Great. And now, this brings us to our final presentation, which might allow us to end on time, even, Dave Vannette who has been organizing all of our festivities all along – go right ahead – has done an analysis of the usability of the websites –

Interesting – usability of the websites of the big three surveys. And, you know, as you know, one of the themes that Steve touched on and that we want to have as a part of this subcommittees report is the usability and access of the data, the documentation and so on. And so, Dave was –

– generous enough to put this analysis together for us.

Alright. Hi, everyone. Thank you, Jon, for the introduction and it’s a great privilege for me to have the opportunity to present, alongside so many prestigious names and fantastic researchers. So, about a month ago, when Jon and I started talking about doing this study and collecting feasibility research, we decided the best approach that we could think of was to actually bring in a bunch of people and have them do some fairly simply, some more complex, tasks with the big three websites, the NES, PSID and GSS. And we did this in our lab with screen capture technology going so that we could, kind of, observe what kinds of patterns emerged in terms of the kinds of problems that people ran into, what works on these websites, what’s not working on these websites. And then, kind of, what kind of principles about usability can we extract from these analyses?

So, what I’m going to show you is, kind of, the outcome of what we observed in that regard. And so, I’m going to try to use, mostly, examples from the big three, but I’m also going to deviate and show some examples from the ICPSR website, which I think is a much better example of a useable website. So, I think you’ll notice, there’s going to be a bit of, kind of, dovetailing, maybe even some overlap with the things that I have to say with what Steve Ruggles just presented in terms of the importance of making different types of data and documentation easily accessible and usable by people. I’m not going to focus on preservation at all, but mostly just on the usability of these things.

So, here’s where we’re going to go today. I’m going to talk a little bit about an introduction to usability, just so that we’re all on the same page about what I’m talking about. I’m going to talk briefly about what survey users want. What are the average users coming to these websites looking for? And then, I’m going to – the core of
the talk is going to be on principles of usability and the current state of the big three survey websites. And so, I’m going to walk through some principles using examples, mostly screenshots, and then a video or two if we have time, of some of these screen capture things that we did. And then, I’m going to discuss some areas for improvement and then, some conclusions.

So, first, why should we care? I think it’s important to motivate this discussion with why we should care. I know, so, Matt DeBell is here. He works with the ANES and actually has a fair bit of influence on when they decide to update their website and things like that. But, I don’t know if the rest of us have direct influence over survey websites or not. Maybe you sit on their boards or things like that and can have influence. But, I think we all should care about this for a couple of important reasons, the first which Steven Ruggles talked about, I think, very convincingly, is dissemination. And from my perspective, another important aspect of this that he didn’t really touch on quite as much is I think it’s really important for the field of survey research to see an expansion of our user base in terms of the people who are using survey data. And usable websites can really influence the ease of access of data and documentation and, hopefully, we can bring new users into these fields so that, you know, we see more use of our data and then, hopefully, more funding of data collection and things like that.

And the second point is a little bit self-serving. I actually use these websites a bit and download data sets and find documentation and track things down and it’s not easy. It’s like these tasks are hard. And so, making secondary research easier to conduct, I think, would also be, you know, an important motivation for all of us, you know? And those of you that have graduate students, those of you that do it yourselves, I think ease of use is really an important factor to consider here.

So, what do I mean when I talk about usability? And I know Jen Romano-Bergstrom is in the room and I’m sure she knows much more about this than I do, but these are, kind of, the principles that – the notions of usability that I was operating under when we were having the users come in and test these websites.

So, the first thing, I think, is pretty obvious and we want to provide relevant and easy accessible information to the users. And I think it’s important that those things go together and that they’re both relevant and easy to access. I think, yeah, so, I’ll leave that at that.
We want to enable learnable routines through usability and so, this refers to having similar tasks on these websites follow similar routines so that you don’t have – so you can apply learning from the browsing or searching that you did for variables to the browsing or searching that you do when you go to the documentation and trying to identify, you know, things in code books and things like that. You want those similar browsing and searching tasks to follow very similar routines.

You want to design efficient paths and I think this refers to, kind of, minimizing the amount of – the number of clicks, the number of search terms that people have to use before they find what they want. You really just want to minimize the distance between people when they arrive at the site and when they get to the data or documentation that they came looking for.

We want to create memorable patterns and this is, I think, referring to having elements not moving around on web pages, like, disappearing and coming back and I’ll show some examples of this later when I walk through some of the principles of usability.

And I think we want to minimize user errors. We want to make it really hard for users to make mistakes when they come to these websites. We don’t want people to get 10 minutes into a task and realize that they should have caught a mistake that they made 10 minutes ago. We want it to be very clear. We want these to be intuitive websites. And, ultimately, I think what this comes down to is we want users to come to the websites and we want them to leave these websites having had a satisfying experience. We want them to come. We want to find what they want and need and then, we want them to be able to make use of that information.

