

Towards a modern mixed-mode Labour Force Survey

Trond Båshus, Statistics Norway

1. Abstract

The Labour Force Survey (LFS) at Statistics Norway is currently a CATI-only survey running in Blaise 4.8. Statistics Norway will phase out this version of Blaise in 2021 at the latest, which makes it necessary to convert the survey to Blaise 5. New EU regulations harmonizing European social statistics will also lead to changes in the LFS questionnaire. These two changes represent an opportunity to rewrite the questionnaire to support mixed mode data collection, but also to simplify the administration of the survey which currently is quite labour intensive. A pilot for a mixed-mode LFS survey was conducted in 2018, and the valuable lessons learned will be an important input for the development work.

This paper describes briefly the current LFS survey and the mixed mode LFS-pilot but will mainly concentrate on the changes planned for the new questionnaire. The goals for the new questionnaire are: 1. Offer a secure CAWI mode to the respondents, 2. A simpler and better structured Blaise questionnaire which should be easier to maintain, 3. More efficient data collection and simpler administration of the survey. We will also discuss changes to our case management system to facilitate mixed mode surveys, and other additions such as two-factor authentication for CAWI-surveys.

2. Introduction

Statistics Norway is at last on the verge to complete the transition to Blaise 5 from Blaise 4.8. We started using Blaise 5 for CAWI as far back as 2014 and made the first steps towards completing the move in 2018, when we started experimenting with CATI in Blaise 5. (Båshus 2016, Haslund 2018) Blaise 4.8 is now expected to be phased out during 2021. In this paper we will look closer on how LFS has been a driving force behind this development, despite being a survey many are reluctant making changes to.

3. Brief history of the LFS at Statistics Norway

The LFS has an almost 50-year history at Statistics Norway, and it has been carried out using a variety of modes through that time. The choice of modes has been dictated by technological possibilities, but also social developments, cost considerations and organizational issues has played a part. In 1972, when it was first fielded, the LFS was a uni-mode, or more precisely a PAPI, survey. The interviewer got lists of households and paper questionnaires in the mail and visited the respondents in their homes. As time went by and telephone coverage got better, the survey evolved into a mixed-mode survey: PAPI and PATI. If the respondents lived too far for the interviewers to visit and a telephone interview wasn't an option, the questionnaire could be mailed to the respondent, in effect making the survey a three mode survey: PAPI, PATI and PASI.

A major change came with the introduction of computer assisted interviewing (and Blaise) in 1996, which in essence worked as a computerized version of PAPI. A few years later, in 2000, a call centre was established. Most one-person households were routed to the call centres, while the rest of the LFS sample was sent to interviewers around the country. (Gravem 2011)

The next significant change in the LFS data collection happened in 2011/2012. A new case management system had been planned and built, and it was decided to move all interviewing to the Blaise CATI

environment, with most interviewers working from the same central database. A system for navigating between the members of the same household was developed, at the same time as each person in a household was treated individually in the CATI system. This is still how Statistics Norway collects data for the LFS. (Båshus 2012)

4. Current LFS questionnaire

Our current LFS questionnaire has been running since 2012 without any significant changes. The main form of interviewing is CATI interviewing from a central database, but offline interviewing is also possible, although the offline option hasn't been used for several years. Since every person in a household is a separate case in the database, only one household member at a time is normally active in the database. Appointments can be made for individuals in the household, and if there are several appointments within the same household, they will all be active in the daybatch. In case of offline interviewing the whole household would normally be sent to the same interviewer, and the household locked for online CATI-interviewing.

5. Current production system for social surveys

The data collection production system for social surveys at Statistics Norway is a combination of an in-house developed case management system which handles projects, samples, interviewers and payments. The system has good integration with Blaise 4.8, but the integration with Blaise 5 is still very basic. The consequence of this situation is that mixed-mode surveys must run in two different system: CATI on Blaise 4.8 and CAWI on Blaise 5. The reason for this is a combination of previously missing CATI functionality in Blaise 5 and lack of developer resources to update the case management system to work with Blaise 5. Working with mixed-mode surveys is therefore quite labour intensive since one must handle two questionnaires and two databases, which then have to be merged when the data collection period is over. (Båshus 2016)

