

# EFAMRO OPEN MEETING

## *Fraudulent Respondents*

4<sup>th</sup> May 2023

Bernd Wachter, President

Debrah Harding, Vice President

Elena Onbright, Vice President

<https://efamro.eu/>

follow us on:



<https://twitter.com/efamro>



<https://www.linkedin.com/company/efamro>



# Opening of meeting

*The aim of the EFAMRO Open Day meeting on “Fraudulent Respondents” is to:*

- Discuss the new global pan-association initiative to address data quality and integrity*
- Discuss the impact of fraudulent participants and technology on data quality*
- Hear from some panel companies on what they are doing to address data quality and integrity issues*
- Learn about some of the techniques which can be used to improve data quality and integrity*



# Agenda for Today

- 1 *Welcome – Bernd Wachter*
- 2 *The global data quality initiative including the MRS project – Debrah Harding*
- 3 *The panel supplier perspective*
  - Erling Eriksen, Norstat
  - Oliver Frangakos, Dynata
- 4 *Some technical solutions – Elena Onbright, MAP*
- 5 *Roundtable discussion*

# *The global data quality initiative*





# The Global Data Quality Initiative

## The specific association led projects include:

- **MRS:** Fraud and bots technology, mobile considerations, panel/supplier data analytics
- **Insights Association:** Language, measurement, standards, and activation; toolkits
- **The Research Society:** Incentives use, language, and guidelines
- **Samplecon:** sampling technology codes, benchmarks and standards
- **CRIC:** Sample Frame consistency, transparency and quality; Guide for buyers of sample
- **ESOMAR:** participant-centricity and the role of the survey design in quality; checklists and the ESOMAR 37 Questions

Other associations are also inputting to the initiative, for example the Austrian association VMÖ, are participating in the UK project



# The MRS Data Quality Initiative - bots and fraud

A multi-disciplinary working group has been created to address data fraud and bot technology:

- Johnny Caldwell - Pureprofile
- Cecile Carre - Ipsos
- Oscar Carlsson - Cint
- Rebecca Cole - Cobalt Sky
- Daniel Culshaw - B2B International
- Mardien Drew – Today Consultancy
- Simon Glanville - Ronin
- Florian Kögl- Redem and the Austrian association, VMÖ
- Leah McTiernan - Ipsos
- Jack Millership – Zappi



# The MRS Data Quality Initiative - bots and fraud

**MRS is focusing on several workstreams:**

- **Terms and definitions to describe bot and fraud technologies**
- **Identifying research sector approaches to combat bot and fraud technologies across modes and methodologies**
- **External sectors' approaches to addressing fraud and bot technology**
- **Identifying legal and GDPR issues for some of the techniques use for addressing bot and fraud technologies**
- **Compiling a list of sources of fraudulent participant responses and working with platforms to have sources removed**
- **Investigating new solutions to combat bot and fraud technologies**
- **Investigating how technology is disrupting qualitative research**
- **Documenting and categorising third party quality and security software solutions**



# The MRS Data Quality Initiative - mobile optimisation

Since 2018 MRS has been facilitating an annual research project to explore the impact that poor mobile design and lack of mobile optimisation has on participation, completion, and response rates in research

Collaboration of four research companies:

- **Dynata**
- **Kantar**
- **Cint/Lucid**
- **Toluna**

Aggregated Response rate data from 2016 to 2021

New for 2022 was an updated approach to the participant satisfaction survey including tests of poor mobile design on participant behaviour and the inclusion of data from 13 markets

The results from 2022 are available here:

<https://www.mrs.org.uk/pdf/Mobile%20optimisation%20research%202022%20FOR%20ISSUE%20ON%20THE%20MRS%20WEBSITE.pdf>