Alright, so what do survey website users want? I think it’s, you know, it’s pretty clear that they want data and I think it’s pretty clear that the big three know that they want data. These are the headers that are on the home pages of all three websites. So, you can see that PSID has a data tab; the NES has a data center tab and a tables and graphs tab, so you can go look at some cross tabs and things like that that have already been created; and then, the GSS has a data download tab, a browse GSS variables tab and a data analysis tab. So, you know, it’s pretty clear that survey users want data and the surveys are acknowledging that and trying to provide that. Now, how well they do that, you know, you just saw Steve Ruggle’s presentation and I think he gave a pretty good representation of the current state of a lot of that.
The next thing they want is documentation and we all know from Steve’s presentation how important this is. So, PSID has a documentation tab also on their home page. NES has made what I think is, maybe, a debatable decision to bury their documentation in the data center. I think once you get into the data center, it’s implemented reasonably well, but it’s not immediately clear when you get to the home page where you need to go. I mean, it’s entirely reasonable to think maybe you need to go to the help center or the library to find documentation. But, really, it’s, pretty much, all here in the data center. And the GSS also has a documentation tab.

The third thing I think that website users want is to come and find examples of ways that these data are used in the literature and I think the websites are all pretty good about patting themselves on the back in terms of publications that have come out of their projects and everybody provides those pretty uniformly.

So, some principals of usability that came out of these studies that we did where we had people come in, are listed here and I’m going to walk through each of these with some examples from the websites.

So, first, standardizing test sequences and this is one of the things I mentioned just a few minutes ago. We want similar tasks to be structured very similarly. So, here’s an example from the PSID. This is in their data center looking at the cross year index. Here, you can see their variables are here and then, here across the top, is the listing of years. And then, this grid down here corresponds, you know, so you can select which year and which variable and you can select that and then you can add it to your cart, right? So, you know, that’s one routine. What we want is, at least if this is going to be routine, that PSID chooses, we want it to be the same across the different aspect of the website. Well, that’s not really the case.

So, what you can see here is now, I’ve gone to their variable search tool and the year is here where you can select which year you want and the data type, section of the code book. And then, when you’ve done your variable search, so here, I’ve searched for income, the results come here. Then, here on the right, is where you would select where – which variables you want and then, you can add them to your cart. So, it’s clearly a very, very different task structure. The routine is very different. So, you have to – if you’re new to the website, first, you learn how to do it in the data
center and the cross year index and then you come to the variable search and you have to learn an entirely new routine. This is just really not an efficient way to use your user’s time.

So, the next thing is reducing user workload. I mean, I know it’s kind of small here in this box. So, this was – I was at the GSS website and this was the result of five clicks with no discernable progress. Five clicks of unfolding, single option radio buttons. So, there’s just absolutely no reason that I should have to click five times to finally get to a point where I have to make a decision about what I want. So, we really want to try to eliminate these just, kind of, blatant wastes of time whenever we can.

We want to design for working memory limitations. So, Steve Ruggles wasn’t terribly optimistic about the NESSTAR system and after having played around with it a little bit on the GSS website. I’m not really either and neither were any of the people that we had do usability testing on the GSS website. And I think part of that is, you know, they have a 51 page guide just to how to get started with using this tool. And this isn’t really a tool that you can use elsewhere. It’s really specific to the GSS. You have to do it on their website. It’s just really, kind of, clunky and it’s a lot to learn and try to remember in order to, like, get a cross tab for something. It’s just too much to ask of people.

So, I think you also want to display directly usable information to people. You don’t want them to have to make conversions or summarize things or go digging for things that are very, very important and that should be presented upfront. So, here, I think the NES does, actually, a pretty good job with this. This is the study page for the panel study. One thing I will point out before I get to this is the header across the top is gone now. So, in order to navigate back, if you wanted to go somewhere other than the data center, you know, they’ve removed that element from the webpage. And so, that’s, you know, one of the principles that I talked about earlier. You don’t want that to happen. You don’t want things to disappear, especially when they’re really crucial to being able to navigate through the site.

Anyway, what they do well is provide a really nice summary of some of the really important features of the data set so, you know, that it’s the number of completions and the mode and information about the way it’s making sure that people know to use the weights and where to get information about how to use the weights, things that, you know, people may not know initially, but that are really important for people to know. And this is directly usable
information that people should have before they start doing data analysis.

So, the next thing is guiding users through documentation and this was one of the biggest issues that we had with the usability testing that we did. So, this is the first page. This is the very top of the first page of the ANES cumulative data file variable codebook. This is the codebook. I don’t know what most of you see. Maybe you can’t even read the text, but there’s not really much meaningful, especially to a new user, that’s going on here. I don’t know what is in this document and I don’t know where anything is in this document. There’s just really – it doesn’t orient me at all to what’s happening. And I think this is really unfortunate because you could open this and it takes a while of scrolling before you start figuring out, kind of, what the pattern is and what’s going on and where to locate things.

I think a better example is from the GSS codebook. So, they have a landing page. When you click on the codebook information, you can open the entire GSS codebook and get it in PDF from, which, again, Steve Ruggles pointed out the problems with having everything locked in PDFs and not linked. So, I’m not going to belabor that. But, you can choose which section of the codebook you want to go to. And so, you know, it kind of breaks it down. And then, they provide this really nice option to go to technical appendices. So, if you’re looking for, you know, study designs or variables and things like that, you can find that and it’s summarized here and you can just directly go to that portion of the codebook, which is great because it’s a 3,500 page codebook.

