NubiS
Researcher Manual

Version 1.1

August 2016

Preface

Supporting every successful survey is a strong data collection tool.

At the University of Southern California's Center for Economic and Social Research (CESR), we have developed NubiS, accessible and versatile software for administering a questionnaire. NubiS:

- Runs on any server, PC, laptop or netbook, as well as on Android tablets or smartphones.

- Works with surveys that are self-administered, face-to-face and via telephone (i.e., all traditional modes of data collection).

- Collects continuous information from smartphones, tablets or other external sources, such as accelerometers, GPS devices and blood-pressure meters.

- Is programmable in multiple languages..

Following NubiS' development in 2014, it has been the foundation of CESR's Understanding America Study, a panel of thousands of households representing the entire United States. NubiS also has been used in several large-scale longitudinal surveys and other projects around the world.

You, as a researcher, want to design a survey and receive its results with a minimum of fuss. We think you will find NubiS provides the underpinnings for data collection of breadth and depth while remaining user-friendly.

This manual is the document's second version, supplanting the initial release distributed in Summer 2015.

NubiS is free software; it may be redistributed and/or modified under the terms of the GNU Lesser General Public License, version 2.1 and later, as published by the Free Software Foundation.

For more information, feel free to contact us. CESR also can help with access to the Understanding America Study's panel, as well as providing assistance for survey hosting, design and management.

We're proud NubiS can support the quest for greater knowledge and wish you the best with your surveys.

Sincerely,

The NubiS team

Table of contents

1. Logging on to NubiS

If you are working as a researcher with a NubiS project, you will have been provided with a website location and a username/password. To begin, direct a browser to the provided URL. There, you will see the login screen:

UAS SMS

Please enter your username and password to log in.

> Username

> Password

> Login

After you enter your username/password, the NubiS researcher interface will open:

UAS SMS   Reports   Data   Documentation   research ▾

Reports
Data
Documentation

There are three options available:

1. **Reports**: See reports on the collected data.
2. **Data**: Download data files.
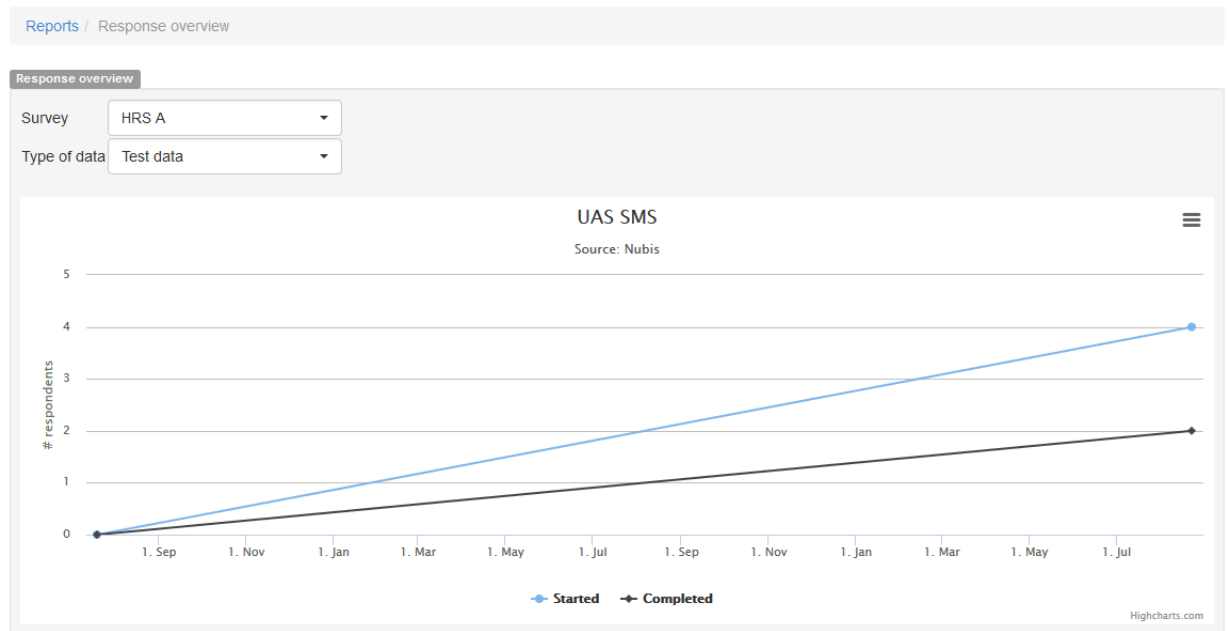3. **Documentation**: Look at survey documentation.

