

# **The future of online data collection in social surveys: Shared learning on the challenges, opportunities and best practice.**

International conference, June 2019.

## **Report: A summary of the agreed challenges, opportunities and best practice for online data collection in the UK in 2019 and the future.**



### **Authors:**

Laura Wilson – Office for National Statistics

Olga Maslovskaya – University of Southampton

Date – December 2019.

## **Acknowledgements:**

The conference was supported with funding from the Office for National Statistics (ONS), the Economic and Social Research Council (ESRC) Secondary Data Analysis Initiative project "Understanding survey response behaviour in a digital age: Mixed-device online surveys and mobile device use" and National Centre for Research Methods (NCRM) Work Package 1 "Data Collection for Data Quality". We are grateful to the University of Southampton (UoS) for hosting the two-day event.

A special thanks go to staff from ONS and UoS who helped to organise the event, they are:

- Penny White (UoS)
- Clare Parsons (ONS)
- Hannah Evans (ONS)

We would also like to thank the key note speakers, presenters and participants for their attendance and contributions over the two days. Your collective efforts made the event a great success.

Thank you,

Organising Committee.

*Olga Maslovskaya – University of Southampton*

*Laura Wilson – Office for National Statistics*

*Andrew Phelps – Office for National Statistics*

*Gabriele Durrant – University of Southampton*

*Peter WF Smith – University of Southampton*

## 1. Introduction:

Online collection of social survey data is becoming more prevalent globally. Societal demands and advances in technology have fueled this change. The online social survey landscape in the UK is varied: some surveys already feature an online mode as part of their offering whereas others are still in the stages of feasibility testing and exploration. However, regardless of stage, we all share the same concerns and are seeking answers to the same questions.

In 2013 the GenPopWeb network ran a series of events to bring together international experts to establish the lay of the land at that time regarding online data collection in social surveys. The network also identified new research questions and potential areas for collaboration across sectors. This network was highly effective in focusing the minds of researchers in this field and empowered them to direct research within their organisations.

In June 2019, the Office for National Statistics (ONS) and the University of Southampton (UoS) held a joint international conference entitled “The future of online data collection in social surveys: Shared learning on the challenges, opportunities and best practice”. Survey experts from around the world were invited to attend and contribute to the conference, in total 78 attended. The purpose of this event was to share international learning and to consolidate the views on and understanding of what it means to operationalise an online social survey in 2019. Further to this, the event concluded with establishing the challenges, opportunities and best practice for the future of online data collection in the UK through an interactive workshop in which all participants took place. The event was effectively an update on the GenPopWeb events. It aimed to understand all the above and to determine whether we had found answers to those original questions from 2013 or whether they were still outstanding.

The following sections of this paper focus on reporting back on the three main themes of the 2019 conference – challenges, opportunities and best practice. These themes were discussed in different formats during the two-day conference. Further insights on each theme were gathered in the last session at the end of the conference; attendees were assigned to sub-groups and asked to generate up to three points under each conference theme. Each group then presented back to everyone and collectively we discussed and finalised the main points. Post conference, the organisers thematically analysed all the comments under each conference theme and the conclusions are presented in the sections below.

## 2. GenPopWeb conclusions:

The 2014 GenPopWeb report (Nicolaas et al. 2014) summarised the knowledge gaps identified by the network at that point in time into three themes (see report for details):

1. Sampling and coverage
  - a. Limited options for sampling frames and methods
  - b. Internet access
2. Engagement and participation
  - a. Respondent's ability and willingness to take part in an online survey
  - b. Lower response rates in online surveys compared to other modes
  - c. Achieved sample composition
  - d. Optimal response strategy
3. Measurement issues
  - a. Measurement quality and how it differs by device
  - b. Adapting questionnaire and questions/ optimising them for online formats
  - c. Smartphone use

The report identified and recommended future research in these areas in order to focus research efforts in a bid to increase the success of online data collection in a mixed-mode environment.

The 2019 ONS and UoS conference concluded that knowledge gaps still exist within the above themes. However, international research conducted between 2013 and 2019 across all sectors on the GenPopWeb themes has reduced the gap for some areas. In some instances, a change in approach to design has refocused research efforts onto another related aspects.