[Laughter]

So, when you do go to the codebook, they further layout the information for you by actually providing a table of contents, which isn’t really a new innovation, but it’s, kind of, an important aspect of orienting your users to the data that you’re hoping that they’ll find. And so, I think that’s a couple of things that GSS has done somewhat right, at least if they’re going to take this approach of having everything locked in PDFs.

So, a few more principles of usability that I’m going to walk through. And we’ll start by looking at, now, the ICPSR website. So, fluid design. This is – it’s, kind of, a minor point, but as somebody who uses these websites, it’s, kind of, important. So, I don’t know if you noticed, but in some of the previous slides, there are bars on the sides, just empty space, and that’s because the
websites were not using a fluid design in maximizing my screen size and resolution. And so, they’ve, kind of, standardized the way that the website will be displayed on your monitor. And so, it doesn’t actually fill the whole screen. ICPSR is using a fluid design and so, this actually just automatically resizes to take into account the resolution of your screen and your screen size, which is really nice when you’re – especially if you’re looking at, like, massive amounts of data or huge documents. You don’t want to be limited to, like, half of your screen.

So, the other thing that I wanted to point out about the ICPSR that’s, kind of, nice is, so, you’ll notice the elements did move and that was one of the things that I said earlier shouldn’t happen. But, I think they’ve compensated for that by the fact that they’ve actually – they’ve color coded everything really nice in terms of their tabs. And so, you can see I’ve clicked on the find and analyze data tab and you know that because you can see in the subheader, they’ve also color coded that. So, no matter where you are in the website, it’s pretty easy to tell where you actually are. So, if you start browsing and wonder which part of the website you’re in, you know, you can just, kind of, make that correspondence and it’s pretty easy to keep track of where you are.

I think the other good thing that ICPSR has done is I think they’ve invested a bit in developing their search capabilities. And I think this is important because a lot of the survey websites right now, are designed to be browsed and that just involves a lot of clicking. And I think, as we moved forward with the internet, we’ve definitely moved, I think, towards much more of search based, kind of, information retrieval model. And so, I think ICPSR has recognized this and they’ve actually built a pretty nice search tool. You can actually enter a literal research question, so they give the example here, “Do children of Asian immigrants speak English in the home more often than children of Latino immigrants?” So, you can actually type that into the search bar and it will come up with the relevant data sets and variables and things like that that you’re looking for. And I think that’s really cool, especially if you’re a new user and you don’t know the lingo, you don’t know the keywords, maybe you don’t know exactly what you’re looking for. You can just type your research question in and it’ll find it for you and I think that this is the kind of thing that would be really helpful for the other – the big three to do.

You know, ICPSR definitely still makes browsing available and you can actually browse by a number of different criteria, which is great. But, I think, you know, improving search algorithms and
search models is definitely going to be the way that we should be moving with some of these web design projects.

So, at this point, I think I’ve got a couple of minutes. I’m going to take a brief video intermission. Wait a minute. I think this is going to work. Okay. I guess I have to do this. So, I actually had four videos. In the interest of time, I’m going to show one of them because – so, what we did was we – as part of the usability study, we had some users do the same task on the three different websites. So, in this case, we had them – we wanted them to locate the variable for income. It’s an entirely reasonable task. It’s probably one of the things that people do most often as it’s a pretty key indicator that people use in a lot of their research and models.

**Question:** [Inaudible].

**David Vannette:** Yes, well, and it’s also in the title of the survey and there’s actually, kind of, an interesting thing that happened. So, I’m going to actually play this at higher speed.

So, you can see the user went to the data page and now, they’re trying to figure out where they want to look for the – to find the variables because it’s entirely reasonable to try to go to the data page, but they had no luck. So, then, they went to the sidebar here, where they found the variables search term, which is actually where you, kind of, have to go to find it. And the most straightforward way is to use this variable search. So, I’m not exactly sure why they haven’t included in that in the data tab, but they’ve made that decision not to.

So, now, they’ve done the search. The website’s really slow when it’s looking through all of the years. And so, they actually cancelled the search and just looked for 2009 because we asked them to find just the 2008 or 2009 variable code for household income.

So, now, they’ve done that search and here, the results are loading and now, they’ll start doing the search. Now, the problem initially here, we’ve got 753 records found for income in the 2009 PSID and absolutely no indication of where – I’m actually going to speed this up so that you don’t have to watch scrolling through – that might be –

**[Laughter]**
– scrolling through 753 variables. So, they’ve given no indication of which are the most important variables and I think that that is a key oversight and it’s one that’s true across the different websites. I think it would be extremely helpful if these websites could give some sort of indication of which of these variables is downloaded most often, published most often, just any kind of indication of what the most important variables are because scrolling through 753 variables is really not efficient and it’s not fun and if I don’t know exactly what I’m looking for, I don’t know exactly, you know, what the variable label is here; and these descriptions are hardly complete. So, now, the user has done a search and this is actually a browser function, not a website function. And I actually read recently that something, like, only 20 percent of internet users actually know that trick, that you can hit Control F and search the entire web page. So, the average user is not going to know how to do that. So, they finally found the variable and got a bunch of other stuff they got thrown in there, kind of, oddly. And so, yeah. It’s just not exactly an efficient process.