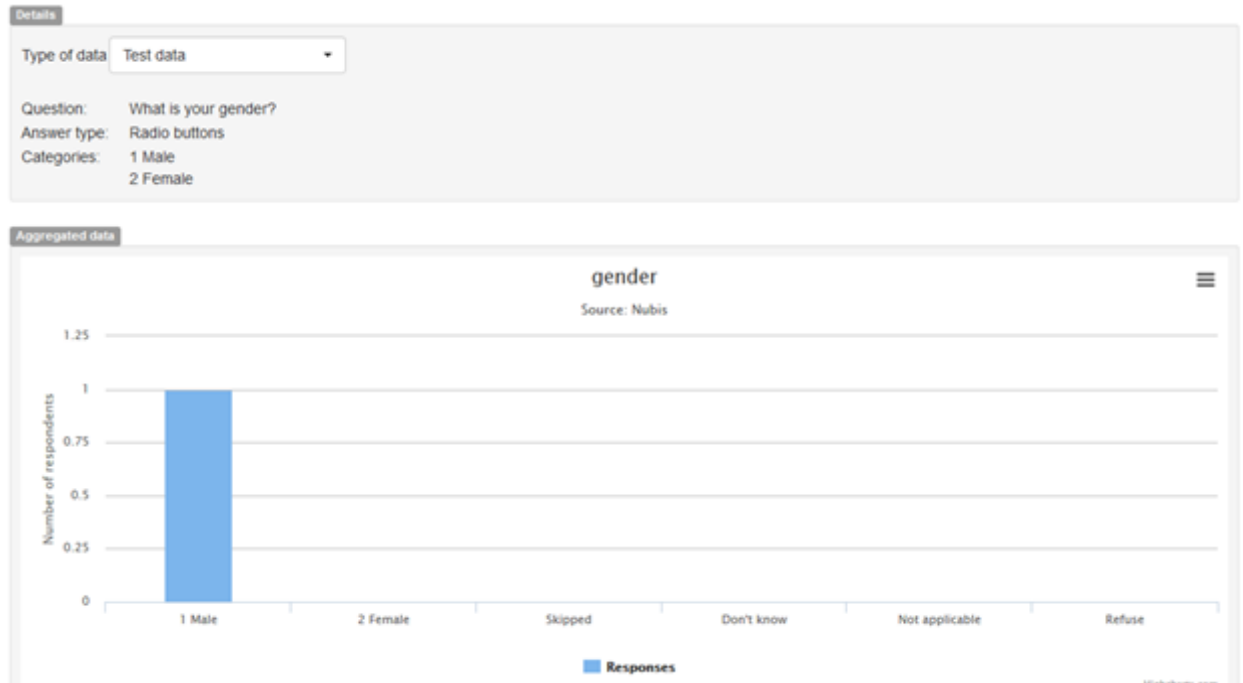# 2. Reports

Clicking *Reports* opens the following screen:

