

Part 7 - Designing and Implementing a Web Component

Andrew L. Hupp - Survey Research Center, University of Michigan

1. Background

As noted earlier, the Health and Retirement Study (HRS) is a longitudinal study of over 20,000 adults age 50 and over that collects health and economic information. Prior to the 2018 main data collection, the HRS has collected its data using traditional interviewer-administered telephone and face-to-face modes. As part of the 2018 main data collection, the HRS introduced a self-administered web response option to the mix.

2. Design

2.1 Sample

A subset (2,247) of cases assigned to the telephone mode for the 2018 wave were selected to receive the web response option. Cases were randomly assigned to one of four replicates (see table 1) based on when all members of the household completed their interview (related lines are assigned to the same replicate). The first replicate was further sub-divided into two parts (a & b). The smaller replicate (1a) was used to ensure the system was working as expected before releasing additional cases. Cases came from across the different cohorts (i.e., age groups; AHEAD is the oldest, LBB the youngest) in the HRS.

Table 1: Replicate Assignment

Cohort	Replicate 1a	Replicate 1b	Replicate 2	Replicate 3	Replicate 4	Total
AHEAD	0	5	12	6	0	23
CODA	0	12	7	8	1	28
HRS	51	153	157	169	16	546
WB	0	93	79	69	20	261
EBB	0	173	178	186	40	577
MBB	0	217	260	242	92	811
LBB ¹	0	1	0	0	0	1
Total	51	654	693	680	169	2,247

2.2 Experiment

As part of the 2018 data collection a two factor non-contact experiment was implemented. Reminder interval timing (short (~6 day) v. long (~12 day)) was contrasted with mode switch timing to phone (short (6-week) v. long (12 week)). Cases were assigned to one of four non-contact sequences.

1. Short interval (~6 days) between reminders; switch to calling after 6 weeks
2. Short interval (~6 days) between reminders; switch to calling after 12 weeks
3. Long interval (~12 days) between reminders; switch to calling after 6 weeks
4. Long interval (~12 days) between reminders; switch to calling after 12 weeks

All cases are assigned to web first. Non-complete cases are followed-up with by telephone (phone is available immediately to accommodate call-ins, but are not actively delivered for calling until the 6 or 12 week mark depending on their assignment). Once contact² has been established, reminder follow-ups are

¹ The Late Baby Boomer cohort was excluded from the web since there baseline interview occurred during the 2016 data collection. One case was selected since it was part of a household with a member in another cohort that was selected.

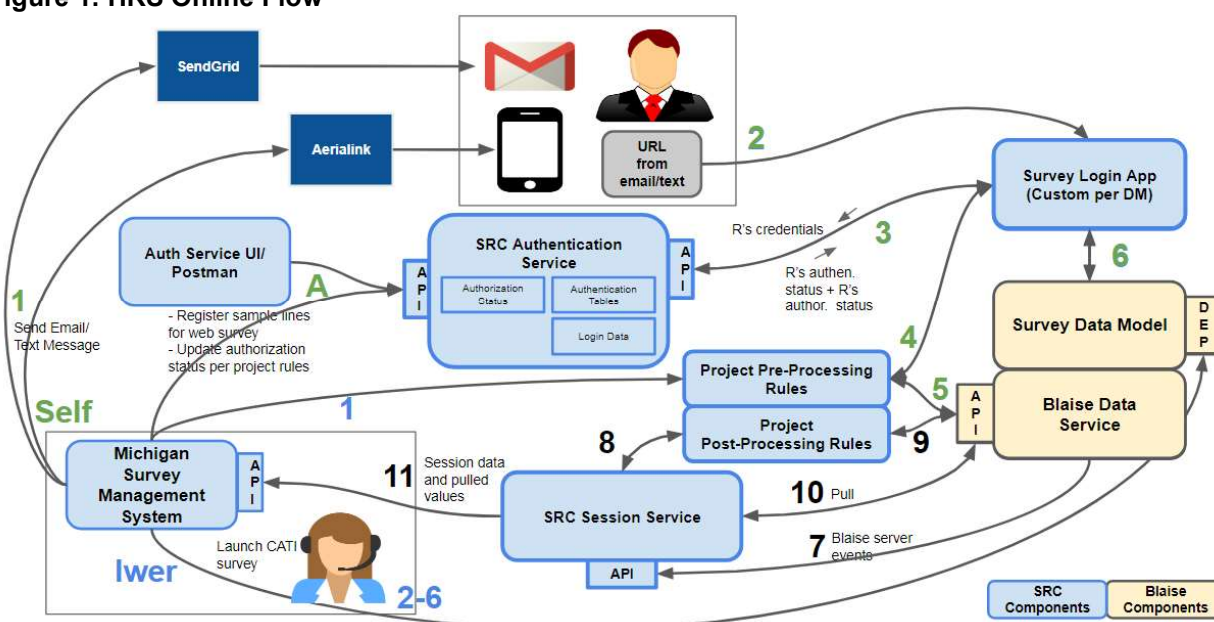
² Defined as completion of Sections A and A2.