Alright. Let’s get back to this. So, the other two websites, the ANES and the GSS, had very, very similar experiences, with just a lot of difficulty finding and identifying the actual variables that you’re looking for. You know, this should be one of the simplest tasks. And, you know, I was playing that video at four times speed, so it seemed – it didn’t seem terribly unreasonable.

[Laughter]

But, when you have somebody whose actually been doing this usability testing for a while, trying to do what should be an extremely simple task and it still takes eight minutes for them to locate a single variable, it’s just not terribly reasonable.

So, in conclusion, you know, I think we do some things well. We do provide a lot of data and I think Steve Ruggles made a great point that, you know, the web has really transformed the way that data can be disseminated and the same with documentation, you know? And I put much of the time here. You know, we provide documentation as much as we can and I think a lot of the newer surveys are, the newer waves, are doing a much better job of putting everything online right away, but there’s still gaps in the documentation. I know, for example, the NES, there are some show cards from, like, the ’60’s – ’50’s and ’60’s that just aren’t available. They’ve been lost forever. And so, if you go – happen
to navigate to that page on the website, it says, “Do you have one of these because we need it”, right?

[Laughter]

Like, just in case you happened to have found one lying around your house from when you’re cleaning, you know?

[Laughter]

You know, and we provide publications pretty well. The NES actually has a nice thing that rotates through on a sidebar showing a bunch of their recent publications and I think that’s kind of cool.

So, we definitely have room for improvement. So, structurally, we can make sure that tasks are structured in similar ways across websites. They shouldn’t be terribly difficult to do. If you’ve implemented it once, you should be able to implement again. You don’t need to reinvent the wheel twice on the same webpage. In terms of design, you know, we can design much more efficient and intuitive web pages and I think there’s no reason for us to not do that. And then, the code books, you know, unlinked PDFs, like Steve Ruggles said, it’s just – you know, we’re just locking away these resources that are absolutely crucial and making it really hard for people to find the documentation that they need in order to make proper use of the data and the variables that we’ve made available to them. And data sets, like Steve was saying, also, just like the structure of the data sets and the fact that they aren’t really compatible with each other over time. It’s really hard to link them. You know, we’re discussing a lot of data linkage issues here and linking with government records when we have a hard time linking our own surveys together. So, I think that that’s also an area where we could use some improvement.

And so, there’s some easy steps that can be taken, I think. And the first one is learn from users of the website, get feedback from them. The second one is learn from each other and look at ICPSR. See if you can emulate those kinds of functionalities that they’ve put on. I don’t know why PSID hasn’t done this already. They’re hosted at U Mich, just like ICPSR is. It seems like they probably have the same web development resources available. And, if they’re not going to do this, I think this is another point that Steve Ruggles was making, if they’re not going to do this, they should push more of their data to ICPSR in order to make it more usable and more available to a broader range of users.
Conduct usability testing when making changes. You know, I think, I know Matt was saying they’re hoping to update the ANES website soon. You know, I think it’s really important to conduct testing as you go so that you know what works and what doesn’t work from a user’s perspective. Don’t just rely on your web design people or the intuitions of the researchers because there’s a pretty good chance that they’re wrong, which is what we can see from the current state in many respects of the big three websites.

And there’s this great resource, so after I had done a lot of the usability testing, having people come in and watching a lot of these screen capture videos, you know, I had a lot of things written down and ideas and observations that I made. And then, I happened to cross usability.gov and they have fantastic resources for how to make useable websites and they provided a lot of the terminology that I was able to, like, put to the, kind of, the descriptions and observations that I had made. So, extremely, extremely useful.

So, the last thing, I wanted to come back to one more principle of usability and this is, you know, we have a lot of expertise in doing surveys. We should actually apply some of this expertise in designing our websites and get feedback from our users and we should get feedback often. We can design good surveys. We can get the information that we need from our users and there’s no reason not to do that. And we can also make use of things like pre-testing with users and making use of Para data. I don’t know if any of the websites are making use of the Para data that they have to find out where users are getting hung up. Where are the pages that people are dropping off and just leaving the website? How long is it taking before people – from when someone enters the website to when they get to what they’re looking for? And are there ways that we can design more efficient paths to help users get to what they want in a much faster way?

So, I’m going to – and I had a video of this and I’m going to end with an anecdote. So, a couple of weeks ago, I was in class with Paul Sniderman, who’s one of my professors. And he came up to me after class and asked if I would please go to the ANES website and find the last year that the economic individual questions were asked and send him the question wordings because he wanted to use them in a study that he’s fielding. So, I said, “Sure, absolutely. I can do this.” Fortunately, I actually had an ANES employee coming in to – a third year graduate student in political science who worked for the ANES for three years – was coming in to do some usability testing for me that afternoon. I was going to have her do the GSS and PSID, just because I figured she knows the
ANES really well and see how well that skill set will transfer to GSS or PSID. But, since I had to find these variables and I had to find them in a hurry, I thought, “Hey, great, I’ll kill two birds with one stone. I’ll see how efficiently and expert can actually use this website and I’ll do a screen capture of that.”