6. Exploratory efforts

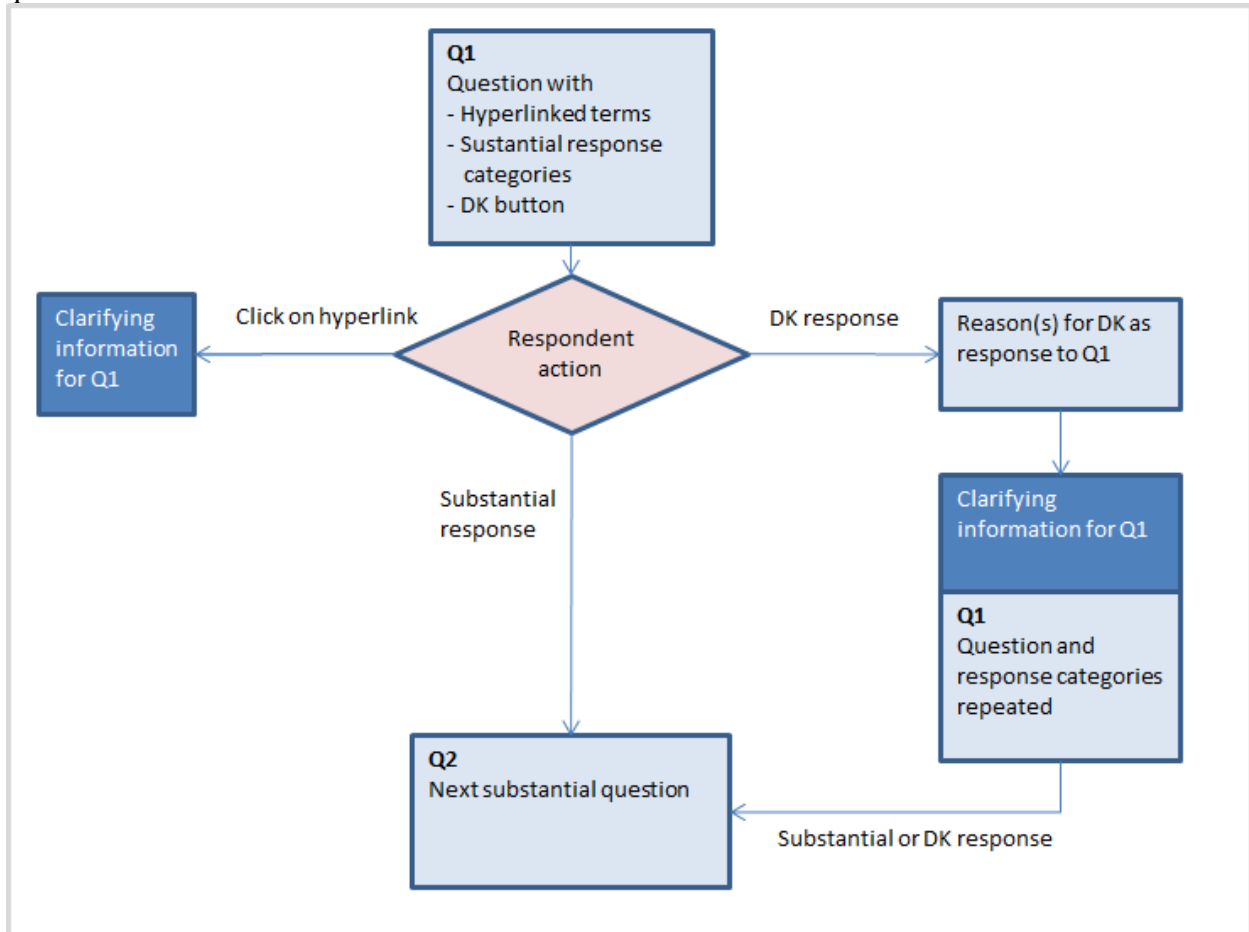
Statistics Norway has considered options for how to move as many surveys as possible to a mixed-mode platform for many years, and many surveys are offered on both CATI and CAWI, although the nature of the production system makes this less than optimal. Also, the nature of the LFS encourages a conservative attitude to changes in data collection. Despite this, we have undertaken several exploratory efforts, particularly under the umbrella of projects within the European Statistical System (ESS).

6.1 ESSnet DCSS

One such effort was an LFS CAWI questionnaire developed in 2012 and user tested piloted in 2013, as a part of the ESSnet project *Data Collection for Social Surveys using Multiple Methods* (DCSS). The aim was twofold. Firstly, to explore technical challenges concerning CAWI-surveys adapted to smart phones, and secondly to find alternatives to the interviewer-respondent dialogue in a CAWI questionnaire.

The survey was developed with Blaise 4.8, using the C-Moto stylesheet for mobile phones developed by CentERdata.

