

We are going to get started shortly. Right now, people are still joining. Just as a reminder, this event is being recorded. Please take a look at the chat, where you will find a link to live captioning and please also let us know what agency organization you are from, in the chat. We will get started in a minute.

All right, let me go ahead and get started. My name is Jean Fox and I am the cochair of the user experience community of practice and I would like to welcome everyone to our 2020 virtual government UX Methods. We have a really good program this year and I am excited to hear from our colleagues in government from across the country at all levels. I want to thank everyone who made this possible, including my cochair, Wendy Stengel, who was there every step of the way. I want to thank the staff of GSA who provided outstanding logistical support. I want to thank everyone who submitted proposals, we got a lot of great ones, and I want to thank the review team who helped plan a fantastic program today. That was a difficult process. Thanks to all the presenters who took time to work on their presentations and share their experience with us, and while the chairs who are facilitating sessions this year. And thank you to all of our people attending the session today. This is why we are doing it and I am excited that we got such great comments this year. We've got a great program, with two sessions today, Tuesday, Wednesday, and Thursday of this week. Don't forget, you have to register for each session separately, because we have separate links for each one.

This has been organized by Digital.gov. The mission of Digital.gov is to transform how government builds, learns, and measures digital services in the 21st century. They do that by providing people in the federal government with tools, practices, and policy guidance they need to deliver effective digital services. As part of their mission, Digital.gov supports the community of practice. Our community provides resources to help those doing user experience work at all levels of the government and is open to anyone with interest in UX and a government email address. Take a look in the chat box, because there will be the information with a link to our websites you can find out more about joining our group. You can find out information about other communities, as well.

Finally, a couple of logistics before we get started. We are planning to have videos of the presentation available in a couple of weeks and the slides for the presentations, as well, probably along the same timeline, but we will keep you posted. We will send information to the community. At the end of the event you will see a link to a survey. Please take time to fill it out, it will help us give you even better events in the future. Keep an eye on the chat, because there will be other informational links in the chat, as well. All of the attendees will be muted, so please address all your questions to the chat, whether questions for the presenters or technical issues. We will have people monitoring that for both kind of comments. As a final reminder, this event is being recorded. So, thank you very much to everyone who made this possible. I will turn it over to Susanne, who will be the chair for this presentation.

Our first presenter is John Pull. John's career spans 28 years, but only to two institutions. Harvard from 1992 to 1995 and the Library of Congress from 1995 until the present. He has been a cataloger, web director, communications director, and he is currently a UX specialist with the Library of Congress design directorate. From 1998 to 2013, in his spare time, he ran a successful

digital art consultancy in London. He has lectured extensively in North America and Europe and has worked with national libraries from six continents. John, we welcome your presentation.

Thanks, very much. Let's share the screen here. My presentation today is called becoming a UX Houdini. In the world of magic, there is perhaps no more famous performer than Harry Houdini. In September, 1907, the San Francisco call proclaimed Houdini, the Leonardo da Vinci of the vaudeville circuit who can break jail with impunity is a real hero. Decades have passed, yet his name is still synonymous with the miraculous. He freed himself from ropes, shackles and chains. No lock could hold him. Squeeze him into a milk can, hang him upside down in a strait jacket, in a tank of water, no less, no ai, fully visible to his audience. Nowhere, fully visible to his audience, and every time he escaped. Most celebrities of a century ago have been forgotten, but today the name still evokes praise that is undeniably breathless.

I work at the Library of Congress in Washington, D.C., and the library has a wealth of Houdini materials including photos, prints, posters, recordings and manuscripts. Houdini is of interest not just for the feats he performed on stages around the world, but how he got there. As a boy he emigrated from Hungary to the United States. In 1886, years before he took the stage name Houdini, 12-year-old Eric Weiss set out by boxcar hoping to earn enough money to lift his family out of poverty. Perhaps his status as an immigrant contributed, as it so often does, to his spirit as an innovator: to use the outsider perspective to discern truths about a situation that others might miss. How does one capture the attention of an audience? When an action takes place, where does the eye go? What is predictable in a focus? Once those things are understood, how can those things be designed and fine-tuned to maximize the odds of a standing ovation? As UX professionals, this is what we do every day. You have a knack for seeing things that others might miss.