So, you know, to make a long story short, we got to the end of our half hour period and she hadn’t found the variables. She was just completely unable to find that and she knew the exact concept that she was looking for, that it was economic individualism and that they had been asked on the ANES. And using all of her knowledge and her ability, she knew the little search trick, so it’s not like she was scrolling forever and ever, wasting all of her time. She was completely unable to find these variables. So, after she left, you know, all was not lost. I was able to go to Google and actually find the question wording using Google, looking for economic individualism and ANES. But, the fact that a user, an extremely experienced user who actually works for ANES, can’t find a simple battery of questions that appeared in multiple years on the ANES, I think, says a lot about the current states of usability of these websites, in general. So, I think there’s a lot of room for improvement.

Andy?

Andy Peytchev: I have two comments for the past three presentations, since we’ve skipped some of the discussion. I think one part that I picked up on your talk was actually not about the usability, but about enforcing things that you think the user should be getting. So, there was a talk by Brady West earlier this year evaluating, it was kind of meta analysis with sampling on how often do – well, published results and articles misusing weights. Now, keeping in mind that he actually used whether they report using weights, not actually whether they used weights. It was an extremely high percentage. So, the nice thing about these websites is that you can, pretty much, force people, before they download the data, the amount of information that you think is critical for them to be able to say and know not to abuse the data and to analyze it properly. And so, I think that is, kind of, one potential avenue for a website that you mentioned that I think is, probably, one of the very, very important ones.

David Vannette: Yeah. Absolutely. I completely agree and I think this is one case where the ANES does a pretty does a pretty good job. You know, they have methodology documents on the study webpage
instructing users how to use weights and things like that and I think that that’s absolutely fantastic and needs to happen a lot more.

**Andy Peytchev:** My other comment was for – and I’ll be very brief – is more about – so, Bob Belli’s talk did something really interesting, proposing that the future should be using these kind of adaptive models and that, on the fly, you could be designing the instruments, which is a fabulous idea and much of that work can be done using existing data. You can imagine that you have two versions of the instrument done completely different with different approaches and you know the model drives you to which is optimal for that respondent, which is a great thing; only that leads to, pretty much, the biggest deficiency in data archiving from the previous presentation and, I think, in terms of the data that we’re missing right now. So, all these big national surveys that have gone through big redesign efforts and big experiments, those experiments don’t make it to the data archive very often. So, they’re examples. And CVS had redesigned pretests prior to ’92. The National Drug Survey had validation studies done. The Consumer Expenditure Survey’s recently played around with split questionnaire design. These are expensive, very expensive and very valuable data that could be used for that kind of evaluation to see whether things would work. Yet, this is exactly the kind of data that’s never archived and we can’t use. So, I think that may be something that would be very valuable in terms of, maybe, that NSF could fund and help because the major surveys would need to take out from their own funding from data collection to do this, so they have no reason to do that.

**David Vannette:** Yeah, I completely agree. That’s a very good point. Josh?

**Question:** Yeah, so, a couple of thoughts. I mean, really love watching it because I’ve had so many troubles, particularly with the census data [inaudible] one in particular, which, kind of, also needs to be added to this list of things that need desperate help. And it’s also in Java, which is even worse. But, one of the things that struck me as a challenge here, and it probably points towards, sort of, a larger systematic way of dealing with these data sets, not just in the context of an individual website, is things like the ANES errata section. And I have, at a couple of points, gotten a good way through a study before realizing, “You know what? I should read through that errata section because I’ve learned by now that, actually, things can be there and there’s one variable that you, one person that you have to drop from the 2006, you know, study for reasons that aren’t even completely clear.” But, these kinds of things need to somehow as we go through this process, especially
if we want to use – move from a browsed base system to a search based system, need to be bundled along with the data set; because if you lose the fact that you’re supposed to drop that one particular case or that a variable was, perhaps, reverse coded or that there is, you know, all of these myriad different things that go wrong in survey data sets, that we all know about because we actually collect this data, but that a lot of the practitioners that use this data later on don’t know about. And stuff like the weights, we’re going to, I think, possibly do ourselves some harm in that process as well. So, I think it needs to be really systemic in terms of how we go about it and, probably, even more broadly based than, sort of, NSF and the big three, to the extent that NSF could be part of a coalition building those kinds of things.

*David Vannette:* Yeah. Yeah. Absolutely.

*Gary Langer:* It’s Gary Langer. Yeah, I would just add that you might want to, in evaluating this, which is really important, and it was an enlightening presentation, to go back to think about the purpose of this because I would suggest that each of the sites you just showed us is not truly a public user website. It’s a parody of a public user website. It’s a parody of a public user website. It might be a professional researcher user website, just for the cognoscenti to use. But, anyone who is a normal human being, not deeply involved in this field, would be completely at sea in a canoe at any of these sites.

*[Laughter]*

*David Vannette:* Yeah.

*Gary Langer:* And that is not necessary. I would just say, if you do a broader look, there are two examples I would suggest that are really very usable and user friendly. One is CHIS, the California Health Information Survey. It’s a terrific website that any schmuck can go in and run cross tabs on. It’s so straightforward, it really should be looked at, I think, and celebrated. Another, although I, full disclosure among the board, is The Roper Center for Public Opinion Research, which also has a much more useable, much more accessible. You can download full data sets if you want to get into the SPSS, but you can also do on the fly cross tabs in a very accessible, clear and straightforward way that none of these even begins to approach.