**This programme is now part of the global data quality initiative**





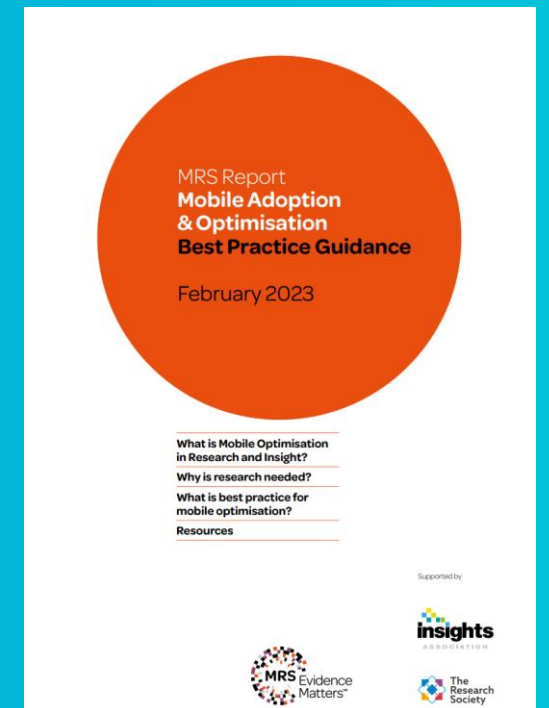
# The MRS Data Quality Initiative - mobile optimisation

MRS has produced some best practice recommendations to help practitioners to produce better mobile design, increase optimisation and to improve completion and response rates:

[https://www.mrs.org.uk/pdf/MRS\\_Report\\_mobileOP\\_0323v5.pdf](https://www.mrs.org.uk/pdf/MRS_Report_mobileOP_0323v5.pdf)

Listen to the webinar discussing the results here:

<https://www.mrs.org.uk/resources/mobile-optimisation-research>

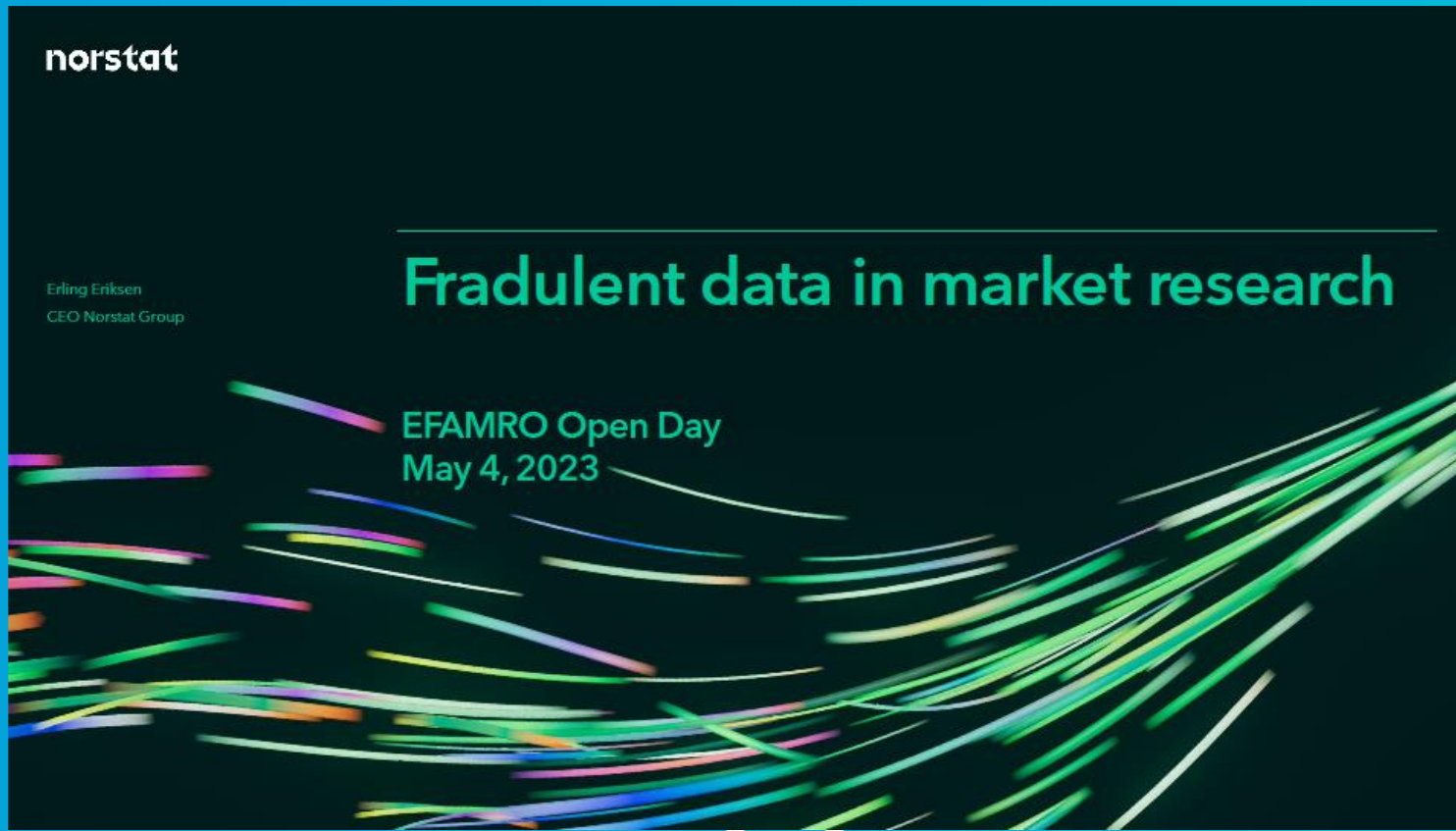


# *The panel provider's perspective*



# *The NORSTAT perspective*

Erling Erikson (CEO Norstat AS)



## Norstat Group

A European leader in data collection for market research

- > 2,300 customers
- > 13,000,000 interviews
- > 2,000,000 panelists
- > 19,000 projects completed/y
- 19 countries with online panels

Established

**1997**

**> 600**

Full-time employees (>280)  
and FTE of part-timers (>320)

**€ > 81m**

Group Revenues 2022

**22**

Offices in 15 countries  
across Europe



## Fraudulent data is not a major problem we experience

### Why?

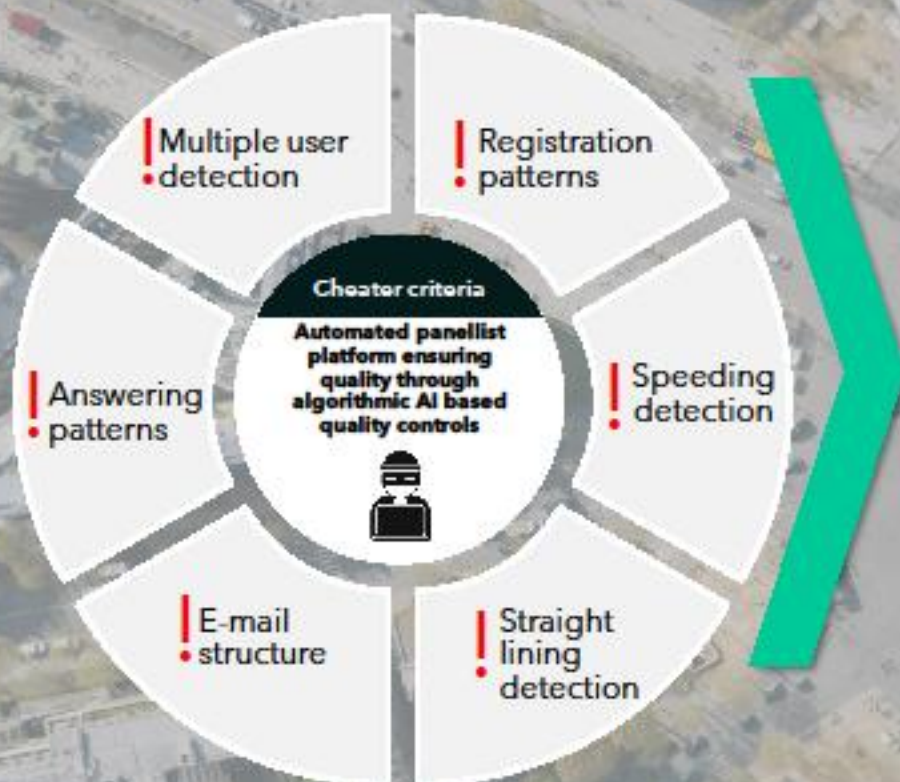
- **No open access** – panel access by invitation only
- **Internal control** over panel recruitment
- **Two-factor authentication** with SMS
- **No cash incentivization** in majority of our panels
- **Self-supplied** for 90%+ of our projects
- The majority of recruitment done through **channels we control from A to Z**
  