Here in Washington, D.C., the subway system typically has more escalators going up, than down, that's because each time the drains into the station, they unload hundreds of passengers in a matter of seconds, creating denser quantities of people exiting than would ever have entering. It really is a kind of magic, so no wonder someone like Houdini would be a patron saint of UXers. To further attach him to our profession, I will point out something else: once he achieved worldwide fame, he used his visibility and influence to expose sham spiritualists. They were abundant at the time. The scammers and fisheres of their day, holding fake seances and swindling people out of their money. Houdini was insistent that it be none that his seemingly impossible feats were in no way enabled by dark arts or sorcery. He wasn't getting help from intangible forces in the great beyond. Instead, he was employing quantifiable methodologies of understanding, manipulating operations and perceptions in the world, creating magic with a reliability that some might call scientific. Likewise, we UXers work in a world that often relegates art and design to the unscientific space of personal taste.

How delightful, then, for our spokesperson, Harry, to campaign for the tangible, the quantifiable. Call it magic if you like, but beneath the wonder is observable human behavior making it happen. And so, having established Houdini as a role model for our discipline, I want to invoke his spirit, to draw on the lessons of his performances in order to help us address some of the less pleasant aspects of our work. I am talking about problem projects.

The ones where your sixth sense is tingling before you even finish reading the charter and you've instinctively put on your risk mitigation cap well before the first standup. As UX professionals in government, perhaps the fact that you're in the public, rather than the private sector, seems to introduce obstacles every bit as constricting as locks, chains, and straitjackets. Bureaucracy and regulations, legacy systems and increased scrutiny can all raise the stakes to Houdini-like levels, but take heart, government workers are the most dedicated, skilled, and nimble petitioners out there. Harry demonstrated those qualities: the impossible becomes possible, as long as you step up and bring it.

Lesson one, shake hands with danger. Don't be put off by the tough project. Avoid recklessness, of course. Don't punch completely above your weight, but then again, you are in the ring because you know your stuff. And, speaking of knowing your stuff, here is something forgotten about Houdini: he didn't start out as an escape artist, he began with traditional magic, as most magicians do. He was pretty good at it, too. The same 1907 newspaper called him a king of card tricksters, a Lincoln Steffans of prestidigitations comparing him to the pioneering political reformer. But card tricks didn't make him famous, so he meditated on what would. Where does the eye go? How do you capture an audience? Which experience will bring a standing ovation? In his case, the audience wanted danger. He didn't cower from it, nor did he launch himself at it Leroy Jenkins style. He calmly extended his hand and gave a firm shake. Here is an example from real life, from the halls of the Library of Congress.

For the last year and a half, I have been the UX specialist on a little engine that could project called Stacks. The Library of Congress has been in the business since the 1990s, providing free access to anyone with an Internet connection. Stacks, though, is a digital library with a difference. It is a space for rights restricted materials that can only be viewed on machines within the Library of Congress campus, including remote workers on the LOC network. Rights restricted materials include books, newspapers and journals that are submitted to the library in a digital form for copyright. In some cases, there are requirements written in the copyright law that dictate restrictions on use. Those restrictions might say that a single item can have no more than two simultaneous users. Users might be able to print paper copies, but not download them. And as mentioned, make it available only within LOC spaces. I hesitate to talk about Stacks right now, having set the stage for a problem project. In our case, the charter was excellent. The assembled team, though albeit a bit small, was highly skilled and eager to work together. The material was diverse, interesting, and the project had the blessing of top leadership in the library. So, what made it challenging? The timeline. Quite short for a project of its size, especially considering the number of unconventional requirements, including finite simultaneous users and print no download. In the way, it had the trappings of a Houdini spectacle. You will be straitjacketed, chained, underwater, upside down and various heads of state are going to be watching. The challenges of constraint, time, and scrutiny, but we also saw the possibility of a standing ovation, from patrons, as well as colleagues. Guided by our project manager, cautious, but undaunted, we stepped up and shook hands and almost immediately we encountered the second lesson to be taken from Houdini.