Reports

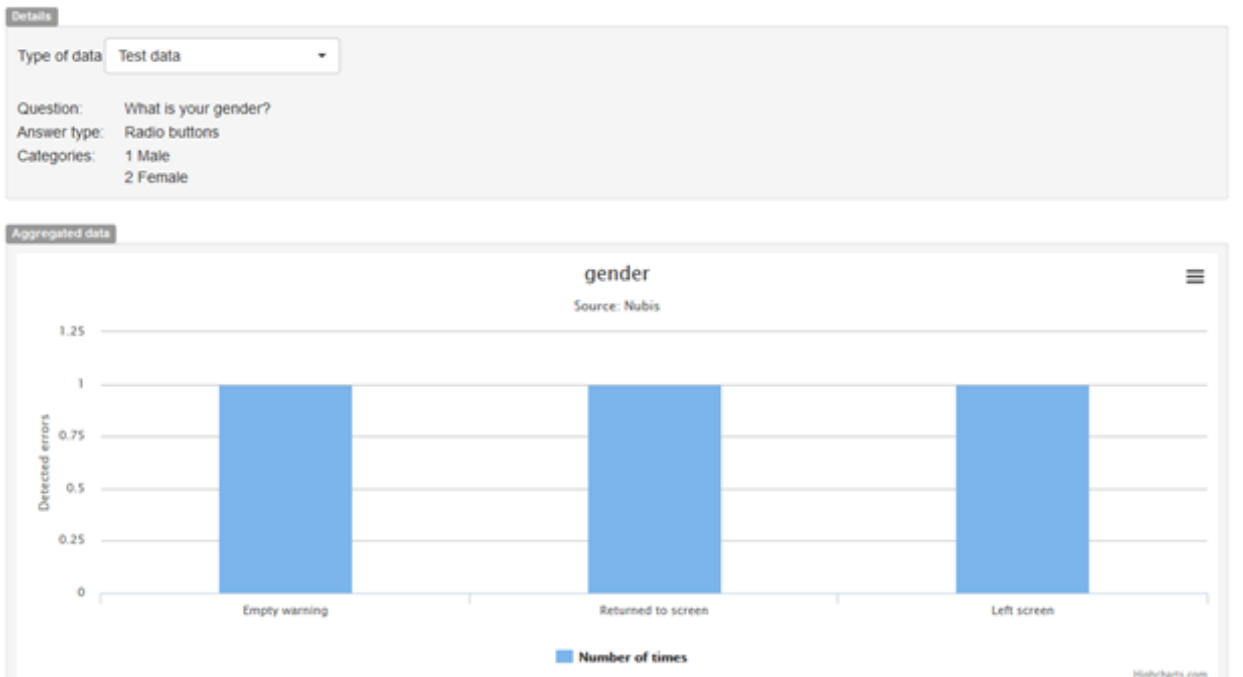| |
|---|
| Response overview |
| Aggregate data |
| Paradata |
| Timings distribution |
| Timings over time |
| Timings per screen per respondent |
| Platform information |
| Contact graphs |

You can view a variety of reports:

1. **Response overview**: view overall survey response rates. If applicable, you can choose a survey and the type of data (actual data or test data; that is, data collected during survey testing.


2. **Aggregate data**: look at results for a particular question.

Type of data Test data

Question: What is your gender?
Answer type: Radio buttons
Categories: 1 Male
               2 Female

gender

Source: Nubis

Number of respondents

1 Male   2 Female   Skipped   Don't know   Not applicable   Refuse

■ Responses

3. **Paradata**: look at paradata for a question, such as the number of times respondents received a "Please answer the question" warning or left the question screen.

Type of data Test data

Question: What is your gender?
Answer type: Radio buttons
Categories: 1 Male
               2 Female

