UNITED STATES GOVERNMENT PRINTING OFFICE

DOCUMENT AUTHENTICATION WORKSHOP

Friday, June 18, 2010

732 North Capitol St., NW

Washington, D.C.

9:10 a.m.

P R O C E E D I N G S

MR. HANNAN: Welcome everybody. Thanks for coming today to our workshop on Document Authentication. I'm John Hannan from GPO. I want to extend a welcome to you. I haven't met everybody but thanks for making time today. We really looked forward to this.

Just a couple housekeeping things to get started with. There are restrooms right down the hall on your left as you go out -- straight out this door, on the left. I think the men's room is first on the left.

We have a transcriber here today, as is often the case at the library conferences, and so later on when we have opportunities for feedback, we have three handheld mics that we'll pass around the room. If you'll please wait for a mic and if you'll pass the mic around and then if you'll state your name and what your organization is, that will help with that. And we'll post the transcription later on at the end of -- about a week from now, just five to ten business days (inaudible). We'll plan to post that for folks.

Lunch, around 11:45 we'll probably break for lunch and that will be on your own. Since the GPO cafeteria is available we'll show you how to get over there to get some lunch. And then we'll reconvene at one o'clock to get started for the afternoon session. So -- and we have a break in the morning and break in the afternoon.

Unfortunately we don't have wireless in the room here, we're working on that but it's not ready yet so our apologizes on that. But there are plug ups for your laptop if you want to run your laptop.

I'm thinking here, Lance did I miss anything?

MALE VOICE: I don't think so.

MR. HANNAN: Okay. Oh, and everybody looks like they've turned their tent cards around so that folks can see and the transcriber can see, that would be great. I'm going through my checklist that Lance gave to me, I think I'm on track here so far.

There's a handout from you which are all the slides. They're up here too and we'll kind of step through that. They're mostly to provide some background and give us some guide to help bring some

feedback from you all for some questions and some topics that we'd like to hear what you all think about these things, as we talk about some of the finer points of document authentication as we look into the future here at GPO. We really appreciate you being here to give us some insight and some thoughts.

So before we get started, thank you, we'll go around the room and just do a quick introduction for everybody. We will go ahead and give everybody -- after the conference we'll provide a list of who was here and we have a website link, it's at the end of your handouts -- where things will be posted later on about our workshops and things.

So with Lances help, I'm John Hannan from GPO, I'm the Chief Information Security Officer at GPO and I have a long, pretty long by now background in data integrity and electronic authentication. Since I've been at GPO about six years now I've been working with people on authentication. It's been really exciting and a pretty fun initiative overall to be able to bring some capabilities to the table (inaudible). So that's me and we'll go around to my left.

MR. SMITH: I'm Steven Smith. I work with GPO for John Hannan. And I support the PKI Instruction.

MR. WASH: I'm Mike Wash, I'm an IT here at the Government Printing Office. I'd just like to thank everybody for coming. As you know, authentication is one of our foundational aspects in our content management activities that we've been working on for the last six years in our federal digital system. And as we positioned, you know, early on, when we were you know, conceptualizing this system, we realized that authentication was going to be a journey.

We have tools today, using Adobe Live Cycle that allows us to digitally signed PDFs, as you all I'm sure are well aware. But it really isn't enough. There are new format types that are coming available that people would like to have signed or in some way capable of being authenticated for their use and we've got to find solutions for that. Plus a lot of the documents that we sign today are very large and subsets of those documents would be nice to have signed as well or authenticated. So a lot of our discussion today is

really to help us understand, from your perspective, what some of the real value propositions are for authentication and what the requirements would be for authentication so that we can frame it properly as we continue to move on this journey.

Later this year we expect to have, and I'm probably taking some of your thunder here, John, but later this year we expect to have more of one of our technical work sessions which we call an Industry Day. And an Industry Day is where we bring the supplier community in to hear our story about what we think is needed so that they can feedback to us and tell us how they think they could help us solve their problems. So this session is really to help us prepare for that, because the last thing we want to do is have a room full of sales people telling us they have the greatest solution since sliced bread, which they all like to do. Instead we like to be able to position what we believe is the opportunity set and the value proposition we need to deliver so that we can really help manage those future conversations to be as effective as possible to provide solutions that will ultimately meet everyone's

needs.

So again, thank you all for coming. I think it's going to be a good session. I really appreciate your support with this.

MS. SEARS: Suzanne Sears from the University of North Texas. And I'm here representing the Federal Depository Library Council.

MR. PRIEBE: Ted Priebe, GPO. I'm with the Library Services Business Unit. I work for Rick Davis. Our unit collaborates with the content originators in terms of getting their support to authenticate collections that ultimately will flow into FDSYS, the Federal Digital System. Thanks for attending.

MS. BAISH: Mary Alice Baish, I'm the Director of Government Relations for the American Association of Law Libraries and obviously the issues of electronic life cycle management, permanent access, authentication, official status, preservation are key components for our community, particularly as we all know, when the official titles are disappearing.

So I would like to thank GPO on behalf of our organization for holding today's event. And also I'm

very excited that it will be followed up by an industry event.

MR. YU: I'm Harlan Yu, I'm a graduate student in computer science at Princeton University. And for the time being I'm at the Department of Labor working (inaudible).