Foster the miraculous. In keeping with the greater vision of patron services at the Library of Congress, we applied a user centered approach to the project. From a UX angle we took the steps you would expect. Stakeholder interviews, cognitive walk-through, heuristic evaluation.

Likewise, developers and subject matter experts were moving forward and before long, we realized something surprising. The impossible afforded us unexpected benefits. The truncated timeline made for beautifully, distraction free collaboration. The pickiness of requirements created solidarity around the elegance of a minimum, viable product. It was as though we were bringing an outsider's perspective to it and uncovering examples where others see obstacles. See escape and acclaim, where others see doom. Create and fine-tune experiences that capitalize on the challenges that sent others running for the hills. Foster the miraculous. And what did the audience think when we launched?

Well, libraries are stereotypically quiet places with a lot of fishing, but not at our demos. I can see in my 25 years at LOC I have never seen librarians this excited. It turns out that pretty much everyone has restricted view material that they have always wished could be available via digital delivery. Manuscripts, audio recordings, foreign language newspapers. The initial Stacks release has been followed by two major releases and current development is underway to support audio and video. It is conceivable that one day Stacks could hold original copyright deposits of motion pictures and if so, there may be clamoring amongst scholars attached to the han shot first argument.

Lesson three, very simple. Be Houdini. Use your training and curiosity to see what others don't. Be undaunted. High-stakes can bring high-fives. Cultivate the unexpected wins along the way. And debunk the belief that user experience is a séance pulling decisions out of midair. Abracadabra. And that is all.

Awesome, John, thank you so much. At the beginning of your presentation, someone likened it to listening to a podcast and I think you continued to pull us in with the rest of your time. So, you are getting applause and the comments and we welcome your questions in the comments, as well. I haven't seen a specific question yet, but continued praise. Someone commenting that storytelling is so effective and I would say that even in your presentation, with your message, that telling a story definitely adds to what you have to add. So, the praise is continuing to pour in. Great story, great job, very pleasant voice. Absolutely. Somebody asked the question if it is recorded? Yes, the sessions are being recorded and will be made available as soon as possible.

So, Kimberly is wondering, John, what was the biggest challenge that you face and how did you solve it?

I would say that it was the timeline, and we were inspired by the fact that what we were doing was meeting the universal goal of libraries, and that is to make materials available to patrons. What had happened was the library had been receiving boring digital, mandatory copyright deposits of newspapers and they were accumulating and we had no way of presenting them to anyone. So, it was an important project, but, as with any with any large institution, there are a lot of different projects. A decision was made that a fast timeline could probably really grease the wheels and make something happen. When we started, there was an immediate sort of head scratching about the timeline. It really was between three and six months and we just banded together and it was an amazing team. Everyone worked together, beautifully. When we were about three months into it, we realized, wow, we are actually going to be able to create something important and the ripple effect is going to be profound.

Here is your lightning round question and that is, if you could talk about a specific instance when you are able to shift someone's mind on UX being magic.

Absolutely, my first experience was doing a major redesign for an intranet and it was for a research organization populated by people who are tops in their field, worldwide. So everyone had an opinion, and it was hard to make anything change. By taking a user experience approach and having quantifiable data to back up the all of the decision-making on the redesign, as well as having pulled in, I think we had 40% of the overall staff, contribute to the different exercises. At the end, when we launched and we had the same 30 people who complain about lots of things say “why did this happen? This is all wrong.” We were able, instead, to point to data and say this is why we did it and we can back it up with more data, if you need.

Great, thank you very much. So, thank you, John. As I noted earlier, the sessions are being recorded and will be available as soon as possible.

So, our next presenter is Angela Smithers. She is a user experience designer at Smithsonian's National Museum of African-American History and Culture. Previously she was one of a number of volunteers on the UN sustainable development goals for national reporting software for the U.S., with the team at GSA. Her fields of design range from architecture and interior design, to graphic and web design. Her passion for both people and design has led the field in user experience and accessibility. Thanks, Angela.

Thank you. I will start by sharing my screen. As soon as the host allows the screen sharing, I can choose the option to share the screen.

Can you try again, Angela?