*David Vannette:* Right. Yeah. Absolutely. I’m a big fan of The Roper Center website. Bob?
Gary Langer: I’m going to defend the PSID as best I can here.

[Laughter]

Gary Langer: Good luck.

[Laughter]

Gary Langer: Well, now you know what kind of fun I had in terms of trying to get that validation data.

[Laughter]

But, the main point that I want to draw and Dr. Ruggles expressed this earlier, is that these data sets are extremely complicated. The PSID data set is both longitudinal and multi-level. You have individuals within households across years with sample following rules in terms of it being an intergenerational study. And trying to-I can’t, you know, if someone’s going to be a novice who doesn’t understand issues associated with what longitudinal means or what multi-level means, they’re going to be lost.

Now, there may be some ways in which that website can provide a tutorial, you know on those issues.

[Laughter]

But, I’m not certain that that it’s really up to the web designers to do that for potential researchers. And I guess that’s – I mean, I’m sure that there are other things that they can do. But, in terms of the historical nature of that, of the data collection, the coding of household members has changed over time. So, there are different codes assigned to heads and wives. I mean, was that all done in ’68? They talk about a head or a wife and now, a female can be a head in some years. But, also, then would be a wife whenever a male partner comes in, regardless of whether or not that male partner is fallible in terms of the sample design. So, it’s a very, very complicated survey. And I don’t know exactly how you are able to convey such very complex concepts in a very easy and usable way.

David Vannette: Yeah. No, I think that’s a very valid point. You know, these are all very complicated data sets and I think, probably, more the PSID than the others. But, I still think that there are some basic principles that they could apply, like the task similarity, right? You should be able to learn how to do a task in one part of the
website. And then, in terms of the design decisions, be able to apply that same learning and that same task structure in a different part of the website, especially, if you’re looking at the same exact variables... I just – I have a hard time excusing that. So, I think that there are – there’s room for improvement that’s not going to, you know, degrade the complexity or anything about the data itself. I think it’s still possible to make it more usable and I think it’s important to get feedback from experienced users for what, you know, what do they like and what do they dislike about actually interacting with these web platforms? Yeah?

**Matt DeBell:**

This is Matt DeBell with NES in Stanford. I just want to say I’m so glad that you have done this study of the sites. I’ve had, for quite some time, an item on my to do list, which is hire some undergrads to do some usability testing of the website. And so, now, you’ve done part of that, so I can maybe halfway cross that off.

**[Laughter]**

And we can start out on the things that you found and then, you know, do it again. It should be an iterative process, as I think you indicated.

**David Vannette:** Absolutely.

**Matt DeBell:** And, you know, I agree with all the points that you’ve made. I think that the difficulty of finding variables is particularly problematic and also, fortunately, relatively easy to fix by creating a tool that will search all of the code books. Our code books are, for the most part, available in text format, which makes them readily searchable, so we don’t have, you know – we have some documentation that’s scanned image files and PDFs, which are impossible to search by any tools that are readily accessible to most people, except the brain. But, we have a tool that’s in beta testing right now that will search the whole thing. And so, that task that the grad student couldn’t complete in 30 minutes, really, you should be able to finish in about 2. And once we get that tool posted online, then the world will be better.

**[Laughter]**

So, anyway, thank you.

**David Vannette:** Yeah, absolutely.
I want you [Crosstalk] here. I want to show you something. Oh, interesting. My internet expired. I think we’re –

Musical accompaniment. I love that. Alright. So, I want to show you something. So, if you want to find a variable on the American National Election Studies website, if you say, “I want to find education. How was education measured in 2008?”, you might think you could go to the data center. But, it turns out there was this lovely tool right here that you can only find if you know to go the search box and type “variables”. And then, you go here. Whoops, sorry. Wrong tool. Whoops. My mistake. I typed the wrong word. It should – actually, “questions” is what I should have typed. “Questions”. And then, you get to questions asked in the ANES. Now, if you look up here at the address, it’s under resources, questions, questions. I can’t find any tab that says “resources”, so there’s no actual way to navigate to this. The only way I can figure out to find it is to type that word in. And what was created here is an either these are PDFs, I think, maybe? Let’s find out. Well, it doesn’t matter. It’s not going to work anyway. These are supposed to be PDFs and these are supposed to be searchable text versions of all the questions asked in all these decades. And so, if you open one of the text files and you search for education, you can find the – or school, whatever word you want to look for – you can look for the question.

So, this was really cool. Guess what? None of the links work now. So, (A), you can’t find it; (B) it doesn’t work. So, I’m hoping it will get fixed soon because it was a really cool idea.

Charles?

Charles DiSogra: Yeah, I just want to go on record because I want to thank the compliment for CHIS, since I was one of the designers of the CHIS website. And I want to say that when we designed that, we designed it with the purpose of being able to have – look at surveys that were going to be taken every other year over time so you can get that trend data. And the other thing is that you can find variables by topic area. And then, when you click on them, it shows you the question that was actually asked and then, you also see the source of that particular question, where it came from.