MR. BOOTH: My name's Harold Booth. I'm a computer scientist at the National Institute of Standards and Technology in the Computer Security Division.

MS. CASE: Pat Case, I work with the Congressional Research Service and I work on LIS which is the Hill version of THOMAS.loc.gov.

MS. MCGILVRAY: Jessica McGilvray, I'm the Assistant Director at the Office of Government Relations at the American Library Association.

MR. GEE: Robert Gee with the Law Library of Congress, primarily working with THOMAS. I'm also doing some utilization initiatives with GPO.

MR. COGGINS: I'm Tim Coggins, I'm the Associate Dean for Library Information Services at the University of Richmond School of Law and have been with

the American Association of Law Libraries efforts (inaudible).

MS. DULABAHN: I'm Beth Dulabahn for the Library of Congress and I'm involved with their digitization programs and the actually digital information infrastructure and preservation program.

MR. ANDERSON: Which is N-DIP (ph) for short. I'm Steve Anderson, I'm the Director of the Maryland State Law Library in Annapolis. And like Tim I've been working with the American Association of Law Libraries tracking authentication issues for some time.

Also, this year in Maryland we did -- well over the past couple of years we've had the occasion to try to find out what some best practices are for authentication for state publications. Our state publications depository and distribution program needs to have some guidance at some point, in my opinion. And we were faced with almost eliminating the print version of the Maryland Register, our federal register equivalent, which did not happen fortunately, but there was no Plan B in terms of authenticating any of this. So to the extent that there's some possibilities of

cross pollination that would be fantastic for me to pick up.

Lastly, on a personal note, I've got young onset Parkinson's Disease so if you see me shake a little bit or something like that, just pretend I'm Michael J. Fox and I'm being famous. Thank you.

MR. TAOULTSIDES: I'm George Taoultsides Research Librarian at Harvard Law School Library.

MS. ZWAARD: Hi, I'm Kate Zwaard, I work at GPO and I work on the Federal Digital System responsible for the digital preservation component which includes content integrity and (inaudible).

MR. HORTON: I'm Bob Horton from the Minnesota Historical Society. I'm the happy recipient of an N- DIP (ph) grant to work with the preservation -- - long term preservation of digital content from state legislatures and I'm an observer for the Society of American Archivists on the National Conference of Commissioners Uniform State Laws Task Force for (inaudible) Law on Authentication and Preservation, religious related or legal -- digital legal documents. Is that what they're calling it?

MS. RUSSELL: I'm Lisa Russell from GPO. I'm the Library Services and Content Management Division and I've been working on the authentication initiatives for a few years now.

MR. WEBER: I'm Andrew Weber at the Law Library and the Library of Congress and I work a lot on THOMAS.

MR. SHULER: John Shuler from the University of Illinois at Chicago. I'm also a member of the Federal Repository Library Council.

MR. GALLUCCHIO: Kevin Gallucchio, Department of Defense. I'm supposed to be some sort of subject matter expert with XML.

MR. MAYER: I'm Jonathan Mayer, I'm a fellow at The Center for Computers and Law at Stanford and a graduate student in computer science and law.

MS. LAPLANT: I'm Lisa LaPlant from GPO. I'm in the Program Management Office and I work primarily on the access and delivery and search portion of FDsys.

MS. DALECKY: I am Selene Dalecky. I am also in the Program Management Office here at GPO. And I'm the FDsys program manager.

MR. DAVIS: My name is Rick Davis I'm the Acting Superintendent of Documents at GPO. I'm also the director of the Library Business Unit that manages the FBLD (ph).

I want to thank all of you as well for coming this morning and give a little bit of background as well as to add to what Mike mentioned earlier. Back in about 2005 the Deputy Public Printer at the time approached me and said, "You know, we've got a growing crisis with digital documents and ensuring their authenticity. We need to stand up a PKI." So being a policy person and nontechnical person, the first thing I did was go to a library and say, "Well what is PKI all about?" Well from there, working with John Hannan, our IT staff, you know, here we are today.

We're now in an environment that many of you see where in the federal sector over 95 percent of all the new publications that come through us, that we make available to the Federal Depository Libraries are digital. In some cases not only are born digital but there is no print equivalent. In a primarily print based environment, when the Superintendent of Documents

sent out government publications sealed in a tied back environment, there was a reasonable assurance that the recipient, whether it was a sales customer buying our product or a library receiving it, had a reasonable assurance that they were actually getting an official and authentic federal publication.

The challenge that we're now facing is when a publication is made available digitally from GPO or another agency and is indexed and Googled and picked off on a number of other sites, how does the user, how does the American public, the researcher know what version to cite for -- in terms of being the official and authentic version.

And we started addressing that issue by applying digital signatures to our content with GPO Access. It's now been expanded through our federal digital system. But I think that, you know, part of what we'll discuss today are some of the additional challenges that we have going forward in this environment dealing with levels of authentication, not only for born digital content, but as we work with digitized content, tracing back a chain of custody and

what chain of custody really means for different levels. Likewise, looking at granular levels of authentication that John's going to talk in more detail about.