Okay. Let's see. Okay. Are you able to see -- a little too much.

I think I saw a slideshow option there.

Yeah, I went to the advanced options and share a portion and that option has been removed.

Okay. Okay. Okay.

Okay, so this is UX via dot voting a story of a novice approach to intercept testing. So, once upon a time in an agency not so far away at the National Museum of African American History and Culture, a UX designer, that is me, and a digital content producer, set out to make something better. Usually at agencies that I have worked for, we are setting out to make something better, as opposed to making something from scratch.

And what were our challenges? We had, as John Pull mentioned, a short timeframe in which to deliver results. We wanted to hear the voice of the user, as opposed to the voice of stakeholders or our own experiences speaking to the design. And we wanted to make sure that we did not push out any deadlines or go beyond our timeline when seeking the voice of the user. We had a budget of zero dollars and we had never done this before. So, like John Pull said beautifully, we were shaking hands with danger. So, lights, camera, action.

We used a method that I will, for the sake of memorizing it, I will call B.L.I.S.S.S, which I abbreviated as Brave, listen, integrate, simplify, and share.

So, B being breathe. L being Listen to advice from other industry experts. I - Integrate improved methods, with which you are familiar. S is Simplify -- answer only one question, if possible. We had tons of things we could have approached the user about, but being such a short timeframe

and the fact that this was very ad hoc for us, we limited it to one question, one design question. And then the final S is share.

So: breathe, don't panic. That might be the first inclination, that I've got this spur of the moment need for user input and I need some sort of organized methodology. But, using the method of listening to other experts, especially via listserv, especially UX-COP, I did reach out to other experts and I heard back from some very nice people who are open to sharing ideas with me at both FAA and another agency, right across the street from me, as a matter of fact.

So, as they advised, I integrated a method I was familiar with, which in this case for this presentation was Dot Voting. I had participated in dot voting at other design sessions, design challenges and even in a university setting at the University of Maryland. Dot Voting, according to 18F comprises of these seven steps, on how to do it. So, taking a proven method and modifying it for our sake, we took these five steps out of the seven steps that were suggested.

So, one, we printed the proposed interfaces that were in question.

Two, we brought plenty of sticky, adhesive dots, to the site where we were conducting our interface testing. Three, we positioned ourselves in an area, a high traffic area, where our audience would be available to us for questioning, for approaching, even for doing a little bit of interviewing.

Step Four, we asked museum visitors to select their favorite web interface by placing one sticky on the back of the paper printout of the interface. We chose to do the back, because we were trying to maintain kind of that absence of peer pressure. We didn't want them to be influenced by the decisions that previous voters or participants had made. Like psychology has proven, we tend to want that affirmation with our decision, by going with a group or crowd or what the majority has already decided. So, putting the dot on the back, as opposed to the front, where it could be plainly visible, was an important, very important step, to our methodology.

And then, five, we identified the interface design with the most votes.

So, our surprising outcome, when we asked our one question, they answered. We were pleasantly surprised that they were willing to participate. When we first went there, we were very nervous and very apprehensive, as to whether or not we would get anyone to respond to our one question. And many people answered. So, we went there with the hopes of getting five, seven, 10 people to respond. They were happy to answer, which was also a surprise. We weren't the annoying person in the mall that grabs you with a clipboard and wants 30 minutes of your time. So, we were pleasantly surprised by that, and not only that, but 75 people answered our one question, within the little bit of time we spent in the museum, which was less than three or four hours.

So, sharing the results. It turned out that we were right about their initial thoughts on the design for the interface. But more importantly than being right or wrong, we had quantified the user experience in a way that we could present to stakeholders and developers, so that they could understand. It was more than just the typical, here is what research has shown, here is what other websites are doing. It was: here is the voice of the user - the user has spoken and we have the users voice in high quantities and the numbers that are undeniable. So, with that in mind, we

shared with stakeholders, management, developers. Not only that, but internally within the Smithsonian, using the platform that we have to share documents and research internally, so that anybody who is looking for a suggestion on what to do and what they can do quickly, they are able to pull up the report and the results of what we had done. And now, we are sharing it with you.