The 2019 ONS and University of Southampton conference website (<https://www.ncrm.ac.uk/research/datacollection/>) contains links to all keynote speeches and all presentations from the conference, these add to the knowledge base on the above themes.

The following sections of this report detail the current knowledge and research gaps.

### **3. 2019 ONS and UoS conference conclusions:**

Below is a summary of the thematic analysis on the challenges, opportunities and best practice for online data collection activities as identified by the conference attendees.

#### **3.1. Challenges:**

##### **3.1.1. Managing stakeholders/ customers understanding and expectations regarding their role in the questionnaire design process (and clarifying what is suitable for the public to interact with).**

A common theme in discussions was that boundaries, or roles and responsibilities, when managing the design and delivery of a survey are often blurred and sometimes not respected. Unfortunately, this can often result in advice from survey experts being ignored and consequently a suboptimal approach or design being pursued. This can make the design and delivery of an online survey challenging. Designing for online brings with it some unique problems which sometimes require tailored or new solutions (compared to other modes) and the expert advice aims to meet these needs.

It is important that we manage stakeholder/ customer expectations of the respondents in self-completion modes. One of the biggest challenges, is to try to ensure that what is asked of the respondent is reasonable. When the expectations from the respondents is too great then successful delivery of a survey is at risk. Sharing insights from research participants with stakeholder/ customers is one of the effective ways to counter this challenge.

##### **3.1.2. Pressured development and delivery environments.**

Attendees felt that the timelines and budgets for developing surveys are shrinking, and the delivery environments are becoming more pressured and intense – “expecting more for less”, (a quote from an attendee). This broke down into the following subthemes.

- a. Limited resources – impossible to achieve everything or to pursue research in the desired way to achieve quality due to the limited resources available.
- b. Decisions not being made on evidence – due to the pressured development timelines and lack of resources it often results in decisions being made at desk without evidence which is not best practice.
- c. Not enough time to test and learn – demands to deliver mean that there is often not enough time to pre-test ideas or options. This pressure is also felt post collection where there is not enough time to analyse and learn from the last exercise or survey round. This

may sometimes be viewed as a 'nice to have' step but is in fact critical to effectiveness and success.

- d. Lack of control – feeling of not being able to influence or shape the development timelines or environment
- e. Limited collaboration between sectors (i.e. Public, Private, Academic, not for profit) due to commercial relationships and pressures – more collaboration in this environment would alleviate some of the pressures but can be difficult due to commercial sensitivities and delivery pressures
- f. 'Save money versus save quality' quandary – in these environments money usually wins, and quality suffers. The effect of this is much more pronounced in online self-administered surveys as there is no interviewer to monitor the process of survey completion. This in turn contributes to large mode effects in mixed-mode designs.

### **3.1.3. Digital Divide: Challenges of current and future devices.**

In 2013 concerns around the use of mobile phones in online surveys were at the forefront of survey methodologists' minds. In 2019, although some organisations have chosen to embrace mobile devices (tablets and smartphones) when designing web surveys, others are still considering their next steps in this space. Those surveys that have embraced these devices are debunking myths about their limitations. They demonstrated that a design approach that includes these devices from the start can yield positive results. This is an example of how a challenge can disappear through an alternative approach to design and development.

An improvement on 2013 is that it is now less common to see invitation letters that explicitly ask respondents not to use their smartphones. However, unless the surveys are designed using either "mobile first" design (i.e. designing questions for mobile use in the first instance and then applying them to other modes) or progressive enhancement, we still generally observe higher drop-off rates on mobile devices.

Smartphones are not the only alternative way to access online surveys, for example consideration may need to be given to games consoles in the future. However, if all surveys are built in a progressive enhancement way from now on, then we may avoid the issues which could be introduced by new devices used by respondents for survey completion in the future.

