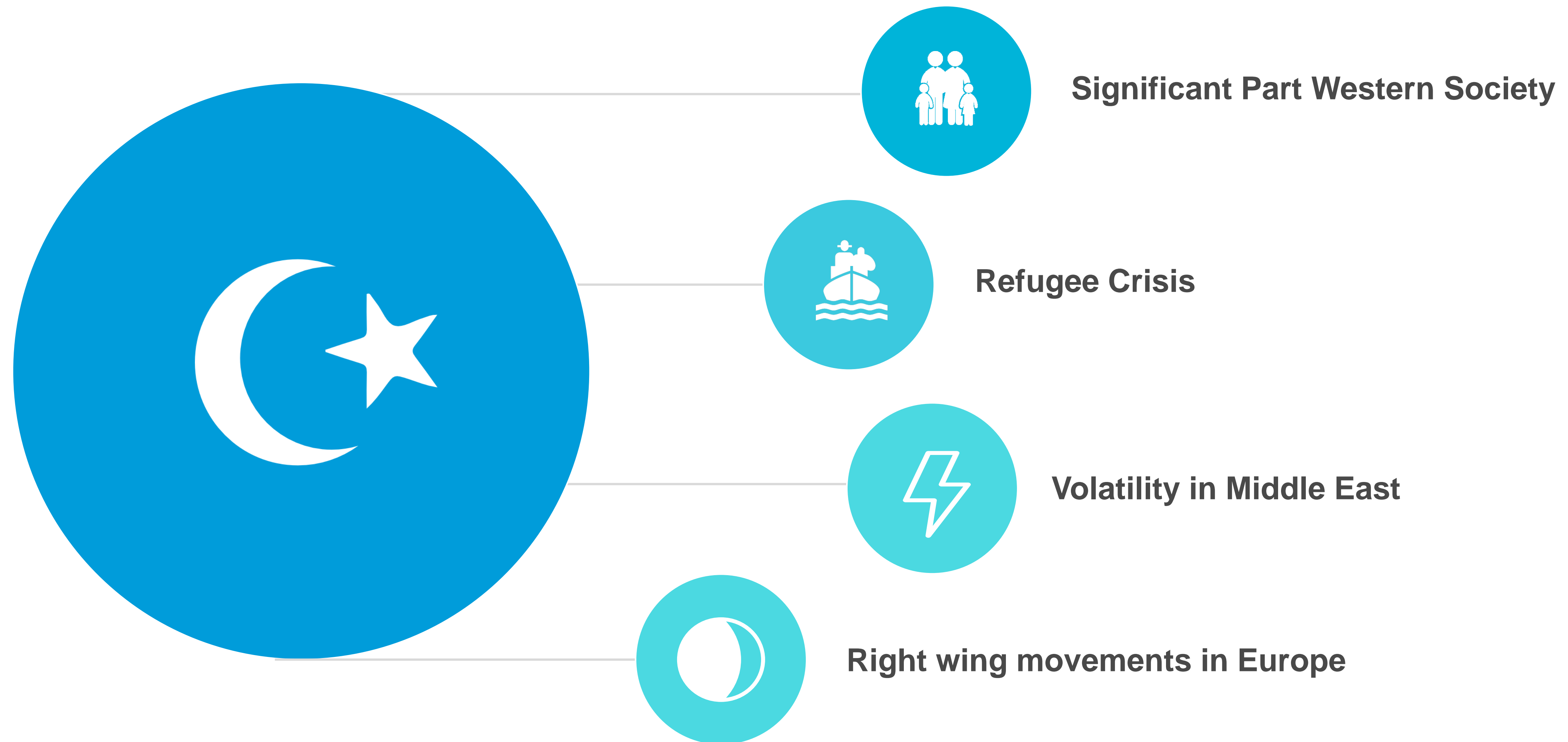
Future designs and possibilities



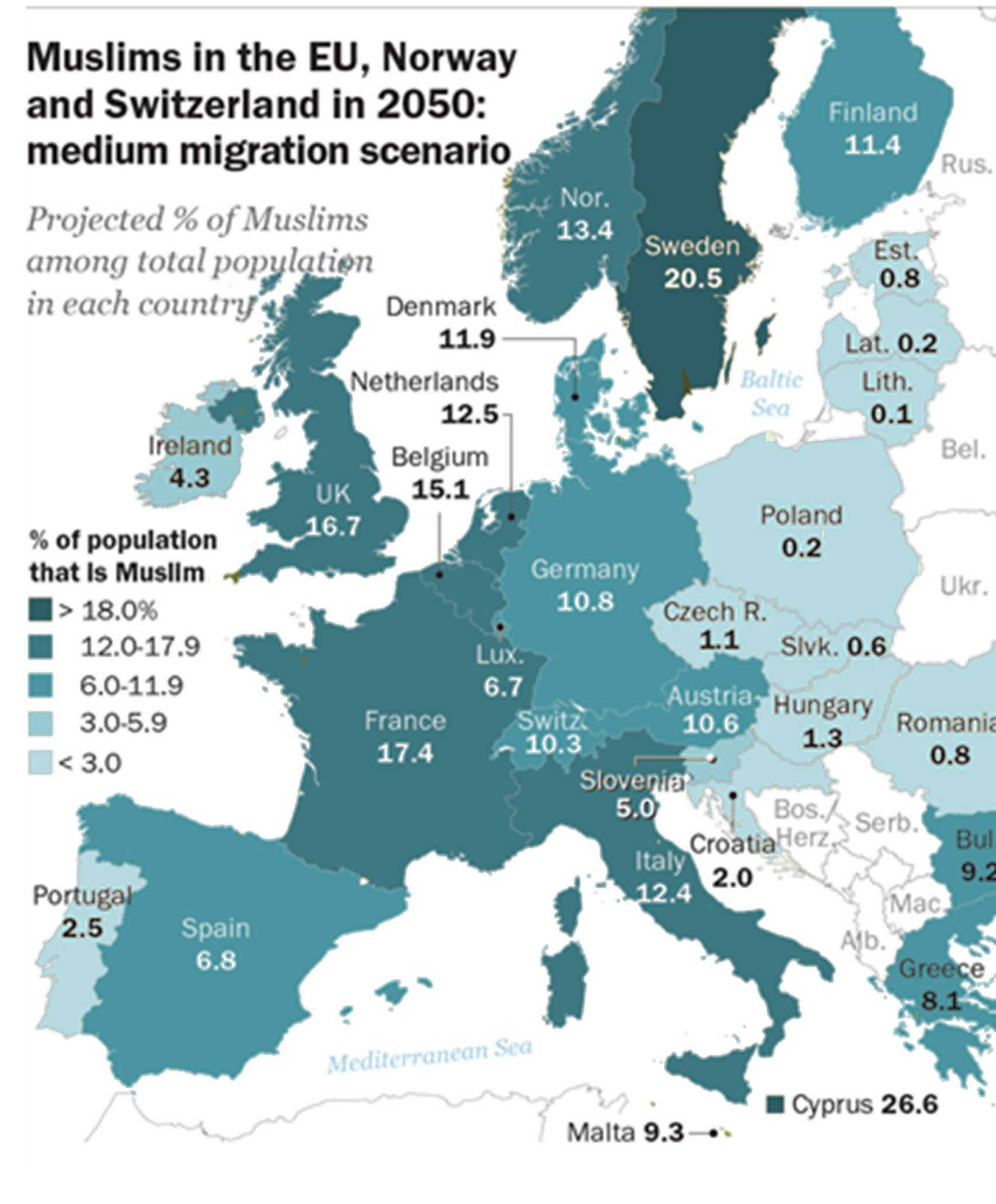
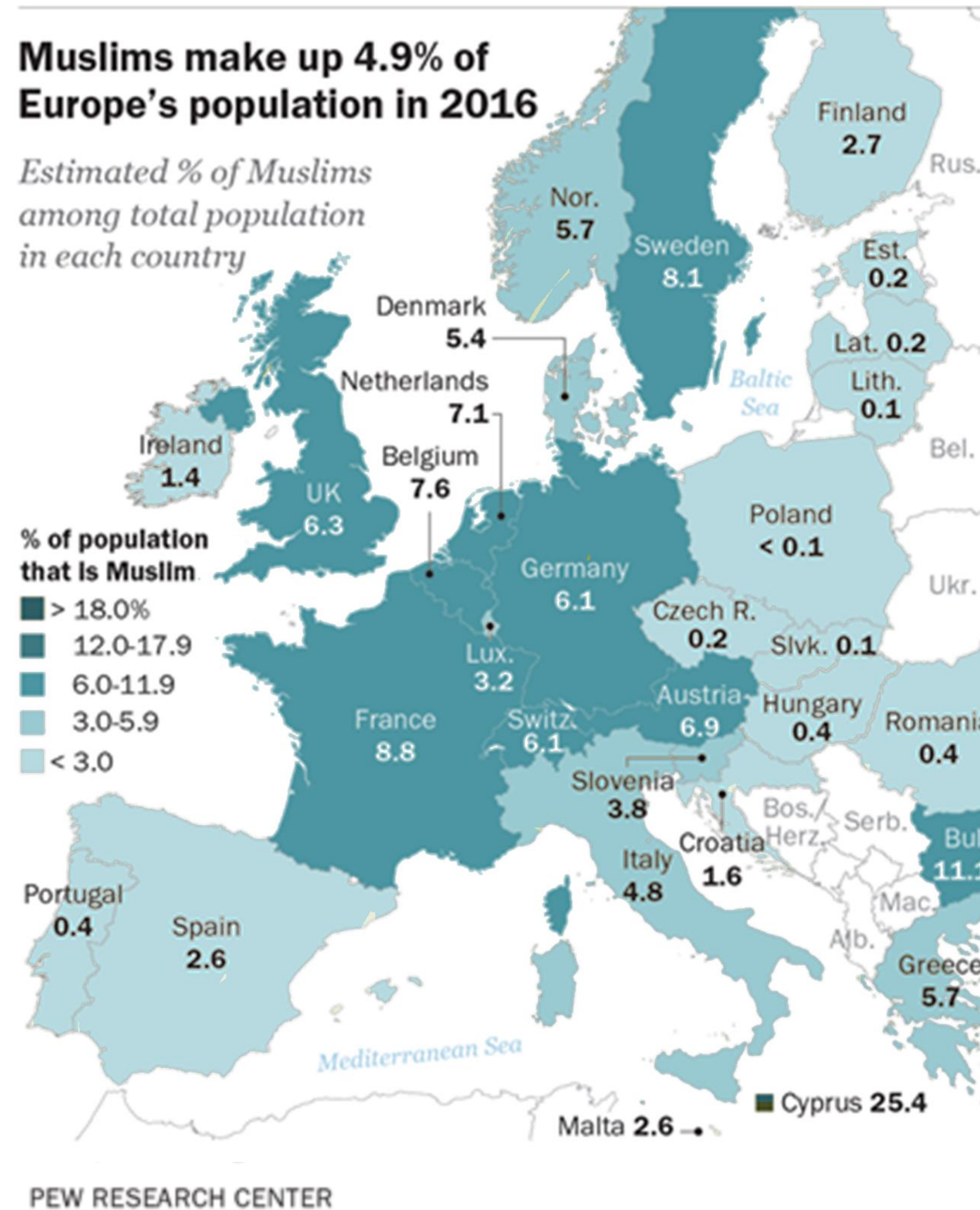
1.Introduction