So, some of you may be familiar with the Thomas-Kilmann model for argument. There is assertiveness versus cooperativeness. Where we were at the point at the beginning the dot voting was at a point or a crossroad where we could possibly compromise. We didn't want to sacrifice the voice of the user for internal compromises and tradeoffs. We wanted to get to a point where we were collaborating, both with the user and with each other. So, we moved from a place of compromise to collaboration, which allowed us to build a stronger design.

And the final, winning interface, according to the user, was what you see before you.

Thank you so much, Angela. We do have one question, which is, when you said quantified, how big was your sample?

We had a total of 75 votes that resulted in a matter of, I want to say, that resulted in a matter of no more than four hours. I think we might've spent three hours there, but as far as the quantity of people who walked through the area of the museum where we were stationed, we didn't do like a click number to see what our potential was. I know that thousands of people move through each Smithsonian Museum in a day, so we had high hopes for catching at least 10 or 15 of them. So yeah, 75 ended up being the total number of votes.

That is awesome. Always better to get more people than less people, right? Great.

Another question is, this will be our lightning round question for you, how do you deal with management approval to do an informal user study, without paperwork, approval, et cetera?

Okay, so for that, we wanted to go by the book at first, but it turned out we approached our internal review board and because we weren't taking any PII, we didn't need IRB approval to do ad hoc research, using this particular method. And then we wanted also permission in writing, because we weren't doing any PII collection and we weren't doing any photographs or that sort of thing, management was like, everything you are doing so far is fine and we don't see a need for a signoff. I know that safety came to the museum to make sure that they signed off on our location being safe for both of us and for possible, I want to say fire escape, in case an alarm was pulled or something like that. But that is the extent. Of our red tape that we had to go to.

Neat, thank you so much. With that, we will transition to our next speaker who is Dr. Anthony J. Schulzetenberg. He earned his PhD from the University of Minnesota in educational psychology and is currently a usability researcher working in the human factors and usability group in the Census Bureau Center for behavioral science methods. There he conducts UX research and evaluations for internal and external clients on paper, web, and mobile class platforms.

All right, thank you, Susanne. I am presenting a collaborative effort with my colleagues on developing a formula to calculate sustainability scores in multitask usability testing.

So first I will touch on why severity scores are important and useful in UX research then discuss how they are currently produced. Next, I will introduce a proposed formula, use a real test to demonstrate on how it works and then wrap up with limitations and advantages.

Anthony, before you get started, could you enter into full-screen mode for us, please.

Oh, sure.

Thank you, I did not mean to cut you off.

No, I appreciate it. Is that better?

I think it is the same. If you go to the third icon, if you see comments at the bottom, and then to the right -- exactly. Use slideshow at the top, perhaps?

It is not allowing me to share my second monitor is the problem.

You just need to select displays. So, go into the slideshow and you will see an option for switch displays at the top left. If you go into the slideshow first. There you go. And then do you see at the top left, swap displays. There.

All right. Sorry for that. Okay, so part of user experience research is testing systems usability.

The test usually requires users to complete tasks while researchers document behaviors, attitudes and reactions. So that purpose is usually to evaluate the system and identify problems that negatively impact the user's ability to complete a task. So Oftentimes, usability problems identified exceed the resources available to address all of them and therefore researchers can help designers and developers prioritize which problems should be addressed. Severity scores provide a means of prioritizing usability problems and guide limited resource allocation.

So the way researchers provide prioritization, falls mainly into two camps. The first being judgment driven. This method relies on the experience and intuition of the researcher to rate usability problems they believe should be given priority. The second way to prioritize usability problems is to use data. This can be done by assigning the category: such as high, medium, or low, or a number to the problem, using a preestablished criteria. While there is a lot of variation as to which data sources are used to determine the usability problems prioritization, some examples include frequency of occurrence, impact on past performance, ease of correction and timing.