We have users who are saying, "It's just not enough to authenticate an entire file that encapsulates the publication. We want to be able to take snippets of information from that publication and have the authentication carry forward for additional needs." So these are some of the challenges that we're going to be facing. And again, I thank all of you for being here and look forward to the discussion.

MR. SCHWEICKHART: Good morning, my name is Reynold Schweickhart I'm the Acting Chief of Staff and the Chief Technology Officer. On behalf of the Public Printer and also the fifth person to welcome you today. This is an important workshop to continue the progress we've been making as we look at GPO's historical mission to really provide authenticated government information which we've been doing in the repository library program since 1813.

On a personal note I've been working in this

space since I came to work for Congress and we worked closely with GPO and the Library of Congress and CRS, including THOMAS. And I also helped form the working group that began the migration of legislative data into XML which was now many years ago. So I'm excited to be here. I think this will be a very productive day and look forward to engaging with everyone.

Let me just go back and -- we've been going around and introducing ourselves Daniel, so let's pick you pick up here.

MR. BENNETT: All right. I'm Daniel Bennett with the eCitizen Foundation. And we have worked with the folks at the House and the House administration on efforts to look at XML for these legislative documents.

MR. SHAPIRO: I'm Arin Shapiro the webmaster for the Senate. I manage the central portion of Senate.gov along with several intranets. I work with a lot of the agencies here on a bunch of issues.

MR. HANNAN: Jim, why don't we go around and start with you on the folks or let's actually (inaudible).

MS. COOKS: I'm Judy Vance-Cooks (ph), Managing Director for Publication and Information (inaudible).

MR. BENDER: Jim Bender, I work in the Production and Engineering Division of Plan Operations and we primarily (inaudible) various GPO publications such as the Congressional Record and the Federal Register and some of the different legislative work that comes down from Congress.

MR. GREEN: My name is Lyle Green, I work in the Office of the Congressional Publishing Service at the GPO. My office serves as liaison between GPO and Congress.

MS. SWIATEK: My name is Kathleen Swiatek (ph) I'm a IP Specialist in Congressional Support at GPO. And I primarily work on XML projects on congressional publication bills and we're starting to report some other documents also.

MR. KAUFFMAN: My name's Mark Kauffman, I'm from GPO. I work for Mr. Wash in IT and I'm the Project Manager for GPO's (inaudible).

MS. SABOL: Hi, I'm Janice Sabol, I work in

Planning and Development in Library Services and Content Management.

MR. HANNAN: Great. Well I think we'll go ahead and get started. I believe I mentioned there are some handheld mics so when it comes time and you'd like to provide some input and speak, if you'll please turn it on and state your name and who you're with. And if folks would help by passing that around that will probably be good.

So with that we'll go ahead and get into it and what we'll do is, this is the agenda for today, kind of a guide for discussion. We've obviously been through the first part.

Authentication background, individuals as well as for automated high volume uses. Standards and methods, we're going to spend a lot of time this morning on standards and methods because that's an area where we'd like a lot of input. And we have some questions, we have some thoughts but also some questions. When we come back from lunch we'll go into chain of custody, type of use cases, re-authentication over time. That's really more just just awareness for

you more so than input, I think, but we're here to listen. So we look forward to that. And then lastly, granular authentication and then we'll wrap up.

A little bit about what's in the scope for today, what's out of scope just because we don't have that much time really for some of these topics. We're really talking and want to focus on authentication for the electronic documents that we've assembled at GPO. What we don't want to spend time on today is authentication as we might apply it to tangible or other products at GPO (inaudible).

A little bit about what we're thinking in terms of the desired outcome. Really we said this a little bit but, the next slide was a slide for your handout, we'd like to get feedback from the various stakeholder constituencies we have here to help inform our decisions that we're going to have to make about future system deployments. You know, we don't lack for a lot of different ideas or methods, what we really are looking for is to see if there's some consensus that will help us make our decision about how to provide the best value for citizens and all of our constituencies

here today. And as Mike said earlier, this will help us form the input for the follow on industry data that we will have. We'll be asking parties about what kind of technical and other systems they might provide to help us in that regard.

So some of the decision topics for today, the first one is does the community really require different levels of authentication assurance on the same content. Some parties are willing to use a les robust means and others requiring the current digital signature, is that really something that's important or not? It's obviously more complicated and expensive for GPO to do, but this is the area of feedback that we'll get in to that we think is important and want to hear from you on.

Next we all know XML is a very important format. We're already publishing or disseminating rather information in XML format. What standards and techniques ought we to use for that format. Again, there's a few different ways to go with that and we'd like to get some input and your thoughts on that.

And what techniques and standards ought to be

applied to chain of custody? There's a lot of different ways to go with that. Again, one of the big things in this regard too is content originators really need to support the electronic dissemination of that to some degree. We can't make that call, I think, on our own at GPO. So there's more about this as we get down there, but that's an area we'd like to get some feedback.

And then lastly, granular authentication. We'll define what that is or scope it at least, for the purposes of our discussion today. It could be many different things. Some areas maybe not as concerned, others very concerned with that, so.

Am I speaking loud enough for the -- okay, good. Most people say I'm loud, so.