**“We would like to
conduct a NatRep
study of the Muslim
Population in
Countries X, Y & Z”**

What creates interest?



What creates interest?



Sampling Challenge

1

Target Population

Registered, self-reported?

2

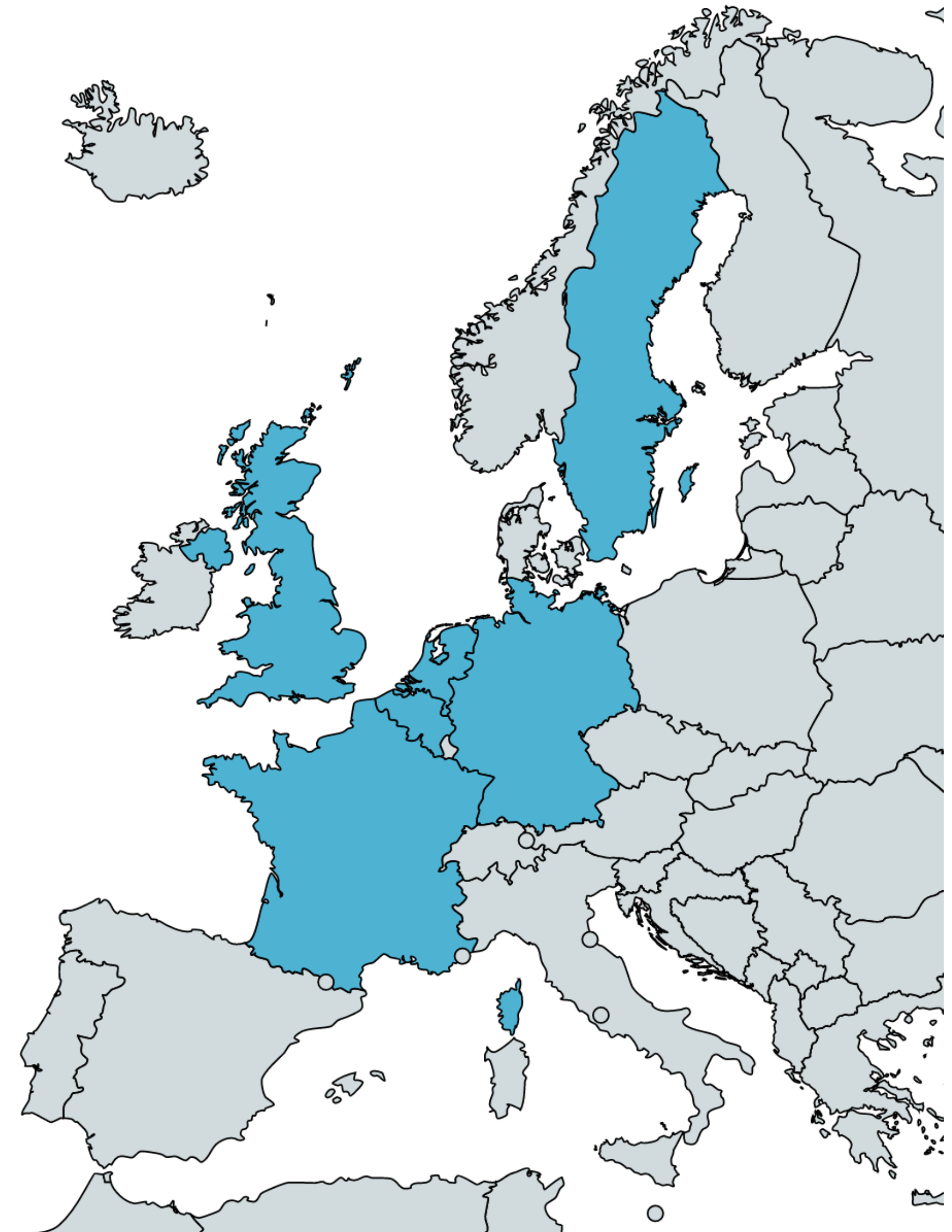
Multi-Country

Comparable results in multiple countries

3

“Representative”

Is there data on Muslim populations? If so, on which level of detail? Weighting-Targets?



2.Literature Review

Previous Approaches




September 2017

Muslims living in the EU face discrimination in a broad range of settings – and particularly when looking for work, on the job, and when trying to access public or private services. The report examines how characteristics – such as an individual's first and last name, skin colour and the wearing of visible religious symbols like a headscarf, for example – may trigger discriminatory treatment and harassment.