So, both current prioritization methods of several shortcomings. Researchers have found that judgment driven priorities are inconsistent across researchers and tests. Others have found the data-driven and judgment method driven scores often do not match. Information on the users affected and what behavior really caused the behavior is lost. Furthermore, the type of prioritization is likely not as persuasive to stakeholders as ones that use data. So, data-driven methods also have their limitations. Some well-respected formulas do not accommodate usability tests that include multiple tasks. Furthermore, some metrics used to determine severity, such as impact on performance, are not always operationalized and are left up to subjectivity. Because most of these methods also involve some level of subjectivity, has resulted in weak correlations of scores among researchers. And like judgment driven methods, data-driven methods can result in a loss of information, depending on the formula being used. Lastly, the use of data-driven methods is inconsistent across researchers and has left the field kind of uncertain asked what method should be used or adopted for different kinds of usability evaluations.

So in efforts to address some of these shortcomings, we were inspired by a method by Hertzum that presents the severity of each problem as a percent of the severity sum, which is akin to R squared. This way we could see all the problems identified. For example, Hertzum states on average, half the severity sum is constrained in the first 22% of the problems. We believe this is especially valuable when communicating results to stakeholders. Unfortunately, Hertzum's formula is not compatible with multitask usability evaluations. To address this gap, we proposed a multitask usability testing severity score method, or MUTSS. With this formula we aim to

provide a reliable and meaningful measure of the relative impact of the usability problem on the system within multitask usability evaluations. Moreover, this method can provide the development team convincing evidence of usability problems and their toll on the system, as well as a tool to quantify their efforts when they are addressing usability issues.

So, to calculate the usability score of the usability problem, MUTSS utilizes a weighted average of the percentage of users who experience a problem and cannot complete the task or %C, and the percent of users that were delayed in completing the tasks or %D. Here, incidences related to failure are upweighted and instances related to increased time, but still completed are downweighted. So scores range between one and 100. Using effectiveness allows for less subjectivity and improved interpretability and moreover within multi-task usability tests some usability problem, sometimes usability can lead to task failure and sometimes can lead to task delay for the same user across the task. The MUTSS, unlike previous formulas, accounts for these nuances, because users can be counted in both task failure and task delay. Perhaps the best way to show this is there an example.

So imagine we had four users participated in a four usability test. From the results we identified and identified two usability problems. We then code which experience the problems each problem and if it corresponded with an inability to complete the task or just a delay in completion.

We see with usability problem one two out of four or 50% of the users could not complete the task and two out of four were only delayed in completion. For usability problem two, we see four out of four 100% of users were unable to complete the task and three out of four or 75% were delayed. So this gives us percent C and percent D. Next, the percentages are weighted and averaged. This results in a severity score for a single usability problem and is repeated for each problem. For this example we have scores of 50 for usability problem one and 93.75 for usability problem two. These scores can be informative in and of themselves. The percent the total problem that it accounts for tells a more compelling story. For example, in this usability test that found 10 problems, we can see the relative impact of each problem across the entire test.

So, using this table and pie chart, we can clearly see that if developers were to address the three largest problems, they can lower the total severity score by over 50%, thus reducing the identified toll on the system by half. Therefore, not only does this provide a severity score for each usability problem, it also helps visualize how much each problem impacts the system, relative to all the problems that were found. So, we apply this method to a usability evaluation on the Census Bureau data website: data.census.gov. So a sample of 19 users from the D.C. Metro region completed 7 to 11 tasks depending on their experience levels. We identified 34 usability problems from bottom-up analyses, but when we applied the MUTSS method to the data we saw that six problems with the largest severity scores accounted for over 50% of the total severity.

So when doing this, we narrowed down problems developers can tackle for the most bang for their buck. The visualization helps communicate the impact of addressing the most severe problems. We can see that many problems we found were not that impactful relative to others,

therefore priority should be given to the problems that account for the largest amounts of severity, or the biggest pieces of the pie.

There are a few limitations to this method that should be noted. First, our formula is not intended for use with usability tests that only have one task, and the scores produced do not provide information for the impact of a problem on a specific task. Also, the formula isn't sensitive to the length of delay, because all the delays are treated the same. Lastly, this likely requires more time in judgment driven decision, because researchers must first tabulate the occurrences among users. So to help make this process a little more timely, we have created a worksheet for anyone who will create the tables and graphs that we present here.