gender

Source: Nubis

Detected errors

Empty warning   Returned to screen   Left screen

■ Number of times

4. **Timings distribution**: review the time spent by respondents completing their interviews.
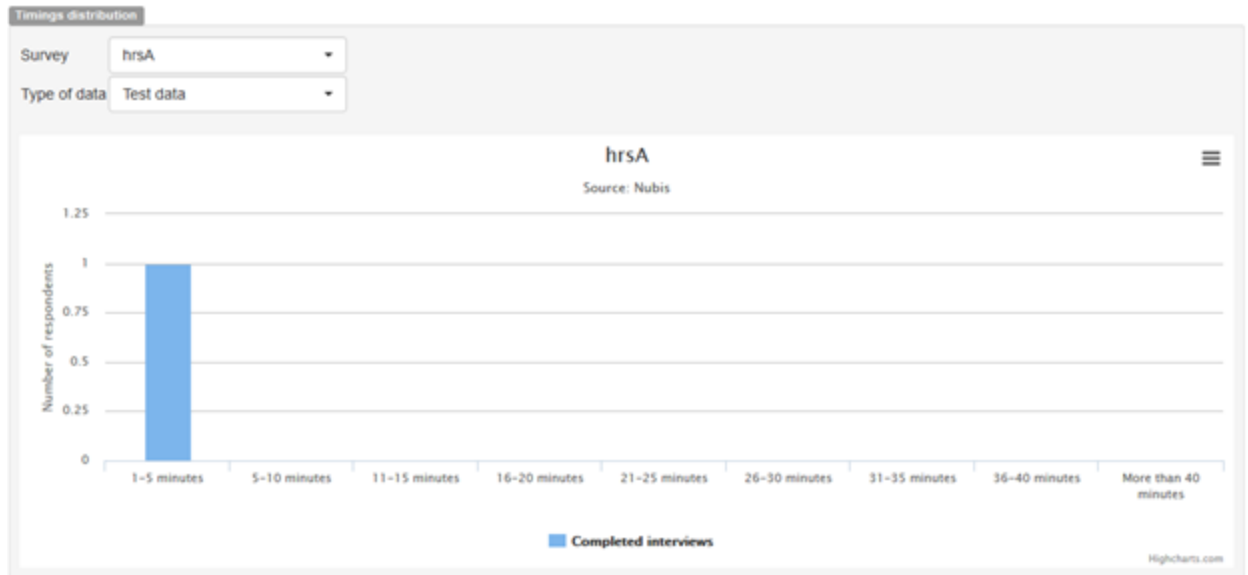
5. Timings over time: shows the time spent by respondents completing the survey over different dates.
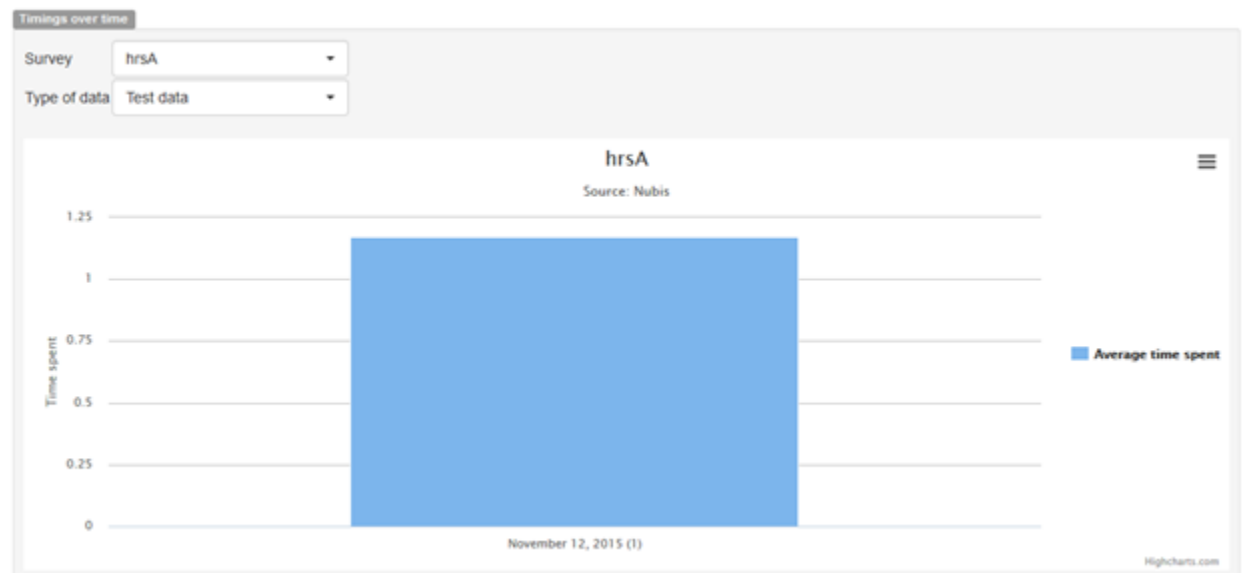