Downloads:

Second European Union Minorities and Discrimination Survey (EU-MIDIS II): Muslims - Selected findings

 [en](#) (1.53 MB)

 INSTITUT
MONTAIGNE

**Un islam français
est possible**

RAPPORT - SEPTEMBRE 2016



ICMUnlimited...



DIRK HALM UND MARTINA SAUER

Muslime in Europa

Integriert, aber nicht akzeptiert?

2017, 72 pp (PDF)

DOI 10.11586/2017029

[Download](#)

Free of charge

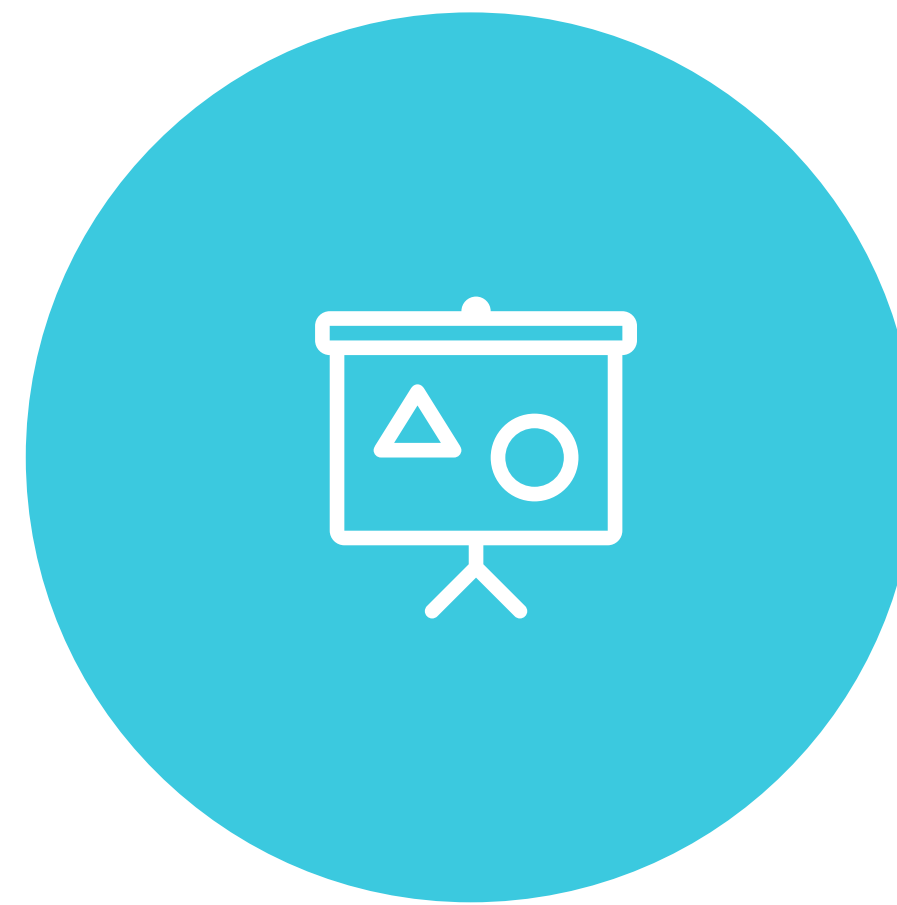
Previous Approaches



White Pages



Targeted RDD Sample



Specific Providers



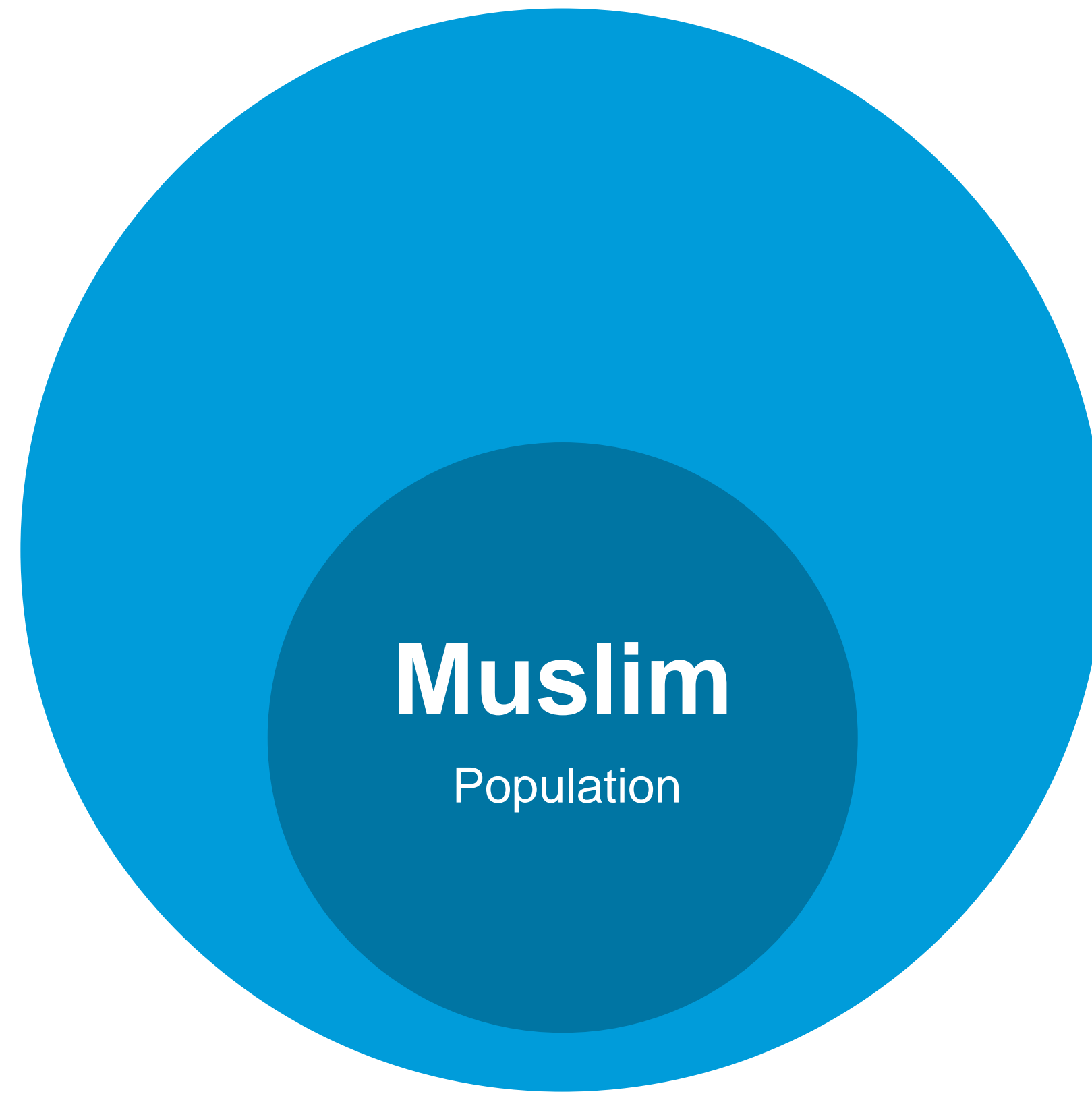
Unknown Methodology



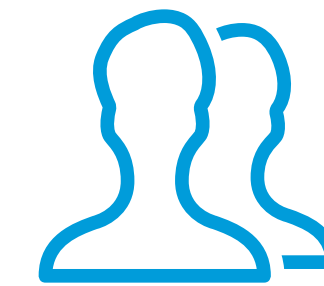
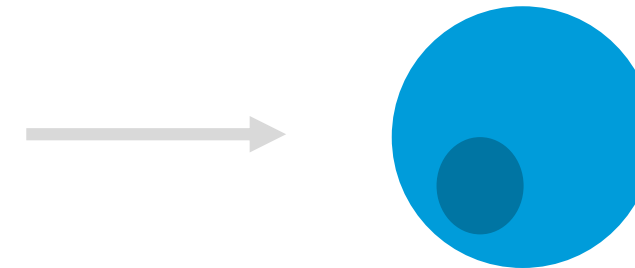
Landline Based