Some background on authentication. As you probably know, we're publishing or disseminating PDF files that are digitally signed using encryptographic digital signature in PKI, and that's a specific kind of method to use. That method uses an open international standard that's really available for all parties to write software code to do. And PDF is the standard as

well, again that people are free to write software codes to process that kind of information.

As we go forward you're going to see those are a couple of key points in all, from our perspective, in all that we do. Not having proprietary methods rather having open, international standard methods that are available to all parties to use. There's some other features as well.

For this we've gotten mostly positive feedback from the stakeholder communities, at least that I'm aware of working and talking with Rick and his teams and Mike and their teams and everyone. I'm not quite as close to all of your communities as Rick and his team and Mike are, but I think generally that's gone pretty well and we're really gratified. We hope that that's provided some value to you.

That's been, from our experience, pretty easy and reliable for users and citizens to really achieve, knowing they have the electronic versions that came from GPO. And if somebody tried to insert one, purported to be from GPO, you could detect that easily anywhere on the planet. So that's been our experience

so far. And again as Mike said, in the first step on our journey with PDF that's been our experience so far, but there's a lot more to it, since we think about other content types and such.

Again a little bit more about the background. We started off in 2008 actually with the President's budget on GPO Access. You all probably know most of this, so I'll go through it really quickly. Congressional bills we started in the 110th Congress with Rick's and Ruth and Lisa and we've a lot of our folks working on that. FDsys data started, it's been doing signing from the get-go in the Federal Digital System or FDsys as we like to use the acronym here at GPO, signing all the collections as the collections are disseminated, the PDF files with digital signatures are available.

So our goals, what are our goals as we see them for authentication? This is to provide recipients of documents or readers of documents with who the source or disseminator of the document was. GPO is the source for the GPO documents. Going back to what Rick said, as trying to provide some value to a larger

community in reducing confusion about what did GPO disseminate, kind of the definitive way to deal with that in an electronic, faceless Internet-oriented world.

Second bullet there, providing some assurance that as you're reading the document it hasn't been altered since GPO put it up on the Internet. So it's also a means of reliably detecting, quickly and easily for users, if it was altered. So that's a goal for authentication programs.

And then lately to provide a method that supports authentic chain of custody so that the chain of custody can be reliably provided not altered as it leaves GPO. And again, a way to reliably detect if it was altered. This is obviously something that's further down the journey for us and one of the reasons we're having today's session (inaudible) about that.

So at kind of a high level those are the three high level business and function goals we're trying to achieve.

Okay, some factors that seem important to us about how we take the next steps on this journey. We

want to use authentication techniques that are strong enough to maximize the length of time the authentic technique is valid. In other words, we don't want to pick a technique that turns out to be so weak that we have to go back or redo it next year. That kind of wastes everybody's time, or mostly ours. And that's not a good thing.

Authentication techniques that work even when you're disconnected from the Internet. That's obviously really good technique but those are the the kind we'd like if we could get them. There are plenty of techniques that require you to be connected to the Internet, but it also works when you're disconnected from the Internet, we think that's valuable but we'd like to get some input on just how valuable that is as such.

Authentication techniques are based on open, published, established international standards as opposed to proprietary method. That's obviously a key question. Authentication techniques that are based on binary data or open data formats rather than proprietary data forms, that's what that last bullet's

really about, open, freely available formats not proprietary formats.

Authentication technique that provides a clear and simple display of who the disseminator of the document was, not some convoluted technique or requirement for 17 different mouse clicks but hopefully something a little simpler. Again, some techniques that they really, for a data format, require a lot of complexity, well -- but we're going to look to and guide -- our decision thinking is guided towards simplicity where it's possible to still achieve the other goals.

And then lastly, a technique that's extensible for chain of custody. In other words not something that we necessarily need to do at the beginning but something that's easily extensible if we want to do that, either from the start or later on down the road.

So those were, to us as we talked about that at GPO, some things that seemed like some of the overarching principles. So at this point we wanted to kind of stop and see if that makes sense or if there

are others that you think we should be considering too in that regard. So I guess at this point we'd like to go ahead and open it up for some feedback or some thought process or thoughts from others.

That mic over the in corner, Harold, that's the one mic where you have to press and hold the button for it to go green, so -- okay, you've already figured it out. Great. Thank you. The other mic has a little on/off power switch at the top and you flip it to power.

Again kind of a little abstract topics there but those seemed like the criteria that we ought to be looking at. Are there others that come to mind? If not we can just move on into and get a little more specific about some of the topics. And you'll have plenty of time to send us feedback too after the conference, by the way.

Yes, Mary Alice?

MS. BAISH: Just one quick question.

MR. HANNAN: And if you'll just -- it should be -- I think it's on already.

MS. BAISH: It's on. Can you hear me? One

quick question on page seven you mentioned that you're not there yet. Where -- tell us where you are right now just in terms of your application goals.

MR. HANNAN: The chains of custody?

MS. BAISH: Right on page -- where you have on page seven, you know, provide assurance that a document, that the source identity is named, that GPO is the source or OFR is the source, that it hasn't been altered. And then of course the very sticky third rule about chain of custody.

MR. HANNAN: Sure, so for PDFs today, those first two absolutely are already provided. The third one is easily providable, there is a little coordination with content originators on the PDF specifications, if you will, providing a place for follow on signatures. We really haven't quite gone there yet but we're fairly easily able to provide that kind of thing, potentially working with content providers in the Congress, for example, or the Federal Register on that, for PDFs.