### **3.1.4. Pace of societal and technological change**

There was a feeling amongst attendees of being vulnerable and that the industry is always playing catch-up due to the pace of societal and technological change. There was agreement that there is a greater need for horizon scanning and cross-industry/ sector collaboration (e.g. Technology and Digital) to get ahead of the changes rather than having to

react to them. The following comments represent some of views expressed:

- “Susceptible to ‘shocks’”
- “Unknown unknowns”, i.e. things that we are neither aware nor understand
- “Fear of becoming out of date or too slow to evolve”

### **3.1.5. Sampling**

The same issues and concerns in relation to sampling exist in 2019 as in 2013 (see the GenPopWeb report (Nicolaas et al. 2014) for full details). In summary:

- a. We can still only currently sample an address and not a named individual (unless subsampling from a previous survey is used).
- b. Initial contact can still only be made via postal communications (unless subsampling from a previous survey is used).
- c. Recent changes to the UK legislation around data linkage and sharing may in part address some of the challenges in the future. However, this issue will continue to exist for some time.

### **3.1.6. Response rates**

Some of the same challenges and issues still exist in 2019 as in 2013. However, some issues have evolved based on research evidence in the interim years.

- a. Motivating respondents – this is still one of the main challenges faced by the data collection industry. Conference attendees noted that we are living in a transactional society and this ‘need’ should be addressed in our relationship with respondents. Currently we take information but what do we give back to respondents in return? Other points raised were focused on loss of trust, rebellion and apathy among respondents. The current climate we operate in is not conducive to response, especially in a self-administered mode. Further work is needed in this area.
- b. Selling the study – further work is needed to learn how to best sell the study and encourage participation. Research in recent years has proven that a respondent-centred approach to the development of communications can help explain the purpose of the study in a way that is suited to the respondent rather than the commissioner
- c. Inclusive design and respecting the diversity of needs – building an end to end survey that is inclusive and respectful of disability and diversity needs is challenging. However, it is integral to the success of the survey and quality (especially if the survey is providing estimates on these parts of society). The challenge is how to design and deliver in this way within the current pressured development environments. Delivery timelines and resources may need to change to allow this to happen. This should be an

accepted new stage of survey delivery, especially when delivering government surveys due to new UK Government legislation (GDS 2018) which stipulates the legal requirement to meet accessibility standards.

- d. Surveying of the offline population – this challenge was raised in 2013 and was raised again in 2019. However, in 2019 there was acceptance that there are solutions for this population and that this demand may diminish over time.
- e. Incentives – the challenge of achieving response remains. Organisations are still focusing their resources and efforts to explore how best to motivate potential respondents to take part, including the use of incentives. No firm conclusions on what approach is the most effective as results are context specific.
- f. Data protection – there is an increased awareness of data protection amongst the public. We need to address that respondent need for reassurance. However, it is challenging due to the amount of advice that we are required to provide that is associated with the new legislation. There are also challenges in how to word the information in a way that is widely understood, accurate yet not off-putting.

#### **3.1.7. Measurement errors and time series**

The same issues and concerns still exist in 2019 as in 2013 (see the GenPopWeb report (Nicolaas et al. 2014) for full details). In summary:

- a. Clarity needed on how to handle mode effects in mixed-mode surveys – further work needed to support the industry and customers in understanding and explaining them
- b. Concerns around breaks in time series and how to understand and communicate these more effectively.



## **3.2. Opportunities:**

### **3.2.1. More data available pre-survey**

Advances in this area (e.g. Big Data, administrative data, area profiling, other data sources etc) mean that there is more information available to us than before about address prior to contact being made. This may mean that with access to these, we can start to explore how to do things differently and begin to address some of the challenges mentioned above, such as contact strategies and content. However, additional time and extra resources are needed to explore these aspects further.

### **3.2.2. Chance to have a fundamental rethink about questionnaire design and do things better.**

- a. There was the view that now is the time, based on new learning and other advances, to refocus on excellent design and go back to basics with simple designs. This includes moving towards a respondent-centred approach to designing surveys.
- b. Interdisciplinarity, new paradigm – in the past, psychologists joined with social researchers to design surveys. Now there is a need to join with digital user experience and design experts to deliver high quality online surveys
- c. Chance for surveys to remain relevant – with this fundamental rethink comes the opportunity for surveys to remain relevant as they will evolve to meet the changes in requirements and respondent experience expectations
- d. Build a case for more experimentation and hypothesis driven research – these have suffered over the years due to limited resources and time.