But, when it was designed, we designed it, actually, for export in the sense that we – with a SAS engine underneath it. We designed
it so that even though it’s pointed at each of the CHIS surveys, we felt – well, it made perfectly good sense – we could point it at any other survey, any other data set and data sets that operate over time, it would be designed to do that. And, in the last year, unfortunately, we had a small stop on this action, but UCLA, who owns this, has decided to, basically, market it, make it available and to other places, particularly states and state health surveys it was thinking. But, I’m thinking bigger than that. I mean, you know, in the sense that a lot of work’s been put there and I want to just go on record because that’s what I wanted to ask Steven about, that he might want to consider taking a look at that and see how it operates. I mean, that work’s already been done.

David Vannette: That’s great. Norman, did you have your hand up earlier?

Norman Bradburn: Norman Brandburn. I actually had two questions. One, it just reveals my ignorance and so forth. Steven Ruggles mentioned NESSTAR and SDA, I think, and saying both of them were, essentially, going out of business or not being maintained and so forth. So, I guess, two things is how dependent are these things on systems like that? And, B, if they’re going out of business or not very good, what’s going to replace them? That’s, kind of, one question.

The other question [Inaudible] same because they may be related, and this is probably to the ANES people, but because the other people aren’t here. Who’s responsible for putting these up? My guess is the PIs aren’t or only in a very general sense. So, I – my intuition about how we got into a situation like this is that it’s like the disparity I mentioned before between the researchers and the operations people. It’s a different set of people who don’t understand – who, either, they’re – what they’re trying to maximize is something different from what we think users/researchers would want coming to it. Common failure, in my view of website, because the people who tend to design websites have other – either they don’t understand what the users really want and this is – I’m over generalizing. But, you know, it’s – I have frustration with lots of websites, not just these.

[Laughter]

So, it seems there’s a structural, there’s a, kind of, human capital structural problem and it may be that, again, it, sort of, seems to me it’s inefficient and almost a waste of money, to expect each of these different things to do their own. I mean, that seems to be what ICPSR or places like that are Roper Center and so forth, are
for. I mean, they have – should have – expertise in how to do these things. They should know usability kind of things and so on and so forth. And it seems odd. In fact, I, until he mentioned it, I thought that the GSS was maintained, essentially, by The Roper Center and IPCSR. And I don’t know who does it now. Tom left, so I couldn’t ask him, but I’ll ask him tomorrow or Monday.

[Laughter]

*Jon Krosnick:* Well, I can tell you that when we were running ANES, we thought we were responsible for putting the data up on our website and maintaining our website ourselves. And we have zero expertise in web design and we had no staff members who had any expertise in web design and we had no budget for web design. And so, we just tried to do it, you know, tried to, sort of, live with – and we did redesign the website during our time. But, you know, I don’t know very well. Matt could comment on what’s happening now in terms of these issues.

*Norman Bradburn:* The situation now is very similar to when Jon and Skip were PIs. I would just add in response to your comment that – Norman’s comment that, you know, it seems like it makes sense for an organization like ICPSR to be the repository for all this sort of thing. On one level, I think that makes perfect sense, but I think there’s a – maybe there’s – I think there’s a distinction to be made between types of users and their objectives. I think there’s a general – a user that would have a general research project in mind for which they might go to ICPSR and search for the appropriate data and hope to find it there. And then, there’s another kind of user that knows that they want something from a particular study. And so, for instance, if I want to look something up in the most recent GSS because I know that they’ve been asking a question for years or something like that, I don’t want to go to ICPSR’s website and have to navigate through other things. I want to go directly to, you know, the greatest authority on the GSS, which is the GSS website, which is managed by the people who run the GSS because I expect that their, sort of, local knowledge and expertise on the GSS will give me the best experience.

Now, maybe that’s not a valid assumption because they don’t think enough about the web, just like it might not be valid or central to the ANES because the web is not our priority. Our real priority is designing the study, carrying out the study, producing the best data we can. But, I think there’s a tension between those two kinds of user experiences and objectives.
Question: Could I –

David Vannette: Yeah, go ahead.

Question: The – these – I think what you mentioned about expertise and website, I think we researchers, [inaudible] and so forth, tend to underestimate what the expertise of web designs and we think it’s something easy to do. It’s like people who don’t know how to do surveys easy. You know, they can put together a questionnaire in half an hour and get on the phone and use Survey Monkey and that’s all you need to do. So, I think there’s a real disconnect between understanding all of the things that go involved to a good website.

One idea, I thought this, maybe, you know – you know, I know it’s very expensive to – I mean, and we don’t tend to budget for it. And so, it’s – this is the result. One of the things that occurred to me, although this is, again, resources, for these complicated surveys, there are really two types of users. I mean, there are people who want just to know, you know, what the trend is in something or other, but in particularly, things like GSS, and have rather simple questions. And then, there are people who want to do deep analysis of varying sorts. Would it be possible to have, sort of, two versions? I mean, a kind of reduced form version for people who just want to look up some things in a way and then, people who really want to do serious analysis? People who do serious analysis, right now, you know, they don’t like it, but they slog through and get stuff out and so forth. But, you know, I think it’s – I mean, but there are other places that have put more – is Randy still – yeah, Randy has put a lot of resources into getting the NOSY in a usable, you know, it’s a complicated survey. It’s the ’78 one and so – and the ’94 one is even – or ’97, whatever it is, is even more complicated in a way. And, you know, they put a lot of resources into making it usable or sorts.