So for PDF the techniques line up pretty well. And one of the subjects for today is some of the

choices regarding XML, dated XML or other formats that people want to talk about.

MS. BAISH: Thanks.

MS. ZWAARD: John, can I jump in reason quick. So a few things that FDsys does --

MR. HANNAN: Can you go ahead and state -- yeah.

MS. ZWAARD: Sorry.

MR. HANNAN: Your name, organization.

MS. ZWAARD: This is Kate Zwaard, and I'm from GPO. So in order what FDsys is doing for authentication, providing the identity of the publisher for the document, we do do that in metadata. So if you're a direct user of FDsys, if you're not receiving that file from an intermediary, and you trust FDsys as a repository, you can be assured that the document you're looking at is provided by that issuing agency from metadata we provide you.

In terms of providing assurance the document wasn't altered since publication, we do provide checksum (ph), a hash using the shop 256 (ph) algorithm for each file we make publically accessible. So if

you're technologically able, you can check that checksum to make sure it wasn't maliciously altered.

And in terms of chain of custody, we haven't gotten as far as being able to make reliable that it hasn't been altered, but we are using the premise data dictionary and premise schema to provide a digital (inaudible) chain of custody from the issuing agency through all of the things we do to make sure that it's, you know, we digitally sign it, we do checksums, we identify the file format, so it gives you that long list of things that have happened to the file since we've received it.

One of the things GPO is also, you know, thinking about is how does authenticity of a document change over time once we're doing digital preservation things to it. So once -- when we're migrating documents to preserve it, you know, it may be altered a little bit. How can we assure users that that's still authentic, that it still contains, you know, the meat of the content. And so that chain of custody allows the user to inspect that and make that decision themselves.

MR. HORTON: Could I just ask one question, and I'm not sure if this apropos, but is there a sort of intellectual premise here that material in a digital format is unreliable? Or is the premise that it's just de facto because it's in a digital format? Or is it reliable until proven otherwise?

MR. HANNAN: I think that's an interesting question. And so this is John Hannon from GPO. The way I would address that is the fact that information's in digital form provides more means for nefarious actors and the ecosystem to be at work if they wanted to. And therefore some additional thought processes and techniques might be appropriate to help the community assure itself when they are looking at a digital data object of something. So yeah, that's how I would look at it.

MR. SCHWEICKHART: Reynold Schweickhart from GPO. I think particularly when you think about the timeframe that GPO's responsible for providing authenticated federal content, sort of life of the Republic, over that span or even if you go back just before Y2K, digital media is actually pretty fragile.

And so it's at this point more fragile than high quality archival paper stored in many locations under correct circumstances. I think that is the reason why we're going to look more at how we do visual authentication.

MR. HANNAN: Let's go to Mary Alice, because I think you were ready and then we'll go to the back of the room.

MS. BAISH: Because I never did get my question, but just based on what you said, I mean the whole issue of the cyber security is a really critical one. And I know Senator Lieberman in fact just introduced a bill on cyber security and DHS is probably going to take the lead, but those are crucially important when our government relies on so much information that is only available an electronic forms.

Just my comment about your goals, you know, are these enough. I think in an ideal world it would be nice if every member of the public knew that they could go to GPO Access and they will find the official electronic version of the Federal Register or a law. But I think the fact is with the growth of Google and

everything else people just go anywhere to look at the information and make the assumption that it is perhaps an official version and they can rely on it and they can trust it and they can use it.

So I think one of the ultimate goals which I would like to see if that whether I download a public law from GPO Access or whether I come to FDsys and look at it there or whether a court opinion is in -- available through a third party commercial publisher, it would be very nice if at any -- under any of those scenarios, wherever the electronic file resides, a user could check the authentication if that's what they needed to do.

MR. HANNAN: Right. I think would be great for the whole system.

Lyle Green had a question. Can you give the mic to Lyle? If somebody could pass the mic. Thank you, Kate.

MR. GREEN: Lyle Green. Just wanted to bring out one other -- I guess a different aspect on the question about the reliability of digital format. The authentication doesn't really bear on the issue of

whether the printed document matches the electronic version. Authentication, you know, is more geared to any alteration to the digital version. But as far as, you know, we make the assumption that the digital version matches the printed product but there is no control in the process that verifies that.

MR. HANNAN: Well, you know, I mean Mike or Reynold do you all want to comment on that? This is John Hannan from GPO. You know, I think the -- Kate kind of spoke to this too a little bit, about internal processes at the GPO in terms of receiving information electronically or otherwise from source agencies as well as the topic that Lyle just brought up about printed things that GPO obviously is control of producing and the electronic version. So there are some things there at GPO that help assure that those things line up.

MR. SCHWEICKHART: This is Reynold Schweickhart. I would just say that in the production process there are quality assurance steps to assure the electronic and paper product are synched up. I think that's a general issue as we look at importing content,

for example, from websites where documents can be posted. We don't have that same deep knowledge about production process at that particular agency to be able to ascertain whether that document, in electronic companionable form in fact is identical or you've gotten a slight revision somewhere along the line.