- Norstat panels **not present in the exchanges** – and we are buying minimal from the exchanges
  
- **Rigorous cheater detection and cheater scores**
  - Logic tests
  - Speed
  - Duplicates

We minimize exposure, limiting bot/cheater entry to a very minimum


We work actively with automated and manual rigorous quality controls



# NORSTAST AUTOMATED CONTROLS



**Cheater score (0p-40p)** generated after conducted controls. Panellists receiving a high cheater score are investigated and unsubscribed from the panel

**Bot detection** through third party software 

**System warning for self cheaters**

**Logic tests** requested ahead of each survey

**Automated controls** are part of Norstat Engagement Centre

## Example panellist

Cheater score calculator: **60%**

Very suspicious: **?**


24 of 40 points: **?**

Country:  SE (Sweden)

Language: SV (Swedish)

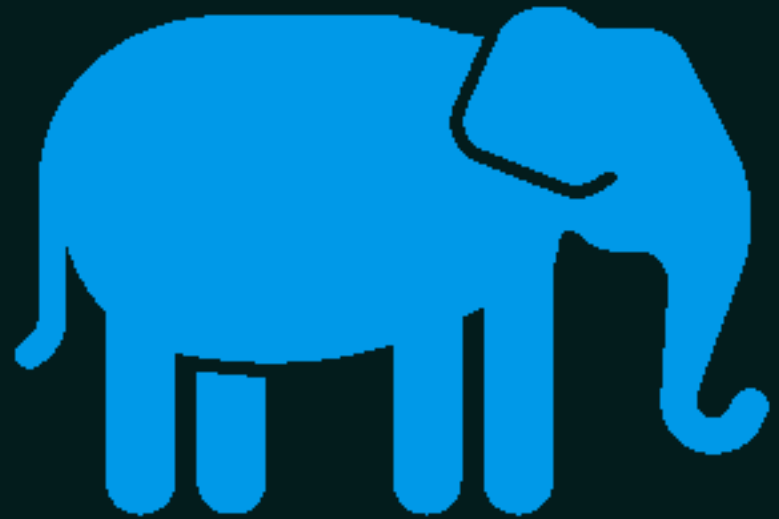
Completes: **26**

Panel member for: 5.3 years

Status: Cheater  **?**

Ingoing panelist controls with outgoing cheater detection eliminating panelists from platform through automated controls as part of the Norstat's automated core platform

The biggest elephant is not  
fraudulent data ...



Skewed data ...

The next bubble ready to burst?





# The *DYNATE* perspective

Oliver Frangakos (Country Manager DACH)



## EFAMRO Open Day - Fraudulent Responses

Oliver Frangakos, Country Manager DACH

The concepts presented in this presentation are property of Dynata, LLC.  
Duplication or dissemination of the information without the express written consent is prohibited.





WE ARE  
INNOVATING IN  
A TRANSFORMED  
WORLD

- Research is transforming to a new normal
- Data scarcity is our new reality
- Long-term sustainability is a key priority
- We must elevate the respondent to be an equal stakeholder in our business



**RESPONDENT QUALITY  
IS A KEY COMPONENT OF  
LONG-TERM SUSTAINABILITY**





**REMOVING THE “RIGHT”  
PEOPLE IS CRITICAL**

# THE NAKED EYE VS. AUTOMATION

RESOURCE INTENSIVE

HIGH RANGE OF ERROR

CAN INTRODUCE UNINTENDED  
BIAS INTO SAMPLE



# BUILDING A BETTER (AUTOMATED) MOUSETRAP

---

 **IMPERIUM**<sup>®</sup>  
A DYNATA DATA QUALITY SOLUTION



# IMPERIUM<sup>®</sup>

A DYNATA DATA QUALITY SOLUTION

**6**

MONTHS  
OF DATA

**17,000**

STUDIES

**20**

MILLION  
COMPLETES SCORED

**1.5**

MILLION FLAGGED FOR  
REMOVAL BY  
QUALITYSCORE™

# WOULD YOU KEEP OR REMOVE THESE RESPONDENTS?

## Respondent A:

Removed with manual approach

Straight-lined 2 out of 7 grids

## What does QualityScore™ tell us?

**Straight-lining:** only slightly below average

No speeding or acceleration

Open-ends (7):