But, again, once, you know, what’s the tradeoff between centrality where you’ve got the expertise to do these sort of these things and distribute it, having different things. And I don’t know what the tradeoff is.

David Vannette: Yeah. So, I think, you know, my first impression to your initial comment about, you know, having kind of a stratified website, like, for the general user who’s just, kind of, coming to look for a simple trend versus the more advanced user who’s actually going to make use of the complex designs and actually do real analyses of these data. I mean, my sense is that that’s, kind of, what the
online tools were intended for initially because I can’t imagine anybody actually going there and trying to do serious analyses on using these web tools. So, my impression is that they put these web tools up so that the casual user could come by and make, you know, a quick cross tab of a couple of variables. But, then, they’ve, obviously, let them, kind of, fall into disrepair and have stopped updating them and keeping them on the cutting edge. And so, they’re, kind of, basically useless at this point.

Okay. So, one more comment from Jen.

Jennifer Romano-Bergstrom:
I just wanted to add to that. We used to hear that when I was at the Census Bureau. We heard that a lot too because when we did user testing, we’d bring in expert users, bring in novice users and, inevitably, the novice users had more issues, but expert users also had issues. So, I was wondering about your participants, actually, in that study, if they were just – who were your participants?

[Laughter]

David Vannette: So, I would say they were, for the most part, they’d probably be considered expert users.

Question: Okay.

David Vannette: They were all graduate students or post-docs in the social sciences at Stanford. And so, they all had some level of familiarity with accessing secondary data online.

Jennifer Romano-Bergstrom:
Okay. And they had lots of issues?

David Vannette: So, these weren’t just, you know, random people off the street.

Jennifer Romano-Bergstrom: Right.

David Vannette: I would classify these as expert users.

Jennifer Romano-Bergstrom: Yeah. See, so even expert users have issues, which is what we used to find with the Census Bureau data as well.

David Vannette: Right.
Jennifer Romano-Bergstrom:
The other issue is in terms of having two different sites because we used to get this recommendation as well. The problem with that is for anybody who’s doing the upkeep, then you have to remember to update two sites every single time you’re adding a piece of information. So, the idea should be that it’s usable enough for the novice user. If it’s usable for the novice user, then the expert users will have an easier time with it as well. And that doesn’t mean we’re dumbing it down. Because people always say that. You want us to dumb it down, but it’s not dumbing it down. It’s just making it usable.

David Vannette: Yeah. I completely agree.

Jennifer Romano-Bergstrom: Yeah.

Jon Krosnick: Okay. So, before we go any further, I just thought we should take the opportunity to thank Dave, not only for his presentation, but for his amazing work throughout both of these conferences, getting us all organized and fed and housed.

[Applause]

So, this brings us to the end. So, this – thank you so much for sticking in here ‘til the end and thank you for your comments and your engagement. This has really been – I think everybody thinks it far exceeded our hopes of how good it could be under the circumstances. The presentations were great. Discussion was great. The worst part for me was cutting off the discussion at the end of every time period and, as you saw, I didn’t do so well with that today and let the schedule get a little out of control. But, we’re – really, the comments were terrific and I hope that – and I will tell everybody this by email who’s not here, if you have additional thoughts on these issues, please do email us because Dave’s next task is going to be to write reports that will be the basis of what our subcommittee gives to the advisory committee, which we hope will have impact here at NSF.

And so, your continued input on these issues will be helpful and we will ask you to look at the text that we write that has to do with your expertise and that may, in some cases, be a lot of text. So, we hope you’ll be generous enough to give us comments and suggestions to improve it. Allen has something to say.

Question: [Inaudible]?
Jon Krosnick: Dave could answer that question. That’s on his to do list. He’s creating a website of the slides and other things.

David Vannette: Yeah. So, we are creating a website. This will be run through IRiSS at Stanford, the Institute for Research and the Social Sciences.

Jon Krosnick: Research and the Social Sciences.

David Vannette: Yeah. So, I’ve been working with them already on getting this put together, so we’re hoping to have a live website with the slides and then, you know, kind of profiles of each of the speakers and, eventually, maybe even the transcripts and audio available to you.

Jon Krosnick: And everybody’s bank account numbers, right?

David Vannette: Yeah, right. Of course, yeah.

Question: [Crosstalk].

Question: Easily searchable?

David Vannette: Very easily linked and searchable. Yeah.

Question: [Inaudible]?

Jon Krosnick: We’re going to have a lot of usability testing before this website goes up, actually.

[Laughter]

So, it could be 2015 before it’s functional, but that’s alright. When it’s up, it’ll be really good.

So, thank you so much and I hope, you know, you’ve gotten something out of it. If you’ve enjoyed it and you want to let Myron Gutmann know that it was a useful experience for you, I know he would appreciate it because he was the person who made this happen. And I wish you all the best in travels home and look forward to seeing you again soon.

[Applause]