MR. HANNAN: That's a very good point. All right, question over here.

MR. BENNETT: Hi, Daniel Bennett with the eCitizen Foundation. One of the things that is going on in the entire federal government is the data.gov and other efforts to put out information and to have that sort of swimming out there. And so the question of authentication in terms of that form of data I think is going to be solved not so much by encryptographic means but by using URLs that can -- where you can find things and be able to cite to them.

I mean people are grabbing things and so they grab a portion of the census data, they grab a portion of other things and they mash it up with what Mary

Alice was talking about a portion of something that's published by the GPO, the question is

how -- when those things get mashed up can you authenticate them. And having a good URL citation system so that people can go back to the https:// and having that embedded within all of the electronic documents, when people grab pieces of it they'll be able to do it. And then the other aspect of that, if you think about it is, then we can externally hash things and put -- and not just have GPO hash them but other people will be able to grab portions of things which will include both the reference back to the authentic version that will be captured as it goes through a long chain of custody, always pointing back.

And then also allowing more people to not only have copies that can be referenced and just see that they're all the same, but do their own hashes so it isn't just the centralized hash that everybody has to trust, but more and more people. I know as specialists we think of this as a much more distributed publishing system (inaudible) other government agencies and commercial organizations putting out data themselves. That's another way to think about it.

MR. MAYER: This is Jonathan Mayer from

Stanford. I guess going to the factors on page nine, it seems to me that a serious concern as we think about this would be what consumer is the data targeting, the authentication practice we're targeting. So certainly extensibility, standards and so on are very important, but the extent to which they're important you could imagine differing substantially based on the consumer.

So for example, if you're targeting your data release to a web intermediary it might be -- well it's important for it to, you know, carefully display in an easy format the authentication mechanism, on the on the other hand, not if you're targeting the casual consumer, then you might want to think more about some sort of commercial software that just displays the blue ribbon and so on. And so as we go forward that seems to me an important factor.

MR. HANNAN: Other questions? Yes.

MR. YU: This is Harlan from Princeton. I'm wondering whether we should be thinking about data that's coming out of agencies as well as data that's coming out of the GPO. So, you know, once we start distributing the authority to sign documents, say at

the Department of Labor, then there's a big key management problem. And so I was wondering whether or not that's in the scope of today's discussion or we're just talking about documents that are coming out of GPO where, you know, we have a private key and we can sign things related to that, to deal with the key issue.

MR. HANNAN: Sure, Harlan.

MR. BOOTH: And I want to add on to that actually.

MR. HANNAN: Yeah, sure.

MR. BOOTH: One of the things that I think --

MR. HANNAN: Could you go ahead and let people know -- if you don't mind.

MR. BOOTH: Oh, Harold Booth. One of the things that I wanted to add on to Harlan's comment is that I think one of the things that you may want to determine is that you have lots of signers and lots of verifiers or do you have just a few signs with lots of verifiers? It seems to me you have lots of verifiers, so I think I've already got that answer. The real question is do you have lots of signers? And I think that kind of dovetails with Harlan's question.

MR. HANNAN: Yeah, and I think the chain of custody topic really speaks to -- yeah, we do have a lot of signers, potentially, if content originators, you know, choose to do that. So I think that's the challenge for the GPO obviously is to be able to, you know, set up a -- or to look at providing and facilitating things that may get extensible to either case, the case where there are lots of signers and obviously there are lots of verifiers.

MR. YU: Is it the role of the GPO to provide a signing service for the rest of the agencies? Is that something that's in the scope of what the GPO might do or --

MR. SCHWEICKHART: Let me --

MR. HANNAN: Yeah, go for it.

MR. SCHWEICKHART: My short answer is if we could make money at we would. Just a couple of comments. Our PKI is cross certified with the federal bridge. So from a chain of custody point of view we could certainly pick in documents that when signed by any federal employee, contractor, anyone who's a participant in that bridge process, one of the things

that we think about in that context, however, is the re-authentication problem of what do you do 100 years later in terms of looking at that chain of custody. Where -- how do you sort of nail down that authentication as sort of documents enter the system, as they become part of the FDsys and so on.

And so I think there's two parts for that, but broadly speaking we're set up to take -- and in fact when we took the first federal budget from the White House, it was signed by a White House employee with a digital certificate on their HSPD-12 card to do that. So we have a -- I think we have a robust infrastructure depending on where we go from a policy direction.

MR. BOOTH: I'd like to ask a question regarding that particular point. How do you know who's the authoritative person for determining that that was the budget? So let's make it more general, for a particular document how do you know that that's the person that's responsible for doing it, as opposed to maybe having -- as an individual how would I know that if there was some ort of general purpose White House

budget certificate that then would be kind of clearly labeled, that's the role that that certificate is intended to play.

MR. HANNAN: Well let's --

MR. SCHWEICKHART: Yeah, I was going to -- let me answer that shortly and then maybe if we get more technical we may want to take it offline and go to other topics. But that's really an out of band (ph) process today. We know who the person at the White House is, the Federal Register knows who at the Department of Labor is authorized to issue regulations and so there's a out of band process.