6. **Timings per screen per respondent**: shows the time spent by a particular respondent on a particular question screen.
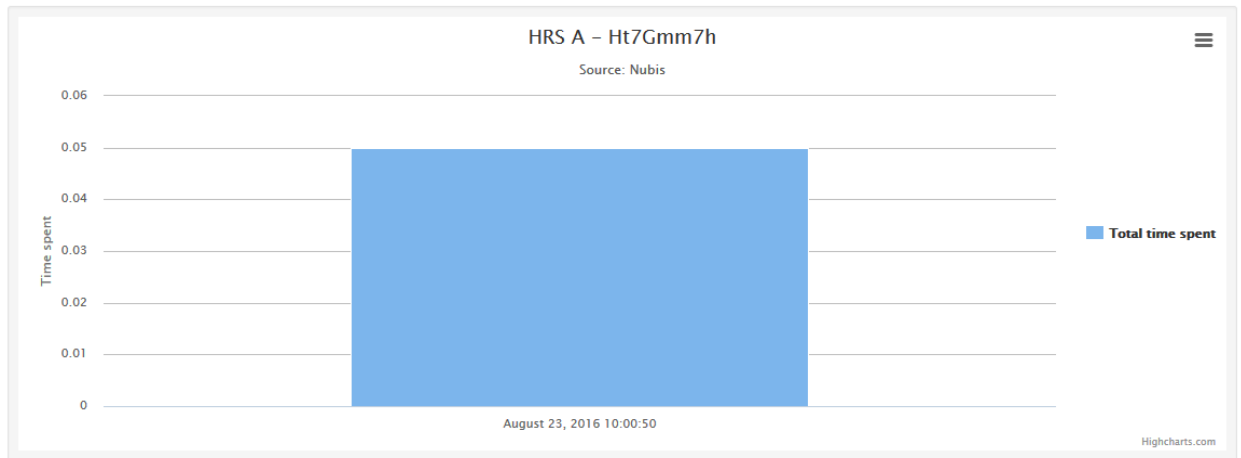
7. **Platform information**: review the platform, browser and operating system used by respondents.
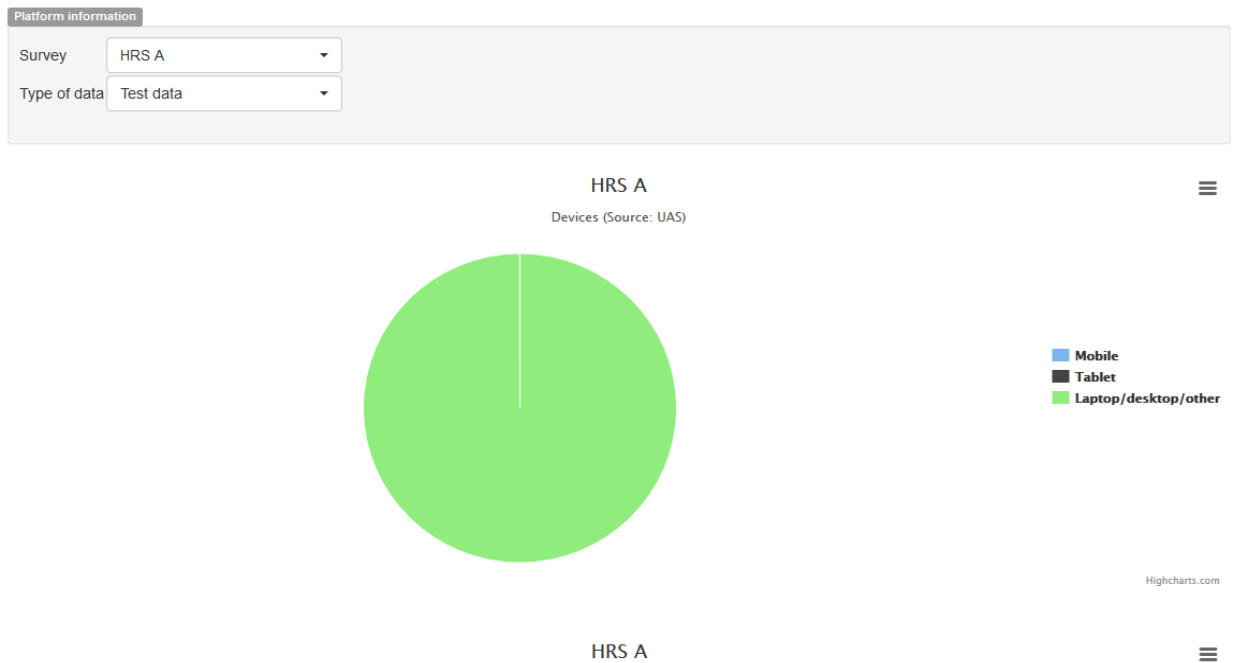
8. **Contact graphs**: show how many contacts were made during telephone interviews.