The survey contained hyperlinks on difficult terms which opened a second window with a definition and an explanation and contained follow-up sequences if the respondent answered “Don’t know” on selected questions.



A random sample of 1500 aged 18-65 was drawn, and 121 people responded to the survey. The questionnaire dialogue experiments showed that respondents are reluctant to use both the hyperlinks and the follow-up to “Don’t know”, but that the hyperlinks were most successful. One conclusion was that additional work had to be carried out to make questions clearer and less ambiguous. (Gravem 2013)

6.2 Mixed-mode pilot (2018/2019)

A large scale mixed-mode pilot of the LFS was completed in 2018/2019. The pilot was conducted in part with support from the European Commission. The purpose of the survey was to:

- Create and test a mixed mode version of the LFS
- Improve and explore best practice for mixed-mode data collection
- Identify necessary improvements for the case management system
- Explore strategies for adaptive-responsive data collection

6.2.1 Questionnaire

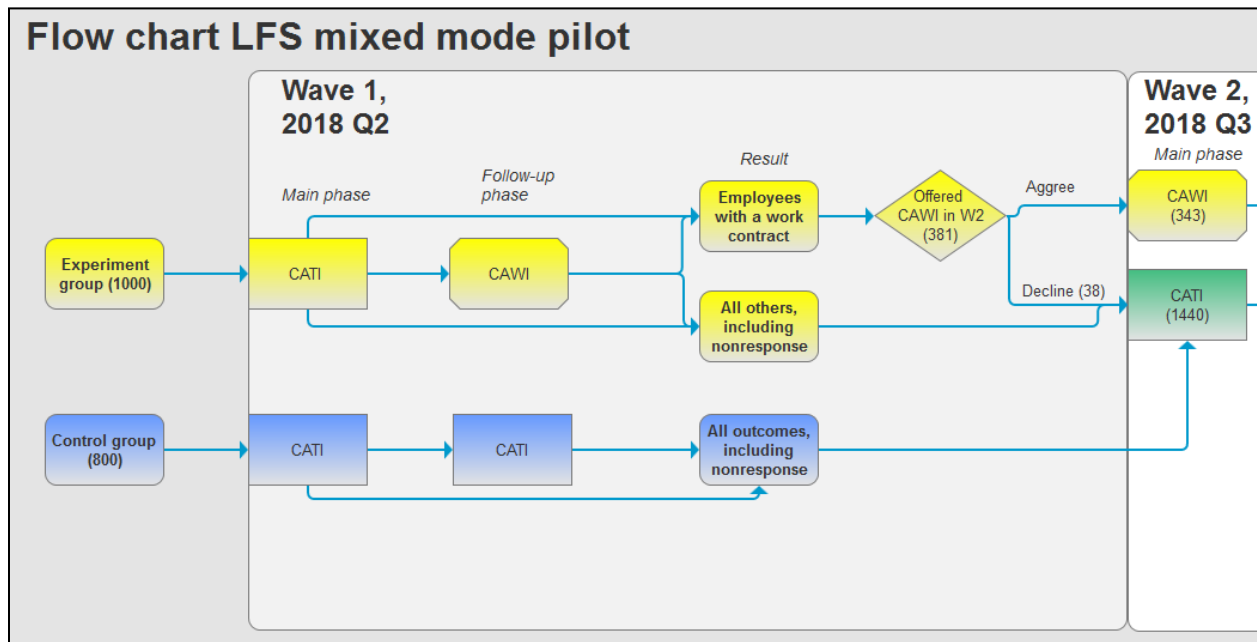
It was decided to convert the current Blaise 4.8 questionnaire to Blaise 5, instead of reusing the questionnaire made for the ESSnet DCSS project described above. The reason for this was that LFS

questionnaire in production is more complete, and that a comparison with the ordinary LFS would be easier if the questionnaires for the pilot and the ordinary LFS were kept (relatively) similar. The conversion tools supplied with Blaise 5 were used, and only minor changes were necessary to get the questionnaire to compile and run. Several questions which could be troublesome in a CAWI context were reformulated and the questionnaire was adapted to two modes, CATI and CAWI, and three layouts: One layout for CATI and two for CAWI. The CAWI layouts were a large layout for desktop computers and a small layout suitable for mobile phones. Since the pilot was to run for four rounds, the questionnaire was incrementally improved each round. Input from an interaction designer in cooperation with survey methodologists proved especially useful to update the look, feel and usability of the questionnaire.