Full Probability Sample

Dual-Frame



Drawn Sample



~3-5%

Incidence Rate

Targeted RDD Sample



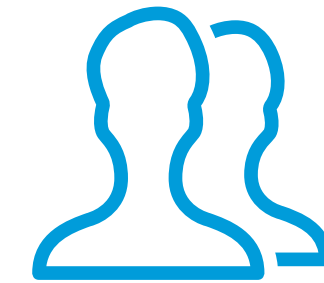
↑ **13.3%**

Rotterdam



↑ **14.1%**

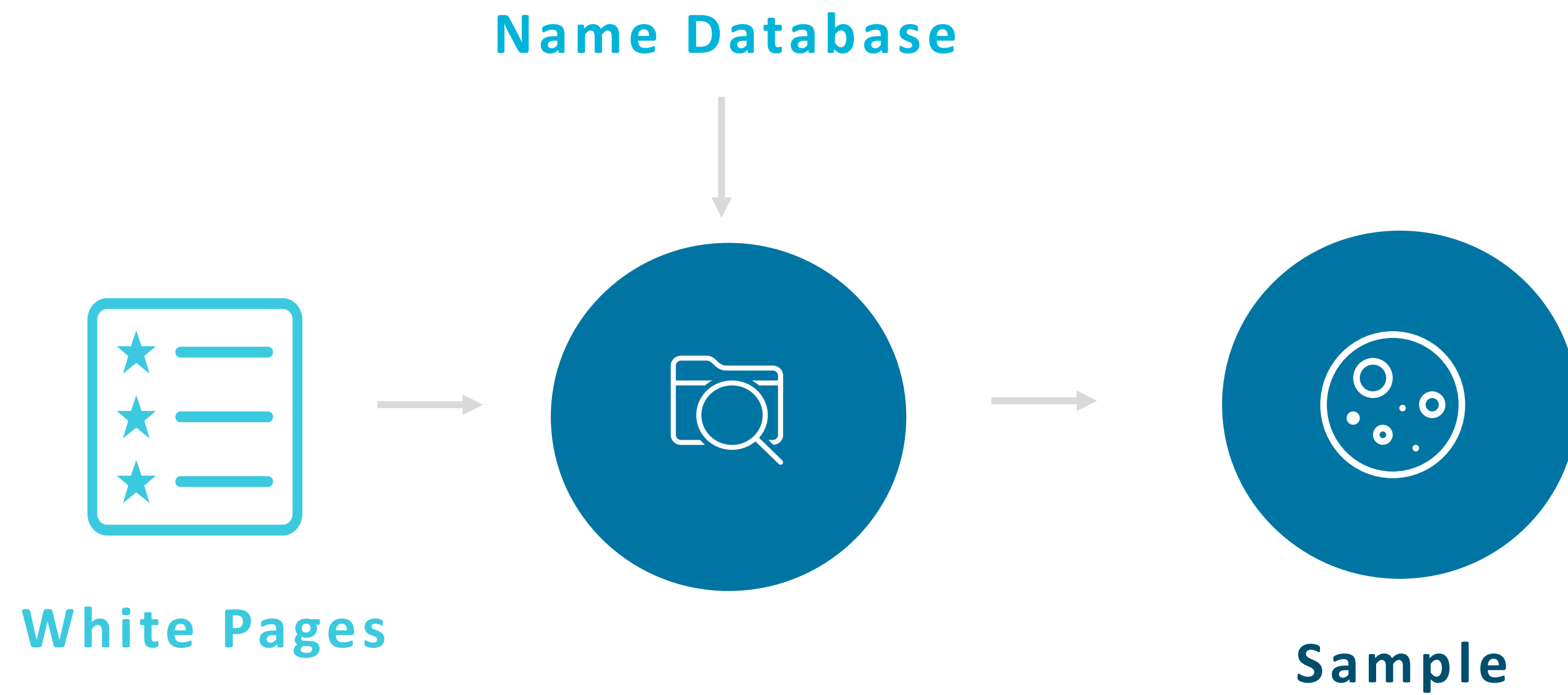
The Hague



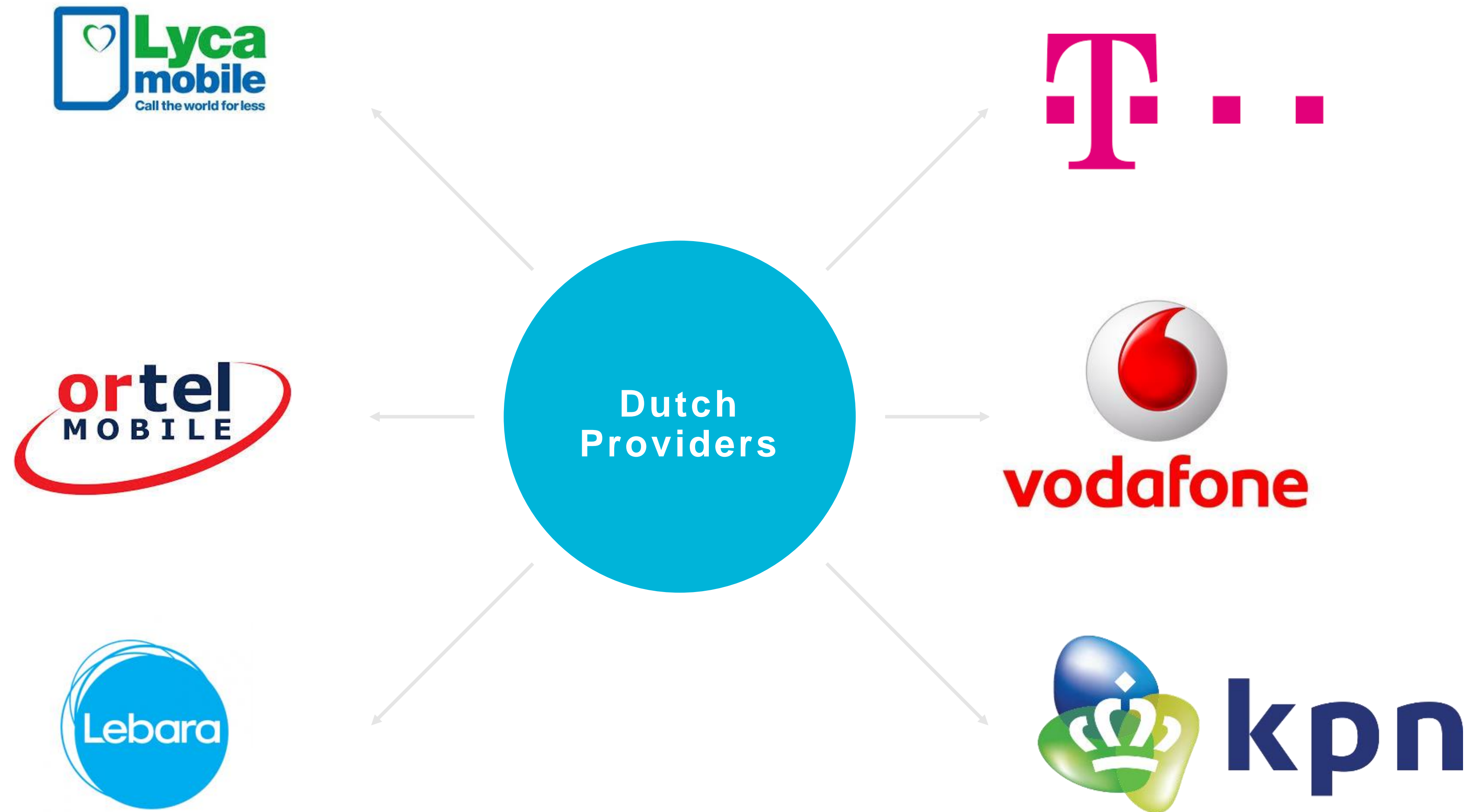