No pasted text; answers were ideal. E.g.,

*The company itself is reliable, but concerns about the demographic structure (population decline)*

*Satisfaction with the expected pension amount*

---

## QualityScore™: 84

Very good! Should have been kept in data



# WOULD YOU KEEP OR REMOVE THESE RESPONDENTS?

## Respondent B:

Kept with manual approach

Straight-lined 1 out of 6 grids

Did not speed through survey

Open-end answers (7) – some duplication,  
but look great overall. E.g.,

*Offers a comprehensive suite of lending  
solutions to address my need*

*Gives me access to exclusive investment products (e.g., pre-IPO  
access, alternative investments, private equity, venture funds)*

*Offers a hybrid model of servicing clients via remote banking and  
in-person advisory solutions (e.g., in-person meeting vs. chat vs.  
video call, on-demand market outlook vs. live webinars)*

*Is a partner that enables me to achieve my life goals*

## What does QualityScore™ tell us?

All 7 open-ends copy and pasted  
(likely from an insurance website)

Did speed through the survey but not  
quite enough to trip a flag

Failed the acceleration check

---

## QualityScore™: 16 (Fail)

Should have been removed from the data

# FOUR TYPES OF RESPONDENTS:

## THE GOOD, THE BAD, AND THE (NOT ALWAYS) UGLY

LAZY CHEATERS

SMART CHEATERS

UNENGAGED REAL PEOPLE

PERFECT PEOPLE



# CATCHING THE “RIGHT” PEOPLE

|                                     | MANUAL CHECKS | QUALITYSCORE™ |
|-------------------------------------|---------------|---------------|
| LAZY CHEATERS                       | ✓             | ✓             |
| COMPLETELY UNENGAGED<br>REAL PEOPLE | ✓             | ✓             |
| SMART CHEATERS                      | X             | ✓             |
| SLIGHTLY UNENGAGED<br>REAL PEOPLE   | ?             | ✓             |

# CATCHING THE “RIGHT” PEOPLE

8%

**POOR  
RESPONDENTS**

INCLUDING 3-4%  
SMART CHEATERS &  
REPLACING 3-4% SLIGHTLY  
UNENGAGED REAL PEOPLE

12%

**“REVIEW FURTHER”  
COMPLETES**

NOT “BAD”,  
BUT TRIPPED A FEW FLAGS

80%

**QUALITY  
COMPLETES**

ACCEPTABLE  
RESPONDENTS

**17,000 STUDIES**



- Consistent Data Cleaning Approach across all your projects
- Significant time savings in the cleaning process



**MAKING THE ENTIRE  
ECOSYSTEM MORE  
SUSTAINABLE...**

**...FEEDBACK LOOPS TO  
REINFORCE ALL TOUCHPOINTS**



AN  
ACCEPTANCE OF  
IMPERFECTION

AKA “FIT FOR PURPOSE”





# QUESTIONS



# *Some technical solutions* Elena

Onbright (CEO MAP marketing research)