3. Data

Data output in NubiS comes in the following forms:

UAS SMS	Reports	Documentation	Data	Unassigned sample

Data
Raw data
Timings
Paradata
Auxiliary data

Raw data is the data directly collected through the survey in terms of answers to questions, flag variables being set and so on. On the *Raw data* screen, you can specify the output parameters NubiS will follow when generating a data file:

Source

Database table	Data record table ▼
Survey	HRS A ▼
Mode(s)	🌐 Self-administered ▼
Language(s)	🇺🇸 English, 🇪🇸 español ▼
Type of data	Actual data ▼
From:	📅
To:	📅
Interviews	All interviews ▼
Clean data only	Yes ▼
Kept data only	No ▼
Exclude hidden variables	Include ▼

In part, these parameters concern from where data should be pulled and under which restrictions. First, you can choose whether to base the export on the data record or data table. In virtually all situations there is no difference, but if you were to insert data into the data table directly during your survey (and not through the usual NubiS internal storage mechanisms), you can only include that data by basing the export on the data table.

Next, you can choose the survey for downloading data, from which interview mode(s) and which language(s). By default, all available modes and languages are selected. You can choose whether

you want to download real or test data, and whether to include all interviews or only those completed. You also can decide between clean data or dirty data. Dirty data is data that may contain answers that are no longer valid, because the respondent answered a question then went back in the survey and altered previous answers, invalidating their response. NubiS will not include such invalidated data if clean data is desired. Similarly, you can say whether you only want to have "kept" data; that is, data for variables for which answers are kept upon survey completion. Lastly, you can limit the number of outputted variables by excluding variables you instructed NubiS to hide in data output.

Upon deciding the source settings, you can specify how the data should be output: Stata or CSV.

You can specify a particular file name; if left empty, NubiS will generate one based on the survey title. Also, you can indicate whether to include the primary key and whether to encrypt it. You may wish to exclude the key or encrypt it if you don't want the person receiving the data to know the identifiers. You can define whether to include variables that have no data associated with them, whether variable names should be lower- or uppercase, and, for downloading a Stata file, whether to include value labels and answer option codes in those labels.

Finally, you can instruct NubiS to mark skipped answers if the survey is programmed to allow respondents to leave a question's answer empty. If **Mark skipped answers** is set to *No*, then in the outputted data file this empty answer will be represented with a period in Stata and an empty value in CSV. But suppose a respondent never saw a question due to a skip pattern. When NubiS outputs data for this question, it does not find a value and outputs it, again, as a period in Stata and an empty value in CSV. With no distinction between the two and unknown whether the respondent did not get the question or chose not to answer, this has consequences for analysis in terms of what can be done with empty values. To make a distinction, NubiS can mark empty answers given by the respondent. This can be done within the same variable, in which case they are outputted both in Stata and CSV as **.e**. This also can be done by creating skip variables, such as **Q1_skip** for *Q1* where the skip variable is set to *1* if the corresponding question was skipped, and, otherwise, to empty.