based on their last activity (e.g. days since they last accessed the survey) rather than at a prescribed time. Contact cases follow the pattern of non-contact sequence one. They may receive the same amount of communications (2 letters and 6 emails) as that group but the timing of those reminders is dependent on their actions (or inactions). If they methodically answer a few questions every day after the first email, they will never receive another reminder from us. If the survey is not completed via web they eventually switch modes to phone to be followed-up with by an interviewer. That transition will always be longer than 6 weeks since there was at least a one delay based on having been in the survey.

3. Systems

The technical ecosystem used for the web component can be seen in figure 1.

Figure 1: HRS Online Flow



3.1 Michigan Survey Management System

The Michigan Survey Management System (MSMS) (see figure 1 in lower left in blue) manages the study protocol. It communicates with third party services to send communications (emails, text messages, etc.) when criteria in the protocol have been met to send the next communication. Interviewers launch Blaise to administer the phone interview from MSMS. Contact attempts entered by the interviewer or created by the system are stored in MSMS for interviewers and managers to see. MSMS also stores information pulled back from the instrument at the end of each data collection session. Some of this information is used to determine the next step in the protocol.

3.2 SendGrid

SendGrid (see figure 1 in upper left in blue) is the third party email provider used to send emails. The protocol loaded in MSMS determines when the next email to a case should be sent. When the criteria have been met MSMS communicates with SendGrid (via an API) and sends the relevant information (email address, email template, etc.) needed to send the email. SendGrid sends status information (sent, not sent, opened (if this feature is enabled), etc.) back to MSMS.

3.3 Survey Login App

The survey login app (see figure 1 in upper right in blue) and its associated services are needed to determine whether a case trying to access the survey should be allowed in. The survey login app provides feedback if the case is not allowed in (invalid credentials, out-of-date browser, survey completed, etc.). The survey login app captures the device and communication (which email) used during the login attempt and the result of that login.

3.4 Project Processing Logic

The HRS has project specific processing logic (see figure 1 in middle right in blue) that does two things. Prior to the launching of a case it checks to see if there is a related sample line and whether than survey has completed sections A and A2. If a case does exist and those sections are complete is takes the shared information between the cases and updates the values based on the first respondents responses. At the conclusion of a survey the logic checks to see if there are any adjustments to the household roster. If a new spouse has been entered a new sample line is generated for that newly entered person.

3.5 Blaise 5

Blaise 5 is used for the instrument (see figure 1 in middle right in tan). Prior versions of the HRS were programmed in Blaise 4.8. Given the mixed mode nature of the 2018 design, the instrument was upgraded to Blaise 5. There is one data model for both the web content and phone content. Most of the survey content is the same across modes, but there are questions/content that are mode specific.

4. Data Collection

Data collection began May 30, 2018 and goes through the first quarter of 2019. The instrument is lengthy. On average, it takes about two hours to complete. While this may seem long, these respondents are members of a longitudinal study and know how long the survey takes.

Given the length, one might expect the respondents to suffer from fatigue and complete the survey in multiple sessions. This however, is not the case. About 55% of completed cases complete the survey in a single session. Over 75% of completed cases complete in two sessions or less.

Respondents are using a variety of devices to complete the survey. Over 85% complete the survey using a PC, about 11% using a tablet, and over 3% using a smartphone. It should be noted that the survey is not optimized for the smaller screen sizes of smartphones. When a respondent lands on the sign-in page, the device being used is detected and a message informing them that the survey is best viewed on a device with a larger screen. A respondent is not prohibited from continuing with the survey on the smaller screen.