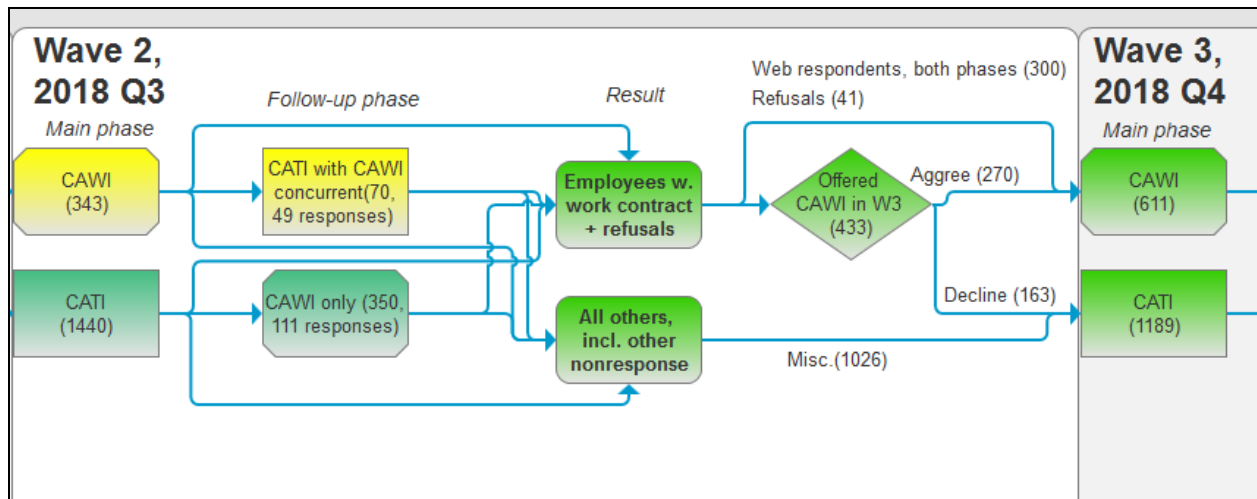
In addition to adapting the questionnaire to Blaise 5, it was also necessary to build an improvised production system to prepare data from previous waves and to prepare register information for the first rounds. This was necessary because it would have required too much work to adapt the ordinary production system for the pilot.

6.2.2 Waves, modes and mode switching

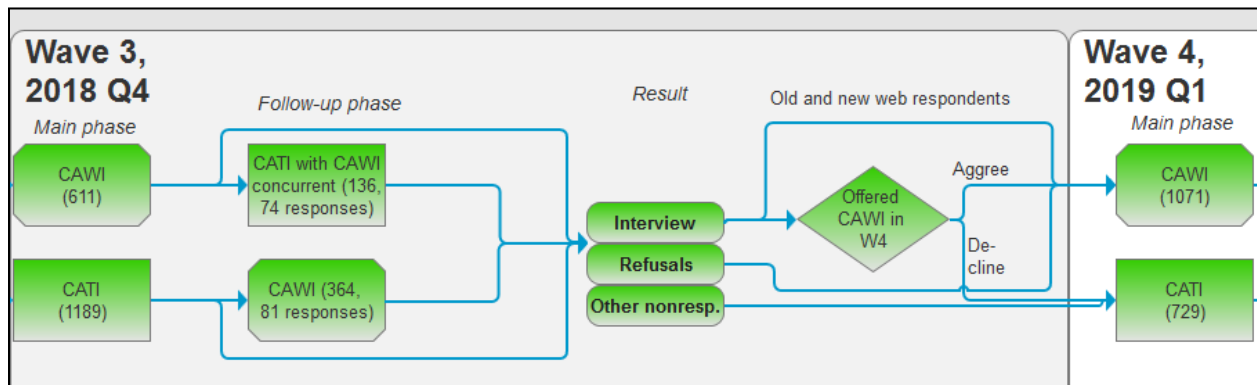
The sample was divided in two groups, an experiment group of 1000 and a control group of 800. For the first wave it was decided to offer CATI as the first mode for both groups, but the experiment group was offered CAWI in the follow-up phase. Employees with a work contract in the experiment group in the first round were offered CAWI mode for second round. The design of the survey was such that the number of respondents being offered CAWI as the initial mode got larger for each subsequent round.