↑ **11.3%**

Amsterdam

List-Assisted RDD Sample



Specific Provider



Sampling Approach Evaluation



White Pages



Targeted RDD Sample



Social Media



Landline Based



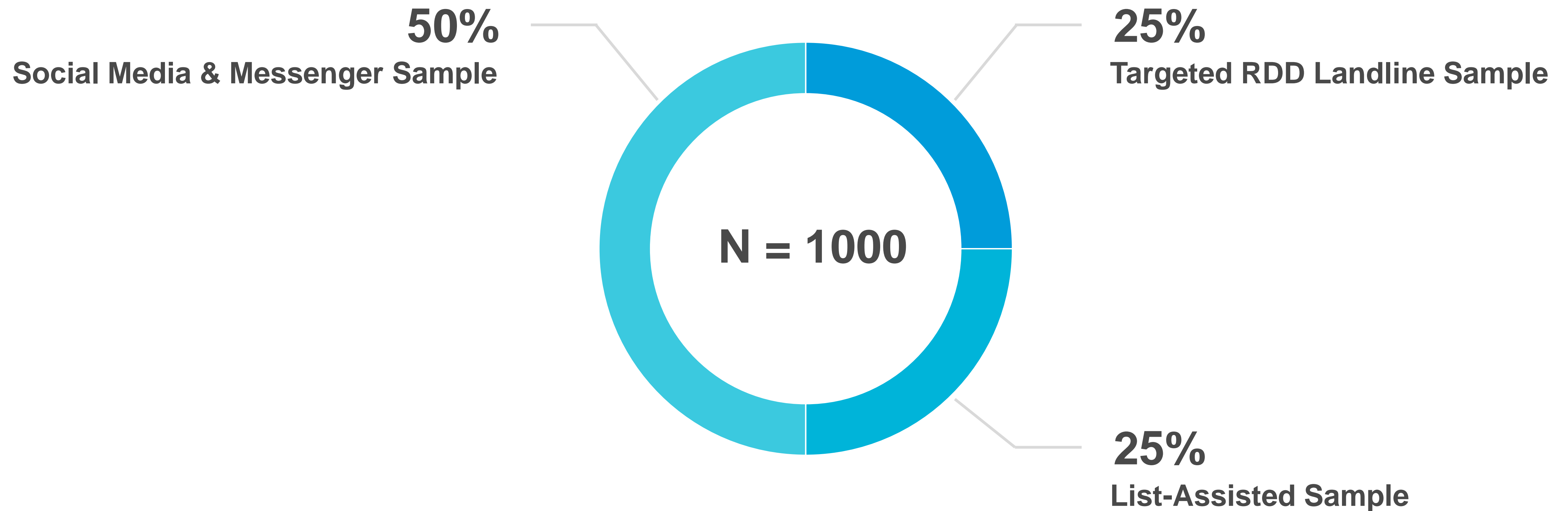
Mobile Based

3. Methodology