Respondents also use a variety of browsers to complete the survey. Over 60% of cases complete the survey using Chrome, about 12% using Internet Explorer, about 15% using Safari, and about 12% using Firefox. It's important to know the browsers being used as Blaise 5 does not support older versions of some browsers. The browser being used to attempt the login is detected. If it's a version not supported the sign-in page informs the respondent of that. The respondent can then call-in an request assistance in downloading a newer version or ask to complete the interview via phone.

Data collection has been successful to date with over 60% of cases completing the survey via the web.

5. Challenges/Lesson Learned

While data collection has gone relatively smoothly we have encountered some challenges along the way. This final section details the challenges, lessons learned, and things a survey practitioner need to keep in mind when implementing a mixed mode study that includes a web component.

5.1 Mode Switching

On the surface this may seem like a straightforward thing to do. On the HRS we have a sequential design where cases are assigned to web first and then switch to phone. As mentioned earlier we have phone available immediately in the event of a call-in requesting a switch from web to phone, but do not actively deliver a case to be called by an interviewer until a specified time. The instrument is designed to take advantage of the mode (web or phone) of administration.

For example, for the phone interview we require responses to all questions, on the web we allow the respondent to skip (not answer) questions if they choose. This creates problems for suspended cases, especially in the switch from web to phone. Since questions in the phone interview are required, any skipped questions from web need to be re-asked over the telephone. This may lead to questions being asked out-of-context or a question that does not make any sense. For instance, the HRS has a series of questions that ask a respondent to subtract a certain number from the previous amount. If the respondent suspends during that battery of questions, they will have no idea how to answer the question since they may not remember what their previous response was during their last session.

The HRS instrument also has questions meant for the interviewer at the end of each section that ask if the respondent needed any assistance. Those questions are now on route in the phone survey and need a response. This adds to the interviewer and respondent burden trying to easily and quickly get back to the last question the respondent was presented.

To address the issue, interviewers need to be trained how to handle such situations, and/or extra coding in the data model may be needed to assign values to skipped web questions.

5.2 Pulling and Updating Information

In our implementation, we pull back and use information from the Blaise instrument to determine the next action in the protocol. We have had instances where a respondent inadvertently started their spouse's line. We have a process for resetting those instruments. Doing this only addresses the Blaise data, not any of the data pulled back to the management system. The data in the management system may also need to be updated if information was used from the instrument that is no longer valid.

5.3 Survey Access

Time needs to be spent thinking through survey access. A good survey login app is valuable. Most importantly, it allows the survey organization to control access to the web instrument. We use it to provide useful information to the respondent. As mentioned earlier, we detect which device and browser they are using to provide feedback that will help in aiding them with having the best experience. One needs to make sure they are passing in the proper information to make sure the correct display is shown to the respondent. During testing we discovered that it was possible to see the interviewer layout set during a self-administered survey. The survey login app was retaining the last layout set used (make sure you test!).

Since we have multiple respondents in a household we spent time thinking about access to the survey. The HRS has some shared preload amongst household members. That information is set in the first two sections of the instrument. If a respondent is currently in one of those two sections the spouse is blocked

until the respondent has either completed those two sections or the server timeout has occurred due to inactivity.

Even though the design is sequential (web first, then phone), we allow a phone interview as the respondent's request. The management system has built in capabilities to only deliver cases that do not have an active Blaise session. We implemented this feature for two reasons, 1) We didn't want the interviewer to bother a respondent who was working on the survey, and 2) we did not want the respondent following along in the web version inadvertently updating data since there is only one data model.

5.4 Timeouts and Redirects

Timeouts are essential for keeping data secure. Initially the Blaise (server) timeout was set to 20 minutes. It eventually was reset to 21 minutes due to an issue that occurred with additional features we created related to the timeout. We alert respondents two minutes (at 18 minutes) prior that their session was about to timeout for security reasons. A pop-up message alerted the respondent that the session was about to end. If the respondent clicked "Ok", the clock reset and they could continue. If the timeout occurred, the respondent was sent to a custom GoTo page that contained a link that redirected them to the survey login app. There timeout was happening inconsistency with our added features so we increased the Blaise (server) timeout to 21 minutes and had a project specific timeout of 20 minutes, with a warning at 18 that triggered before the server timeout happened.