### **3.2.3. New ways to get data**

Consider, research and plan how to use and embed the following into survey design and delivery:

- a. New technologies, e.g. smartphones, Artificial Intelligence (AI), Global Positioning System (GPS), photo, voice/sound etc.
- b. Administrative data, data spine and data linkage
- c. Paradata
- d. Machine learning techniques

### **3.2.4. New ways of inviting respondents to surveys**

With alternative sources available to us also comes the opportunity to experiment with alternative ways of inviting respondents to take part in our surveys. This could be taken as far as to explore using a more tailored approach to contact strategies and content from the outset based on detailed knowledge of the samples at the selection phase, otherwise often known as 'targeted contacting strategies' or 'targeted sampling techniques.' Including but not limited to:

- a. Machine learning used to identify best approach strategy
- b. Mail versus email versus SMS contact

### **3.3. Best Practice:**

#### **3.3.1. Collaboration**

There was an overwhelming view that more collaboration is needed and importantly across all sectors. As outlined in the Challenges section, this can be difficult. However, it is critical to success, remains relevant and encourages innovation. In discussions at the end of the conference suggestions were made to restart the GenPopWeb network to facilitate this cross-sector collaboration.

#### **3.3.2. Sharing of best practice:**

In particular:

- a. More guidance is needed about new modes of data collection
- b. Definition of quality and transparent assessment of quality are important
- c. Sharing what has not worked is just as important as what did work
- d. Sharing results quickly and less formally to get the research out there for others to be able to learn
- e. “once established, formalise it” – quote from the workshop.

#### **3.3.3. Inclusive design: accessibility and consideration of offline population**

Our surveys should be inclusive, and this should be built in from the start and not be an afterthought. The benefits of inclusive design extend to all respondents which is a bonus of the approach – the adaptations often make the experience clearer for all.

#### **3.3.4. Create respondent-centred (or user-centred) surveys**

Going back to basics, using mental models and insights from interviews and public testing to design and refine our surveys. We must not underestimate the importance of the ‘experience’ aspect of the respondent-survey interaction. Introducing self-completion means that we can no longer rely on highly trained interviewers to achieve response and to provide a good experience and that we need to refocus our design efforts on the respondent versus the data user in order to deliver successful online surveys.

#### **3.3.5. Adaptive or tailored design/ approach – not one size fits all**

This applies to:

- a. Overall survey design – important to design bespoke as each survey comes with its own context and set of unique challenges and opportunities
- b. The contact approach/ method, e.g. tailor the medium, messaging and mode

- c. Question design – for device and mode

### **3.3.6. Mixed-mode data collection (UK context)**

There is an accepted view that the future of survey data collection in the UK will have mixed-mode design, and that online mode is another mode to add to the options to choose from when designing a new or redesigning an existing survey.

### **3.3.7. Continuous review, forward thinking, looking to the future**

We must do this in order to remain relevant but to also ensure that what we are delivering is robust and meets the respondent and data user need. This is challenging to do on top of delivering in our day jobs. However, it is vital to success.

### **3.3.8. Evidenced based approach to design**

Due to time and budgetary pressures we have gradually slipped away from this approach. This conference concluded that there is a need for:

- a. More testing and experimentation
- b. Hypothesis driven concepts, theory leading to experimentation
- c. Test, learn, fail fast, iterate – i.e. agile development
- d. Marginal gains

### 3.4. Next steps:

During the discussion at the end of the last day of the conference it was suggested that the GenPopWeb network should be restarted. Since then a follow-up meeting has taken place with the original leads discussing the idea. Steps are being taken to explore options and the feasibility of this happening. Attendees of this conference will be updated on the outcome of this discussion.

### References:

- GDS** (2018) *Understanding new accessibility requirements for public sector bodies*. <https://www.gov.uk/guidance/accessibility-requirements-for-public-sector-websites-and-apps>. [Accessed on 24 August 2019].
- Nicolaas, G., Calderwood, L., Lynn, P., Roberts, C.**(2014) *Web surveys for the general population: How, why and when?* ESRC National Centre for Research Methods. Final Report. <http://eprints.ncrm.ac.uk/3309/3/GenPopWeb.pdf> [Accessed on 20 August 2019].