Altogether, the MUTSS method has many advantages over other methods. First, it takes into account multiple tasks, allowing for an accurate depiction of user experiences. Second, it reduces subjectivity and increases reliability. Third, it treats problems that coincide with task failure more secure than delayed completion. And the MUTTS method provides a calculation of the relative proportion of the total severity for each usability problem, which helps aid in the interpretation when looking at the entire test. Reliable scores and meaningful calculations allow stakeholders to quantify their efforts with the design team as they tackle usability problems. And finally, our Excel worksheet and code helps save time and does the calculations for you.

In conclusion, we believe that usability problem prioritization should be less of a gut measure and use more standardized methods. By using weighted effectiveness and efficiency as measures, our proposed formula can help researchers produce more reliable and meaningful severity scores. Lastly, we believe that reporting relativity reporting burden for each problem can improve the utility of severity scores and aid in communicating results to stakeholders.

So, thank you for your time. If you want the Excel worksheet R code, please email me and I will share it. I will take any questions or comments now.

Awesome. Anthony, thank you so much. We will keep going, so Margo has good time for her presentation and I would invite you to drop your email address into the chat, so people can have it from there and then if we have time at the end, we will go back and gather up the questions.

So, moving on to Margo. Margo Kabel is a user experience analyst in the Veterans Health Administration. She is the pharmacy engagement lead in the division of human factors engineering and her projects includes medical systems, such as electronic health records, and web and mobile-based products for both clinical staff and veterans. Margo, we look forward to your presentation.

Thanks a lot, Susanne. First a shout out to the organizers for setting this up as a virtual summit. I really appreciate the opportunity to present today. Let me jump right into it. We got a little agenda here that we will cover. Sort of an introduction on the problem of the lost socks, human centered design and agile. Requirements checklist. And some summary and takeaways.

So, here we are. Let's start with talking about what is the lost sock problem? I think you know what it is, but let's say it is when you toss three pairs of socks in the laundry basket, but only find two in the dryer and maybe the others are stuck to a T-shirt or on the floor by the washer, but the point is, without some vigilance, requirements are like socks that can go astray during product development. So I am going to talk about a way to address that problem. Those of you who have

already abandoned socks given to the weather, bear with me. Irote this when it was colder outside.

My experience with the problem is in the context of implementing medication information standards across VHA, veterans' health. The standards are part of a policy called essential medications directive. This addresses adverse events associated with inaccurate and incomplete medication lists and also confusing pharmacy terms. It sets some standards for the kind of information presented to staff and patients and it has a really broad scope. It impacts all applications that include meds, allergies, adverse drug reactions and prescriptions. It also takes a user centric approach, which explains my involvement.

So, here is a little example here: An image of a prescription label, from a VA medication. The standards are at work here. You will probably see this on your own prescription labels. Ibuprofen 400 milligrams tab. The VA standards dictate that we only use generic names. You have the unit, 400. You have a measurement of strength - 400. You have units, then you have a form, which is a tablet, abbreviated here. Just by way of understanding a little bit about what this is about.

So, the problems. Providing standards for consistent data and understandable language is really critical for safe and effective health I.T. products, but what we found out was the granular and iterative nature of agile software development makes it really easy to lose track of components of policies, standards, and requirements in general, without some vigilance. That was me trying to channel Alister Moody, there.

So we follow a human centered design process based on the ISO standards. In our office, human factors engineering, we work with business owners to often specify requirements for these kinds of projects. So, VA also follows agile processes, where requirements are collected and turned into user stories. An example of a user story is, as a veteran I want to see a list of all my medications, so that I can manage my medications at home. So these user stories are prioritized and stored in a product backlog, and as a result, user stories can become separated from their original context.

Here is a bigger picture of the process with human centered design as bookends to the Agile process. I can focus on the diagra at the top. The callouts, or details, you can look at later. So, some context, as we said, can be lost just moving from callout number one callout number three in this product backlog. Then you have Sprint planning, you have a Sprint backlog, that is pulled from the product backlog. People actually write code. You have meetings to review it. You create an increment of code for some part of this functionality on this product. Over time, these increments are collected, and here is the important part: the product sponsor wants to release what has been developed, but as a subject matter expert whose role is making sure there is full compliance, you need to find all of your requirements socks. Are they in the laundry basket of collective increments? We can hope. Though for example, maybe the dev team included the medication name and strength, but not the units. So maybe it just said ibuprofen 400 tablets. So, you can imagine, that might be a problem. That could be easily misinterpreted.