Figur 1 Phase plan for first and second wave



Figur 2 Phase plan for second and third wave



Figur 3 Phase plan for third and fourth wave

6.3 Findings

The main findings of the pilot are that the potential cost savings offered by a mixed-mode CATI/CAWI survey, compared to a single-mode CATI, are considerable. The response rates as well are comparable to single-mode CATI, bearing in mind that the LFS is compulsory in Norway. 53% of the respondents in the pilot would prefer to answer the questionnaire on CAWI, rather than CATI (preferred by 29%).

Another finding is that the transition to mixed-mode data collection requires investments in important infrastructure such as the case management system. Despite the potentially large cost savings in moving to mixed-mode, as documented in the pilot, it can be surprisingly difficult to get funding and resources to do the necessary upgrades to core systems. (Gravem et al. 2019)

7. New questionnaire

The development work for a new LFS questionnaire is now well underway, and the main reason that this is happening in 2020 is the *Framework regulation for the production of european statistics on persons and households* or *Integrated European Social Statistics – IESS*. This new regulation requires significant changes to both the questionnaire and the production system behind it, changes which would be difficult

to implement just by modifying the current Blaise 4.8 LFS questionnaire. Also, it is an opportunity to move forward to a more up to date mixed-mode data collection, building on our own experiments and pilots surveys, and other institutions' experiences.

The case management system at Statistics Norway, Sivadm, has been in severe need of an update to support mixed-mode surveys and Blaise 5 for many years, but a general lack of IT-resources, and not least lack of will to give Sivadm priority, has led to a situation where Sivadm does not support our data collection processes in a way we want. A situation that has forced us to use considerable resources to improvise temporary solutions. Luckily, the new regulation affects several important social surveys, not only LFS, but also EU-SILC and others. The most pressing issues and missing functionality in Sivadm are now being addressed, and we expect the system to be ready for production this fall.

7.1 Standardization

We have used Blaise 5 for CAWI since about 2014/2015 and have during the last couple of years tried to harmonize and standardise the look and feel of our CAWI-surveys. A common resource database is used in most of our surveys, and we have also common templates (consisting of a blax and several incx files). At the moment we have a simple uni-mode template supporting only CAWI, and another mixed-mode template supporting CATI, CAWI, and at some point in the not too distant future, CAPI. Standardized templates are necessary for the integration with the case management system to work.

Typically, most mixed mode survey will be made with three different layouts: two for CAWI and one for CATI. The CAWI layouts are specifically made for either smartphones or desktop browsers.

Work on a layout suitable for interviewers was started when we first tested CATI in Blaise 5 for the survey *Governing and Experiencing Citizenship in Multicultural Scandinavia* (GOVCIT), this work was again taken further in the LFS pilot which we conducted in 2018/2019. (Haslund 2018) The interviewers haven't been entirely happy with the usability of the interface, especially when it comes to navigation within the questionnaire, and we have therefore made an effort to make it easier to use a keyboard. In most places, it is now possible to navigate just by using a keyboard or more specifically the numeric keypad. Among other improvements is a Click-to-call system, so that the interviewers don't have to enter the telephone numbers manually any more. This has been a long-standing item on our wish list.

Ringebilde

Statistisk sentralbyrå **Ringebilde** Statistics Norway Norsk

Ringehistorikk **Endre kontaktopplysninger**

IO-nummer: 50028
IO: DUCK LETTI, alder: 15, kjønn: kvinne
Telefon 1: 99999999 **Ring 1** **Ring 2** **Ring 3**
Telefon 2:
Telefon 3:
Adresse: KOSEKROKEN 4
Postnummer: 6060
Poststed: ANDEBY
Kommune: 1301
E-post: letti@duck.no

Periodenummer: 4
Kontaktperiode:
Delutvalg: 3
Tidligere runder:
ID-nummer: 10028
Passord: aaaaaaaa
Antall ringeforsøk:

Melding til intervjuer:
Avtale:
Avtalemelding:

Ringemeny:
 1 Intervju
 2 Ikke svar
 4 Avtale
 5 Frafall, avgang, overføring

Neste **Send e-post**

Figure 1 Dial questionnaire with Click-to-call buttons

Arbeidskraftundersøkelsen

Statistisk sentralbyrå
Statistics Norway

Arbeidskraftundersøkelsen

Norsk

Vet ikke (F8) Nekter (F9) Start Slutt Ringebilde Avtale Send e-post

Spørsmål: skjema.sekvens01a.innl1
io_nummer: 50028, navn: DUCK LETTI, alder: 15, kjønn: kvinne, telefon: 99999999, periodenummer: 4

Jobbet du noe i uka 20. - 26. januar?
Vi er interessert i alt arbeid, helt ned til én time i uka.