When the settings are to your liking, click **Download**. NubiS then will generate the data file based on the raw data as it is at that moment. How long this takes depends upon the size of the

survey and the number of interviews. When done, NubiS will prompt for the generated file to be downloaded.

You also can download a file containing detailed timings information:

The main data sets created already have each interview's begin and end time, but this might not give an accurate picture of how long someone took to fill out the survey. For example, a respondent may have logged off and returned the next day, or went out for a moment and came back 10 minutes later. The timings file NubiS generates tells how long each respondent spent on each individual question screen. Based on that, it totals the overall time spent by the respondent on the survey. This information can be used, for example, to inform whether a respondent actually read the question. If the average time spent by all respondents is 45 seconds and someone answered the question in three seconds, then it could be reasoned this individual did not properly read the question.

Note: the listing of time spent per individual question screen includes all captured time measurements. In contrast, the average time spent by a respondent and the average time spent on a question exclude any time measurement exceeding a five-minute period (to use a different cut off period, re-configure NubiS' *getTimingCutoff* function in config.php).

Other types of data files that can be downloaded contain paradata and auxiliary data. The screen for downloading paradata is as follows:

Source		
Survey	HRS A	▼
Mode(s)	👤 Face-to-face, 📞 Telephone, 🌐 Self-administered	▼
Language(s)	🇺🇸 English, 🇪🇸 español	▼
Type of data	Actual data	▼
From:		📅
To:		📅

Format		
Filename		File extension automatically appended
File type	Stata	▼
Type	Raw (CSV only)	▼
Include primary key	Yes	▼
Encrypt primary key with		Leave empty for no encryption

Download

Paradata can be downloaded in a processed or raw format. In its processed format, a Stata or CSV file is generated containing variables summarizing how often a particular type of error occurred for a specific question for an individual respondent. Also provided are variables capturing how often the respondent left the question screen and came back again (if applicable). In contrast, the raw file contains all collected paradata strings. These strings follow this format:

||FI=1447279216935||ER1:answer1=1447279218058||ER1:answer2=1447279218058|| MC:801:197:1:nextbutton=1447279218083

As the example illustrates, a paradata string comprises a series of actions separated by two vertical lines: ||.Currently, there are the following types of action:

- FI: browser tab displaying the survey received focus. It comes with timestamp in milliseconds passed since midnight of Dec. 31, 1969.

- FO: browser tab displaying the survey lost focus (*e.g.*, the respondent opened another browser tab or another application altogether). It is timestamped in milliseconds passed since midnight of Dec. 31, 1969.

- MC: mouse click on the survey browser tab. The first two numbers after **MC** indicate, respectively, the X- and Y-coordinates. The third number identifies the button pressed (*1*=left button, *2*=middle button, *3*=right button). The string after that identifies the component clicked (if the component's HTML element had a name attribute). Again, it is timestamped in milliseconds passed since midnight of Dec. 31, 1969.

- **MM**: mouse click on the survey browser tab. The first two numbers after **MM** indicate, respectively, the end X- and Y-coordinates. The string after that identifies the component clicked (if the component's HTML element had a name attribute). Lastly, it is timestamped in milliseconds passed since midnight of Dec. 31, 1969.