MR. HANNAN: In today's world, and there are ways of moving that into an electronic world with education, some communication about that. These are challenges that we've always had in the tangible world and solved them quite easily and I think we'll be able to do that, although it will take come some communication and there will be a change in process for people. Jonathan (inaudible) for a second.

MR. DAVIS: Rick Davis, Government Printing Office. Adding to that point, from an agency

perspective, at GPO on the XR519 certificate it's authorized by the Superintendent of Documents which is me. In terms of our policy that we have written, in terms of other federal agencies, we actually ask that it be at the head of the agency level or a designated authority by the head of the agency so it's coming down from the top down.

MR. ANDERSON: Steve Anderson for the Maryland State Law Library. I'd like to reiterate what Reynold said about making sure that the chain of custody in effect and the authentication principles are in effect for the life of the Republic. I think looking at the long term and the longest tail we can on these documents is going to be very important. Coming from a private law firm background to the State Law Library several years ago and finding that we had 400 year old books on the shelf was an eye-opener. And I want to make sure that -- it would be nice to make sure that whatever we do is going to last, you know, as much as possible.

Going along with that, I don't mean to suggest that we do take any type of blocks approach in

terms of having duplicate copies of an authenticated document because that just gets weird. However, I think being able to have some type of very quick, easy, on the fly redundancy measure and enabled to pour things over between system and system is very important.

MR. HANNAN: Thank you.

MR. BENNETT: I think that --

MR. HANNAN: Can you just for the transcriber's benefit, restate your name.

MR. BENNETT: Sure. Daniel Bennett with the eCitizen Foundation. I think having drafted the first legislation on e-signatures, electronic signatures, I think we're having a little bit of potential confusion but I think we might as well just talk about it, in terms of authentication because they're signing something so a lot of stuff which is legislative, people co-sponsor, it gets signed by the President, there's authentication for the purpose of just signing, which is separate than the authentication for the purposes of things like identity for the purpose of the chain of custody to know who was there. But it is a

weird thing once you move into electronic and you have people's -- what look like physical signatures and representations in the future perhaps of their electronic signature, that you will get into a bind of how you deal with signatures in that sense for authentication versus authentication of making sure the document itself is authentic.

So it's something -- we might as well start understanding and thinking about it. But for now probably just want to put that to the side and just understand that we're talking authentic documents and chain of custody issues and not the other.

MR. YU: And if I could just add to that really quick. I also want to make sure we're clear on, in terms of terminology when we talk about authentication I think that that basically guarantees, you know, who that document came from. There's another topic or another concept of integrity which is what hash functions give you. Right? So when you use a hash on any pile of bytes all that gets you is integrity, it doesn't give you authentication. It doesn't tell you who hashed that document, who made

that document, but a cryptographic digital signature would give you both.

So I was concerned in the earlier slide when you talked about less robust means, I was worried that if you use something that was less robust like a hash that didn't involve any (inaudible) material, then you would only get integrity and you wouldn't get authentication.

MR. HANNAN: That's exactly right. We have some slides to try to get that very point to kind of comparatively analyze some of the techniques out there.

So this has been information that at GPO we're pretty familiar with. You know, and it kind of -- - Harold's question really sparked that thought, which is good, about there are many ways to electronically sign a piece of information. The Electronic Signature Act, it's technology neutral, it gives lots of different options for how you can do that. And so that is different from the electronic authentication of a document, although interestingly enough you can kind of use some of the same concepts in both arenas. So we'll talk through more of that today, a little bit. I think

that's a great comment.

MR. GALLUCCHIO: Hi, this is Tim Gallucchio from the Department of Defense. I just wanted to warn you that adding a long tail on this document signature is a worthy goal and should be done to the fullest that you can. But you should also -- you should also realize that this is, you know, you will have to resign eventually, whether the (inaudible) gets broken or just when time expires. Computing is getting better all the time, faster and faster, newer craze so eventually the (inaudible) is going to expire and you will have to resign and maybe we'll all be back here or our children will be back here at a similar workshop discussing similar things.

MR. HANNAN: We do have a topic later talking a little bit about that thought process of re-authentication over time, not so much to get a lot of input, because -- although we're happy to -- again, it's more just to a brief on what we're doing. Great comment, yeah we're aware of that. Some techniques can be more involved in that regard than others.

MR. GALLUCCHIO: And I also -- I don't know,

the way I see this, I take it you would provide a method that supports authentic chain of custody. I think you want to propose this standard, you know, you're providing a standard and then you're going to have different companies implement this standard and then the user will go to this webpage and then they'll click on it or --

MR. HANNAN: Well, I think that's interesting because we're not really a standards body, you know, so that really is why it's stated that way. I appreciate your comment, but we're really not in the business of creating international open standards for everyone on the planet. That's why we've got a lot of folks here from different constituencies so -- again I think GPO is ready to take advantage of the best thinking and possibilities that are there to help bring value to the community. And this workshop is an example of trying to get some of those. Thanks.

MS. BAISH: Very quickly I wanted to comment on your point, John. Mary Alice Baish, American Association of Law Libraries.