So, given the known “losiness” of potential of user stories, we found it is best to evaluate against user requirements and not user stories. That leads us to the next thing, how do you do that?

Enter the checklist. This is a spreadsheet. Anyone could make this. This is a detailed view where we list each requirement in column one, with detailed guidance, then we have an option to enter yes, no, partial, not evaluated, in the second column. The third column has comments to explain the whatever you put in the second column, then we capture screenshots. So, we work through our particular guidance. There are 18 elements, so we work through all of those, complete these details views, and then we have a summary view. It is a nice snapshot of this state of compliance. You can use this throughout the software development cycle and see progress over time. Well hopefully, this is what we hope for. So we track the business owner information and then we have totals for each one of the possible statuses of each one of these elements and then as you can see with version numbers, we have pages set up so we can track all of those over time.

So now let's go back to the HCD and agile process diagram. Now, with the checklist added, you can see it over there on the far right-hand side, you have a much better idea of what requirement socks are actually completed and missing items, like the units for the medication strength, can be easily identified and put back in the process. So, here we are over here. You've gone through your checklist, you've come back to the business owners and said : “here are the things we've got, here are the things we are missing. Can we find them back in the process?” We can send them back in, put them through the laundry again and have some confidence that we know when they come through there or not.

So, in summary, providing standards for consistent data and understandable language is critical for safe and effective health I.T. products. However, the granular nature of Agile really throws a wrench in our attempts to keep track of all of those things. So this is our solution, having lost many socks in the past, we are now going back to requirements and not user stories, conducting iterative evaluations, using a checklist of defined standards or requirements. This provides quick feedback to the team with a visual summary of the degree to tell them where they are. You’ve got a snapshot and you know.

Takeaways, really helps me keep track of all of these small, but really important pieces of requirements. As you can see, pharmacy requirements, I would describe as details on top of details. These checklists can be used to evaluate just about anything, really. Something you're thinking about acquiring and you can use them to evaluate products as they are developed over time. But, more specifically in our case, it has increased the consistency of information across our VHA products and this increases usability, patient safety, and reduces cognitive burden on users.

So, I've got references. I have a question, how many socks that I start with? What color are my socks?

Margo, thank you so much. The question we have gotten his how are you presenting requirements and acceptance criteria and user stories? Do you refer to the checklist or does each user story have its own acceptance criteria?

The user stories have acceptance criteria. We tried to make those pretty detailed, so all the requirements are covered, but sometimes, but they get interpreted differently and parts tend to fall through the cracks. It may be a function of pharmacy that has just so many detailed

components, that it is especially problematic. But that is why I have started going back to the actual requirements in this case, to make sure we have an end to end check.

Thanks. Let us come back to a question for Anthony and I am going to combine two of them. One is, with with an N less than 30, is it really worth being this rigorous with your statistics. It seems like a likert scale would be just as valid and easier to calculate. And a similar question to that is, what you do in cases where you only have only one or two users users you are allowed to test?

Yeah, so we're not claiming any sort of significance, so of course usability tests will have a small sample. Where this benefits the researcher is the waiting and putting a different emphasis on task completion or task delay. So a likert scale wouldn't really capture that, though by looking across that, weighting it and giving percentages, is where we argue that this would be more consistent for stakeholders. If you only have one or two users, one, it might not be much. The number of users isn't limited, but it is more the tasks. If you have more tasks, you can gather more data. That would be more useful and probably be more representative more rative in the output.

Great, thank you. So, thank you to John, Angela, Anthony, and Margo. And Jean, take us out of the session. Remember that there is another session this afternoon to come back to.

I just want to thank the prisoners. This was an awesome session this morning and a really awesome session and a great way to kick off the summit. As Susanne said, there are five more sessions to the Summit and you see in the chat, links to register if you haven't already. I look forward to seeing everyone in the future sessions, so please join us. Thanks very much.