- **ER**: error recorded for a particular input box. Each error action comes with the component's name and a timestamp in milliseconds passed since midnight of Dec. 31, 1969. These names relate to the value of the **displayed** field in order of appearance. For example, *answer2* corresponds to the second field in the *displayed* list. The following errors currently are logged:
    - ER1: no answer given
    - ER2: text (string/open) question answer pattern not satisfied
    - ER5: text (string/open) question answer not enough characters
    - ER6: text (string/open) question answer too many characters
    - ER10: numeric question outside of range
    - ER11: custom question outside of range
    - ER16: numeric question not a numeric answer
    - ER17: integer question not an integer answer
    - ER24: text (string/open) question answer too many words
    - ER25: text (string/open) question answer not enough words
    - ER27: set of enumerated question not enough options selected.
    - ER28: set of enumerated question not exact number options selected.
    - ER29: set of enumerated question too many options selected.
    - ER30: set of enumerated question invalid sub-combination of options selected.
    - ER31: set of enumerated question invalid combination of options selected.
    - ER32: multi-dropdown question not enough options selected.
    - ER33: multi-dropdown question not exact number options selected.
    - ER34: multi-dropdown question too many options selected
    - ER35: multi-dropdown question invalid sub-combination of options selected.
    - ER36: multi-dropdown question invalid combination of options selected.
    - ER37: more than one inline field answered.
    - ER38: not all inline fields answered.
    - ER39: not enough inline fields answered.
    - ER40: too many inline fields answered.
    - ER41: not exactly enough inline fields answered.
    - ER42: option selected, but inline field not answered.

- ER43: more than one question in the group answered.
- ER44: not all questions in the group answered.
- ER45: not enough questions in the group answered.
- ER46: too many questions in the group answered.
- ER47: not exactly enough questions in the group answered.
- ER48: not all questions in the group have a unique answer.
- ER49: not all questions in the group have the same answer.
- ER50: invalid answer option code entered in enumerated text box.
- ER51: invalid answer option code entered in set of enumerated text box.
- ER52: question answer not equal to specified value.
- ER53: question answer equal to specified value
- ER54: question answer smaller than specified value.
- ER55: question answer smaller than or equal to specified value
- ER56: question answer greater than specified value
- ER57: question answer greater than or equal to specified value
- ER58: set of enumerated/multi-dropdown question answer not equal to specified value(s).
- ER59: set of enumerated/multi-dropdown question answer equal to specified value(s).
- ER60: set of enumerated/multi-dropdown question answer smaller than specified value.
- ER61: set of enumerated/multi-dropdown question answer smaller than or equal to specified value
- ER62: set of enumerated/multi-dropdown question answer greater than specified value
- ER63: set of enumerated/multi-dropdown question answer smaller than or equal to specified value
- ER64: string question answer not equal to specified value (case sensitive).
- ER65: string question answer equal to specified value (case sensitive).
- ER66: string question answer not equal to specified value (ignoring case).
- ER67: string question answer equal to specified value (ignoring case).
- ER68: date/datetime question answer not equal to specified date.
- ER69: date/datetime question answer equal to specified date.
- ER70: date/datetime question answer before specified date.

- ○ ER71: date/datetime question answer before or equal to specified date.
- ○ ER72: date/datetime question answer after specified date.
- ○ ER73: date/datetime question answer after or equal to specified date.
- ○ ER74: time question answer not equal to specified time.
- ○ ER75: time question answer equal to specified time.
- ○ ER76: time question answer before specified time.
- ○ ER77: time question answer before or equal to specified time.
- ○ ER78: time question answer after specified time.
- ○ ER79: time question answer after or equal to specified time.

The screen for obtaining an auxiliary data file does not differ much from those seen so far:

Source

Database table	Data record table
Survey	HRS A
Mode(s)	🌐 Self-administered
Language(s)	🇺🇸 English, 🇪🇸 español
Type of data	Actual data
From:	📅
To:	📅
Interviews	All interviews
Clean data only	Yes
Kept data only	No
Exclude hidden variables	Include

Format

Filename		File extension automatically appended
Include primary key	Yes	
Encrypt primary key with		Leave empty for no encryption

Like for raw data files, you can define the conditions for the source data to use, as well as the output conditions. NubiS will generate a CSV file for downloading, and the file will show per variable per respondent in what interview mode and language it was collected.

4. Documentation

To have a better understanding of the survey and the data, you can look at the survey documentation:

The dictionary provides an overview of the variables within the survey and their answer criteria (for example, answer options, or minimum and maximum). The routing gives an overview of the survey programming and helps understand survey skip patterns.