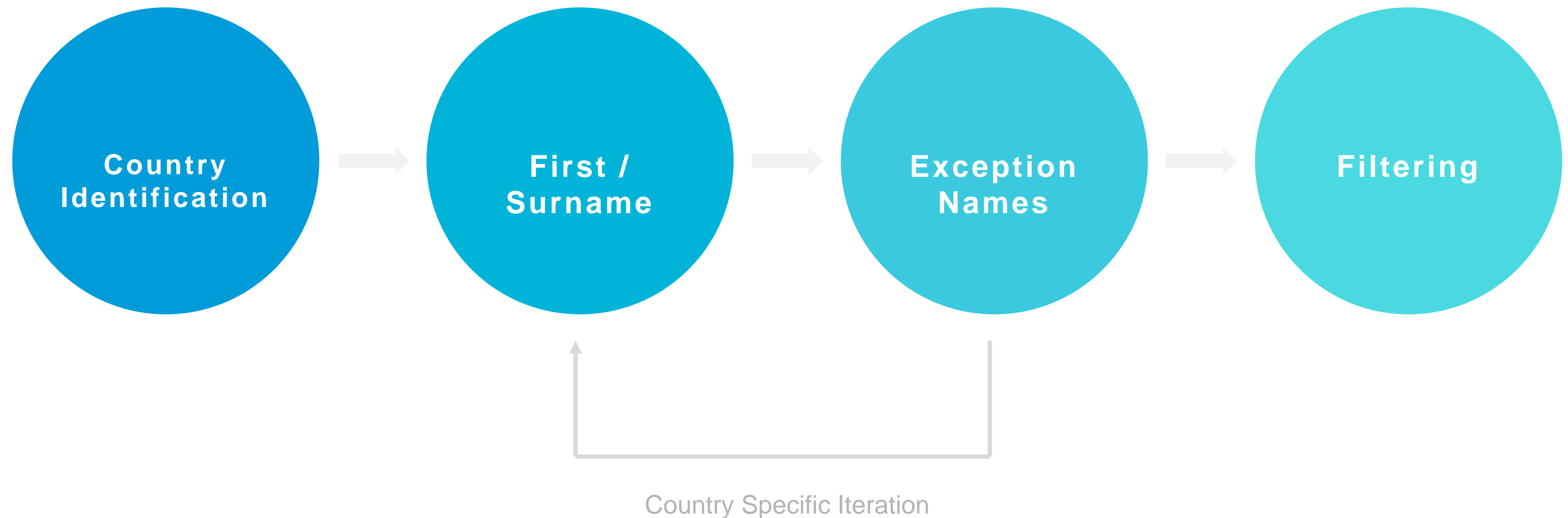


**“How do we compose
our sample?”**

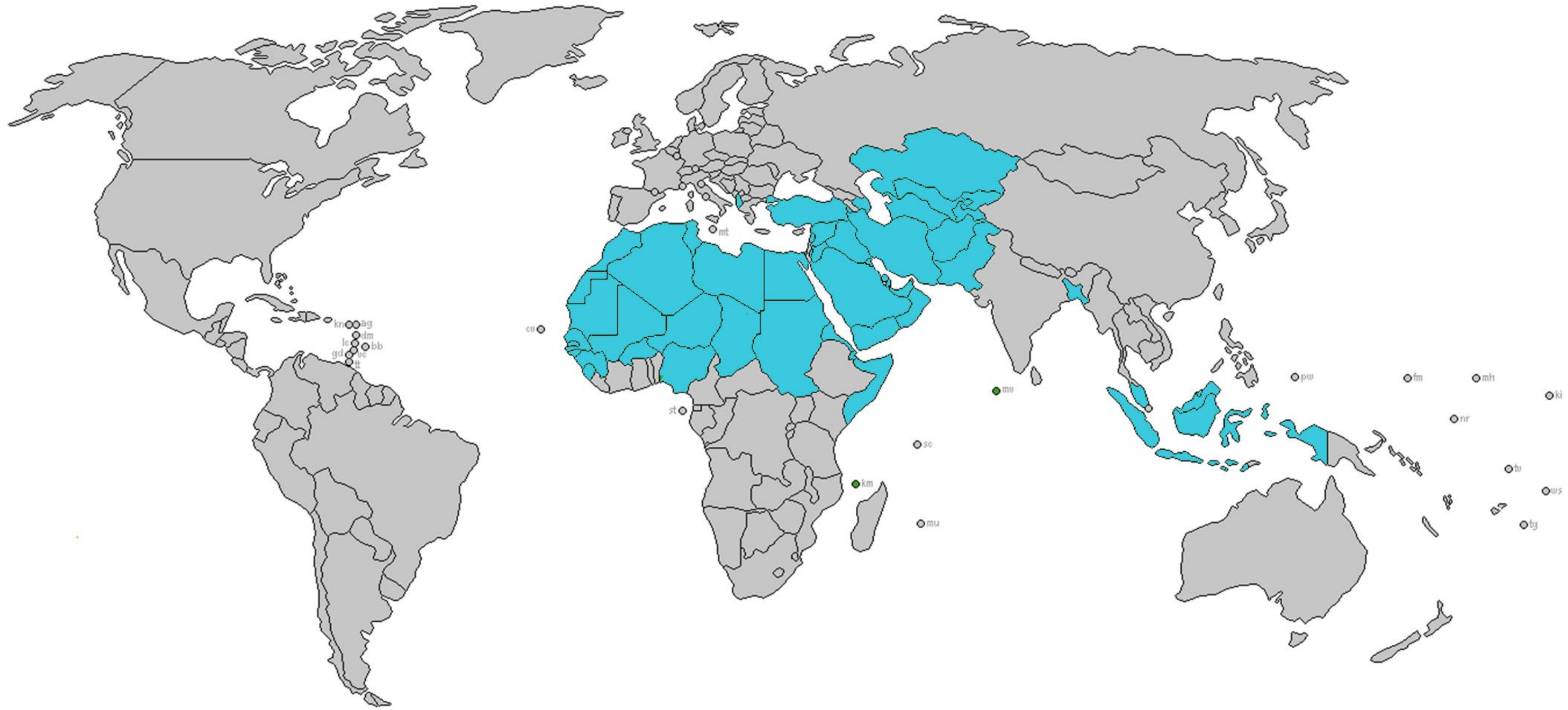
Sample Composition



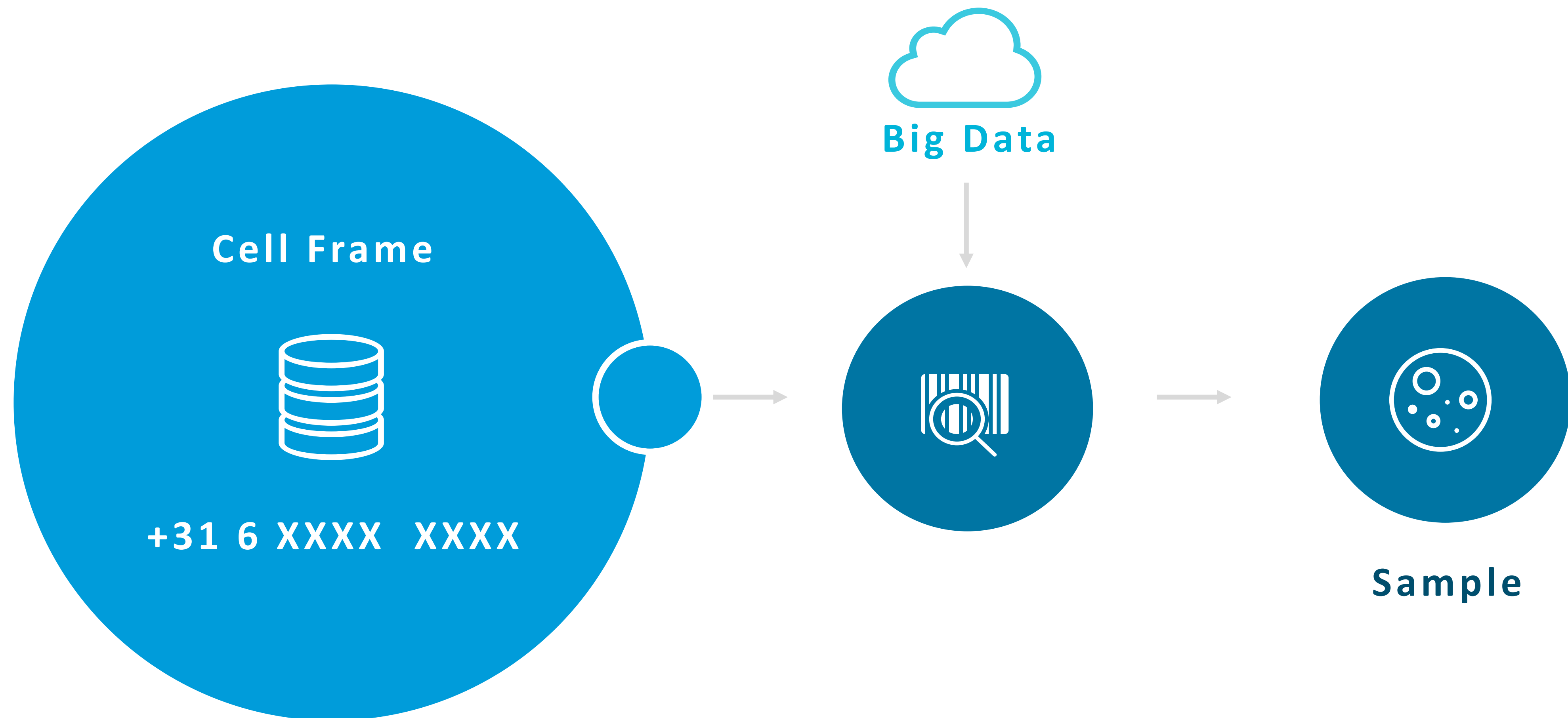
Onomastic Approach: Name Database



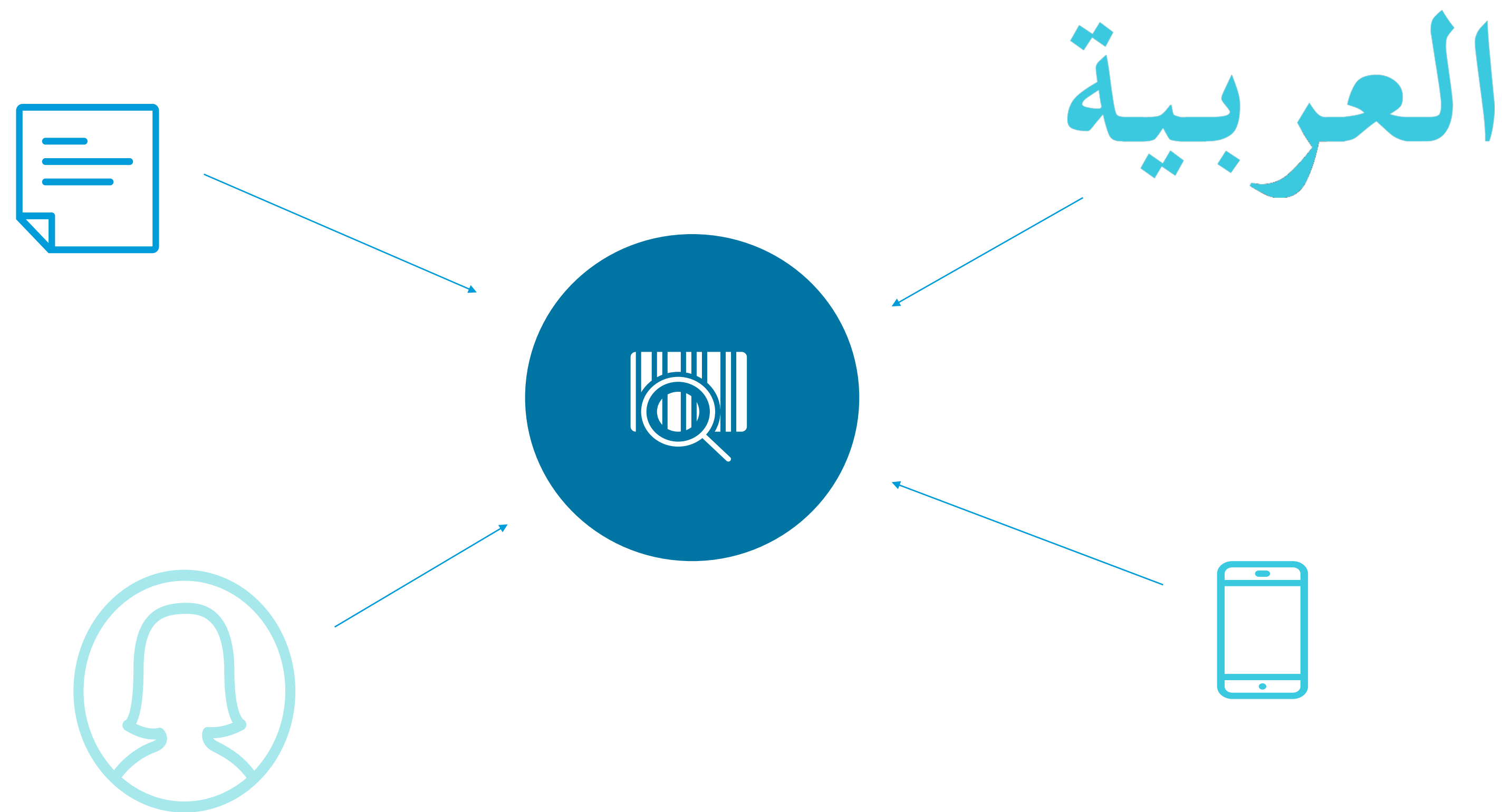
Majority Muslim Countries