Velg en verdi 1 Ja
 2 Nei
 Vet ikke
 Ønsker ikke å svare

Forrige Neste

Figure 2 Example of a CATI questionnaire layout

The current LFS Blaise 4.8 questionnaire is quite old. Much of the code was written in 2006, but it also contains code going as far back as 1996. The questionnaire is rather complicated, and it does not help readability that many Blaise developers, with sometimes significantly different coding styles, have been working on it. One of the aims with the new Blaise questionnaire is to make better organized and more readable code, which also is better commented and documented.

Even though the questionnaire will be ready for mixed-mode in January 2021, it will be used a CATI only questionnaire, possibly for the first year. The questionnaire will be flexible enough for parallel or sequential mode designs. E.g. CAWI first with follow up on CATI, CATI first with follow up on CAWI, or both at the same time if that is desired.

The sample will change from a family sample to a person sample, but we will interview the persons belonging to the respondents' household in one of the waves. The Blaise questionnaire will not contain a household roster, and the household will instead be established on the basis of register information. The questionnaire will initially only be available in one language, but because of legal obligations, it must be made available in both written standards of the Norwegian language: *bokmål* and *nynorsk* when it starts on CAWI. English will also be an option, and possibly Polish. We anticipate that the new translation features in Blaise 5 will offload some of the work related to translation (copy and pasting of text) from the Blaise programmers.

As mentioned before, the case management system, Sivadm, is being updated to handle CATI in Blaise 5. The most crucial problem has been updating contact information such as addresses and telephone numbers between the two systems. Another improvement will be to make data about appointments available in the case management system. This will for instance make it possible to send out automatic reminders about hard appointments on SMS or e-mail, but it will also serve as a useful indicator of the future response potential in a survey.

A separate Blaise questionnaire is currently prepared quarterly, requiring us to prepare and install a new LFS Blaise questionnaire four times per year. From 2020 we hope to reduce this to one per year, easing the overall workload. Subsequent cases and waves can easily be uploaded as needed during the year. Data contained within the sample and external files will ensure that the questionnaire work properly.

7.2 Future improvements

For the use of CAWI on the LFS to be a viable option, the security is very important. Especially because the LFS will contain much information from registers and previous waves, which the respondent is asked to confirm or update. Two-factor authentication is therefore essential. The pilot conducted in 2018/19 used a form of secure login, and this worked well with very high response rates, at least in combination with a survey which is mandatory.

More advanced and automated use of audit trail and call data, for error detection and as an input to adaptive-responsive design are under development. Currently we are creating modules to parse unstructured paradata and call data from Blaise into Python Pandas Dataframe. DataFrame is a 2-dimensional labeled data structure with columns. Further, we will visualize those data with Plotly Dash.

Data delivery from the current LFS is completely automated and is in the form of a flat file which is exported from the Blaise database every night. This file is then imported into another database. In the future we will instead write the data directly to a database, removing an unnecessary step in the data delivery process.

8. Conclusion

The LFS data collection processes generally are conservative when it comes to changes, characterized by stability for usually a decade or more. Usually only small adjustments are made (and permitted). At the same time, it has also been a driving force behind experimentation and exploration of new methodologies and technologies, which in turn has benefited other surveys. When the new LFS survey is in place from 2020, we can perhaps expect the cycle to repeat itself with a long period of stability, while we can continue to prepare for and adapt to a changing technological and cultural landscape.

9. References

Båshus, Trond. *A system allowing sequential interviewing of household members, where the household members are individual cases, within the Blaise CATI-framework*. IBUC, London, 2012.

Båshus, Trond. *Blaise 5 at Statistics Norway*. IBUC, The Hague, 2016.

Gravem, Dag F. *Towards an integrated mixed-mode case management system for the Norwegian LFS*. Paper presented at the Workshop on Labour Force Survey Methodology, Wiesbaden 12-13 May 2011

Gravem, Dag F. *Error prevention through interviewer emulation? Introducing questionnaire dialogues in the Norwegian LFS questionnaire*. Paper presented at ESRA conference, July 16-19 2013, Ljubljana, Slovenia.

Gravem, Dag F. et. al. *Report from Statistics Norway's adaptive/responsive mixed-mode LFS pilot. Cooperation on Multi-Mode Data Collection (MMDC). Quality Improvements for the Labour Force Survey*. European commission. Publication forthcoming, 2019.

Haslund, Jan et. al. *Experiences with Blaise 5 CATI and multimode at Statistics Norway*. IBUC, Baltimore, 2018.