- Survey Programming
- Data Management
- Additional Services





# Technical Solutions

- Sample selection
- Pre-study
- In-study
- Post-Study

| Feature                           | Single-browser | Cross-browser |           |
|-----------------------------------|----------------|---------------|-----------|
|                                   | Entropy        | Entropy       | Stability |
| User agent                        | 6.71           | 0.00          | 1.39%     |
| Accept                            | 1.29           | 0.01          | 1.25%     |
| Content encoding                  | 0.33           | 0.03          | 87.83%    |
| Content language                  | 4.28           | 1.39          | 10.96%    |
| List of plugins                   | 5.77           | 0.25          | 1.65%     |
| Cookies enabled                   | 0.00           | 0.00          | 100.00%   |
| Use of local/session storage      | 0.03           | 0.00          | 99.57%    |
| Timezone                          | 3.72           | 3.51          | 100.00%   |
| Screen resolution and color depth | 7.41           | 3.24          | 9.13%     |
| List of fonts (Flash)             | 2.40           | 0.05          | 68.00%    |
| List of HTTP headers              | 3.17           | 0.64          | 9.13%     |
| Platform                          | 2.22           | 1.25          | 97.91%    |
| Do Not Track                      | 0.47           | 0.18          | 82.00%    |
| Canvas                            | 5.71           | 2.73          | 8.17%     |
| WebGL Vendor                      | 2.22           | 0.70          | 16.09%    |
| WebGL Renderer                    | 5.70           | 3.92          | 15.39%    |
| Use of an Ad blocker              | 0.67           | 0.28          | 70.78%    |
| AmlUnique                         | 10.82          | 0.00          | 1.39%     |
| Screen Ratio                      | 1.40           | 0.98          | 97.57%    |
| List of fonts (JavaScript)        | 10.40          | 6.58          | 96.52%    |
| AudioContext                      | 1.87           | 1.02          | 97.48%    |
| CPU Virtual cores                 | 1.92           | 0.59          | 100.00%   |
| Normalized WebGL Renderer         | 4.98           | 4.01          | 37.39%    |
| Task (a) Texture                  | 3.51           | 2.26          | 81.47%    |
| Task (b) Varyings                 | 2.59           | 1.76          | 88.25%    |
| Task (b') Varyings+anti-aliasing  | 3.24           | 1.66          | 73.95%    |
| Task (c) Camera                   | 2.29           | 1.58          | 88.07%    |
| Task (d) Lines&Curves             | 1.09           | 0.42          | 90.77%    |
| Task (d') (d)+anti-aliasing       | 3.59           | 2.20          | 74.88%    |
| Task (e) Multi-models             | 3.54           | 2.14          | 81.15%    |
| Task (f) Light                    | 3.52           | 2.27          | 81.23%    |
| Task (g) Light&Model              | 3.55           | 2.14          | 80.94%    |
| Task (h) Specular light           | 4.44           | 3.24          | 80.64%    |
| Task (h') (h)+anti-aliasing       | 5.24           | 3.71          | 70.35%    |
| Task (h'') (h')+rotation          | 4.01           | 2.68          | 75.09%    |

## ❖ Digital Fingerprinting and Fraud Identification

- Examine potential respondents before you engage them – are they a bot, known fraudulent person, or a duplicate
- A digital fingerprint contains a set of data that identifies browser setup and device as unique. Once browser and/or device is analyzed, **the fingerprinting software saves the fingerprint data server-side, outside of the user's control.** This allows internet users to be identified and tracked, even when they take evasive measures against cookies. These identifiers mainly relate to browser and device but can be used to pry into personal data and internet browsing habits. Digital fingerprinting has made advanced user and device tracking without cookies a reality and has become extremely hard to control or regulate.
- A digital fingerprint is created with various information extracted from one's browser and device. These data points are calculated to provide a digital fingerprint – a digital identifier that is unique: IP address; Device MAC address; User-agent string; Clock information – used to cross-verify location alongside IP address; Web browser plugins ; TCP stack variation; Installed fonts on device; JavaScript objects; Internal application programming interfaces (API); Device information such as screen resolution, touch support, OS and language; Flash data; List of mime-types; Timestamp; CSS information; Hypertext Transfer Protocol (HTTP) headers.



# Technical Solutions

## In-Study

### Survey Script Controls on Question Level

- Add scripts at question level that checks how much time it has taken the respondent to complete each question → screen out.
- Add Text Analytics Tool that measures and scores a respondent's engagement in real-time by analyzing open-end responses and checking for proper grammar, response length, profanity, copy/paste, and other attributes.
- Add script based on time stamps controls the amount of time that the respondent took to complete the entire survey → screen out before end page.
- Add script that sends the panel supplier the fraudulent panelist reason for screen out as a variable in the pass-back link.



# Technical Solutions

## Post-Study

### Data scrubbing

Data scrubbing tools provide recommendations based on poor or fraudulent data to be examined and removed.

- Answer patterns
- Click-through behaviour
- Keystroke analysis
- Inattentiveness
- Duplicate response identification
- Machine responses
- Video viewing validation
- Copy/paste behaviour
- Speeding.

# *Roundtable discussion*



*Other topics for today?*

*Future Open meeting topics?*