Generation of Social Media Sampling

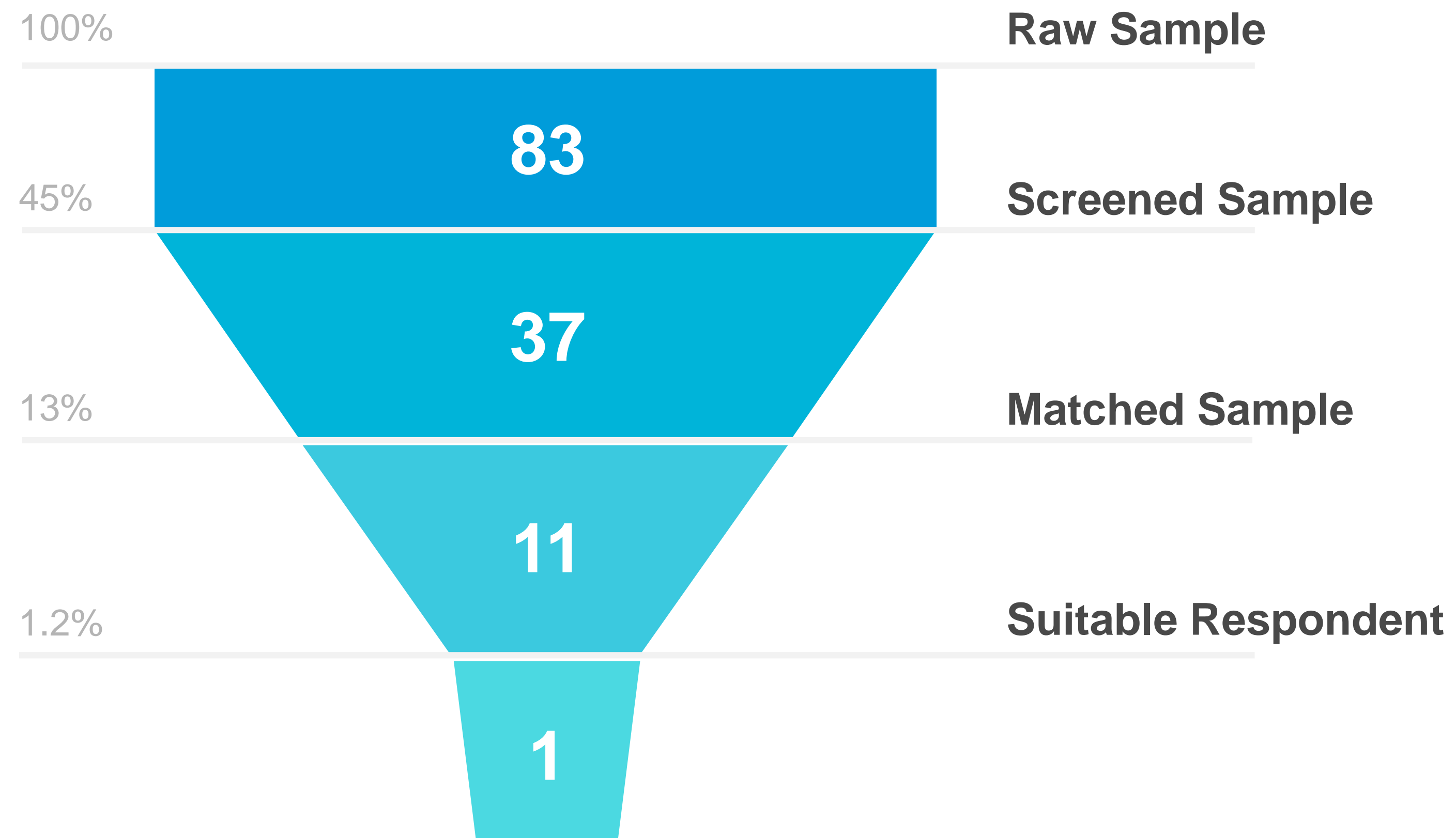


Which Public Sources are used?



4.Fieldwork & Sampling Results

SAMPLE GENERATION



Comparison of Generation

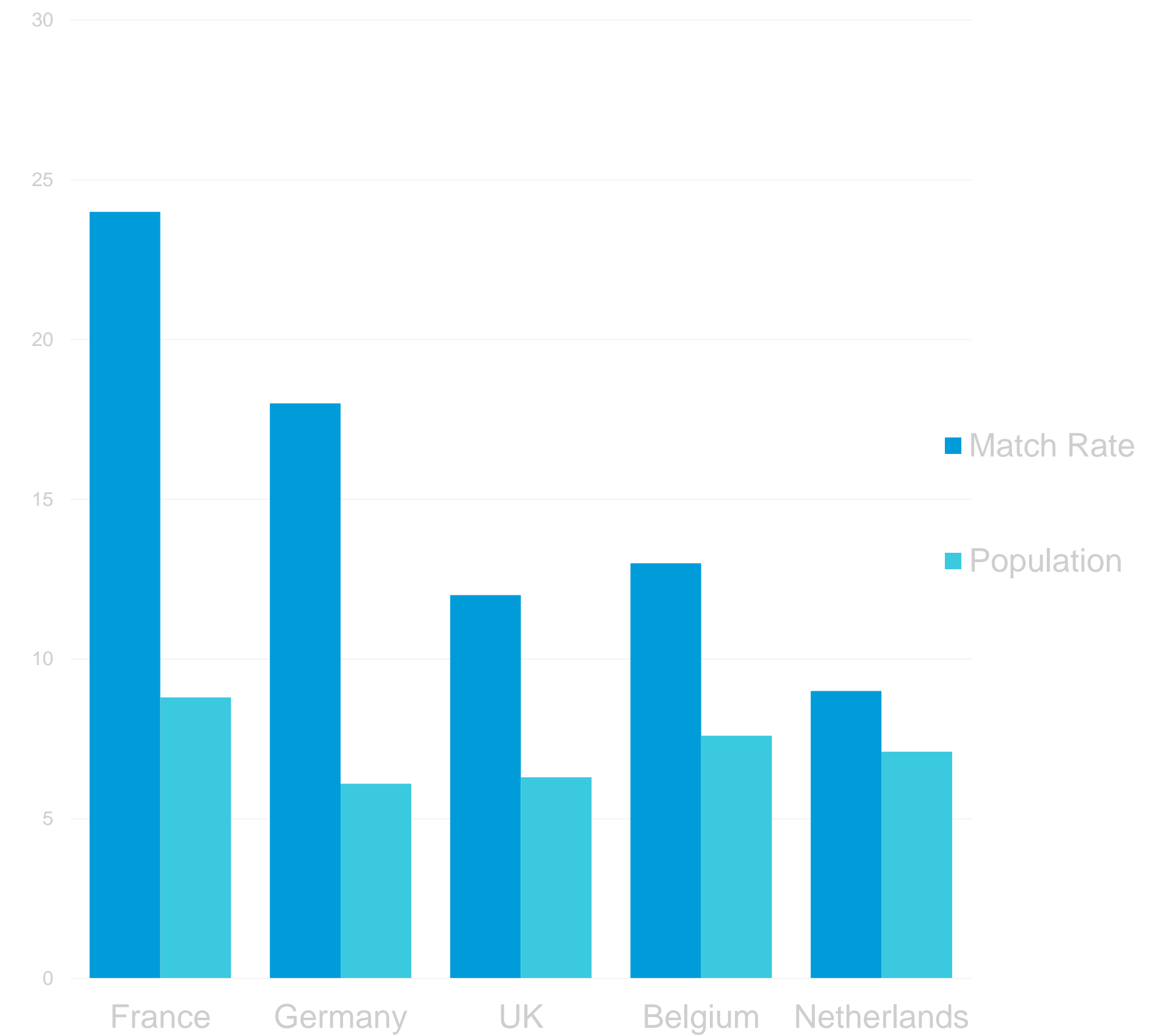
Example Netherlands:

11.1 Profiled records needed to get 1 suitable record

Respondent Identification

53% to 68%: Not much differences between countries

→ Self-Reporting



Fieldwork Results

CATEGORY	TYPE	CENUS DATA	SOCIAL MEDIA	LISTED
GENDER	Male	49.1%	63.8%	41.8%
	Female	50.9%	36.2%	58.2%
AGE	18-29	10.4%	44.7%	5.1%
	30-49	30,5%	43.1%	16.8%
	50+	59.1%	12.2%	78.1%

Fieldwork Results

CATEGORY	TYPE	CENUS DATA	SOCIAL MEDIA	LISTED
Education	Lower Sec or Less	31.3%	24.6%	12.7%
	Upper Sec + Non-Tert.	40.6%	43.3%	39.1%
	Tertiary	28.1%	32.1%	48.2%

Fieldwork Results

Zuid-Holland

Factor 1.7

Noord-Holland

Factor 1.2



↑ **23.6%**

Language Problems Social Media



3.8%

Language Problems Listed

5.Conclusion & Moving Ahead



**“Why cant we just
use an online panel?”**

ADVANTAGES AND DISADVANTAGES

Advantages

- Reach of first generation
- Larger coverage
- Younger Age groups well represented
- Geographical coverage

Disadvantages

- Age Max: ~70
- Bias towards online
- Dependency on name database
- Only access to public profiles

Future: Full Probability Frame

Full Probability Frame

Heading towards a full single frame.

Removing non-qualifier

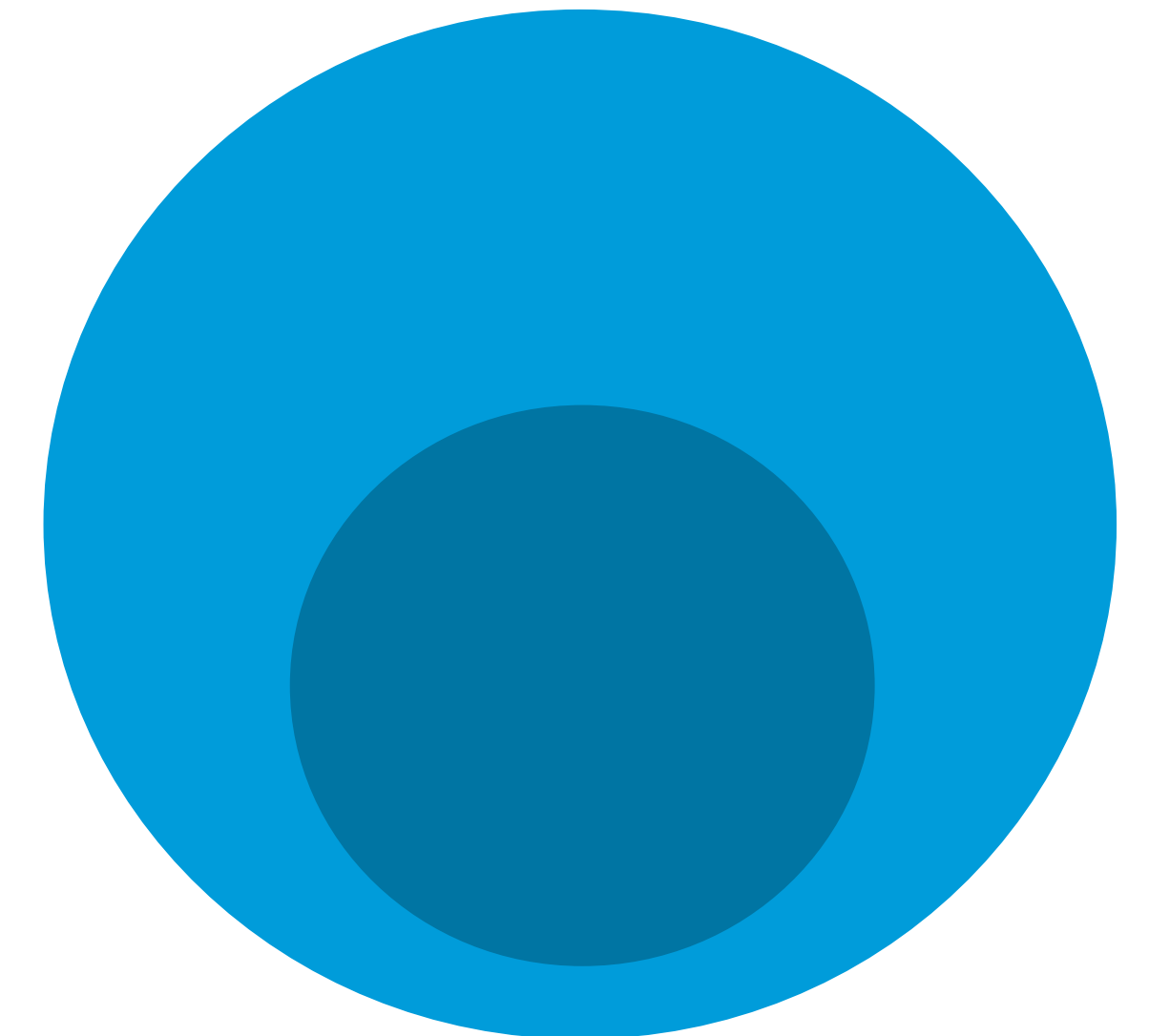
Reverse Logic: Screen-out those that do not qualify.

Onomastic Approach

More detail to local population data. Not just full name but also consideration of first / surname.

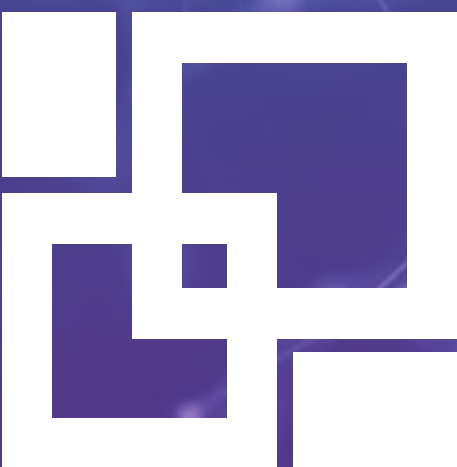
Increased Matching Rate

Addition of more data sources to increasing matching rates



Thank you!

Premier. Sample. Provider.
Sample Solutions



GDPR Compliancy