Going back to your point is that blue ribbon

on top enough for most people? Yes, the answer is probably yes. What GPO has done in terms of your progress on digital authentication is based on what France did enacting a law in 2004, making their official Gazette the electronic version, authenticated, secure and capable of using in a courtroom as the official version. And what France does is actually it has two separate databases for its official Gazette. One is a very user friendly Google-type database and anybody can go and look for information and find it. And they're just looking at -- for information only, they don't need to have that authenticated chain of custody to print off and to take along with the printout of the regulation or the order into the court.

So I think given the fact that you have made progress on your public and private laws and the congressional bills, certainly from our perspective we would like to see you keep up that work, and your top priorities obviously being the electronic legal information. And I'm sorry that nobody is here representing the federal courts, but I'm sure you're having discussions with them, hopefully on this very

topic. Thank you.

MR. HANNAN: Thanks very much. Just one quick comment and then we'll move on, because I think this has been great input and that's what we expected was a lot of different viewpoints based on the diversity of the stakeholders. Is that the comment that Jonathan made earlier is something we are aware of and thinking about is that the needs of the individual user citizen and the needs of large automated intermediary processors information, those can be quite a bit different. So we're trying to think about what's the best achievable, practical way that we can try to help the community in that regard.

With that, if everybody's comfortable, let's move to the next slide. Thanks, that was a lot of really good input. And so that is this slide which is number ten in your handout set. And it's thinking beyond what we've been doing. And we've already talked a little bit about this, you know, information beyond information captured in PDF files.

The digital signature approach has been fairly effective, again for at least some parties in

the ecosystem. You know, in that regard it kind of seems like it may be something to think about in the next types of data or such, so that was just one observation from our point of view.

A question is, you know, is a second assurance level technique like hash based schemes, that may not do as Harlan alluded to, as much as cryptographic digital signature techniques, are those still valuable and should that be another channel that we think about offering or supporting or facilitating somehow for the community. Kind of a complicated question actually and there are a lot of different facets to it. We have a slide that kind of compares some things on the next table. What you get, what do you not get? What do techniques tend to naturally support without a lot of other application business rule processing which might vary from person to person or organization to organization? What techniques kind of lend themselves to certain capabilities that we talked about before?

So let's go to that table on the next slide real quick and then talk through that just a second,

then we'll come back to what do people think about that question of hash based techniques and such.

So this table is, you know, pretty simplistic. So please, there -- we tried to make this something that would be -- something we could at least talk to for today but we basically -- on the far left, the methods we basically, hash based methods using HSA or SHAW-256 (ph) and other kinds of techniques, which provide a digital thumbprint of a particular set of digital bytes. And then there are the digital signature techniques that, you know, we use today in the PDF world.

So across the top row there are some different factors that seem to be important or interesting for our decision making as we try to decide how best to support the community in this regard. So the first one was there's a trusted third party involved somehow and with hash methods, you know, at the end of the day there has to be some other parties involved for trust. Is it only one or multiple third parties? Well at the end of the day there's some other third party involved. You know, could the third party

be GPO? Well, you can -- for hash based schemes it seems like it could probably invent one if we wanted to. I'm not saying whether that's good or bad, there are probably thoughts on both sides of that equation.

The next column is does it use open standards? Again, facilitate the ability of many parties to innovate around it (inaudible) proprietary, obviously that slows innovation (inaudible). Yes, for hashes and yes for digital signatures. So on the first -- those first three areas it seemed like they're pretty much each other technique and, you know, has these features to them.

That fourth column is does the method show the publisher, disseminator of the document -- publisher might not be the exact right term, so forgive me, I'm more on the technology side, that may be a loaded term. But to Harlan's point, hashed based methods themselves don't do that alone, there has to be other business processing rules, information to do that. Does the digital signature provide that capability? It easily can be used for that so it kind of in and of itself provides you a ready-made way of

doing that. So that seemed to be one where the methods kind of were different.

The next one about facilitating a chain of trust, it seemed to us that hash based methods, you know, just by themselves really did, for the same reason they don't really support who the source is, they really don't support, you know, obviously going back further into the supply chain of a digital object, you know, who's been handling it or where did it come from along the way. It's not to say it couldn't, but in and of itself it really doesn't. It has to have a lot of other coding and techniques and business process rules on top of it.

Digital signatures on the other hand seem to have kind of a ready-made way of providing a way to quickly get a sense of that. Not to say that there -- you know, there have to be some understood business rules around the concept of the signature is in this place, that means, you know, first party with the document, here is the second party. There would have to be some rules of engagement. And we talked about that a little bit earlier about, you know, how do we

really know, in a tangible world, who the source of a file was that gets bound into a book. I mean there's -- you know, there would have to be some business rules around any process.

So again, the digital signature seemed to be a little bit more ready-made, not that it's perfect but any means, or solves all the problems. Hashes, not that it couldn't do it, but didn't -- you know, kind of baked in didn't have as much ready-made for that. So a little bit of a difference there.

Free client software available for users, again this is the citizen user. Yeah, it seems like that's pretty easy on both counts. Is this method included in the (inaudible) guidance for electronic authentication? You know, digital signatures, yes. Hash based methods, no. Now that particular publication is really oriented a little bit more to who's the source of something. It's that topic you mentioned earlier, Daniel, about you know, the subtle difference between did I sign, and my signatures on this document versus it's the official one that got bound into a book.

So to be fair, in this publication please Harold jump in and add things if you want to, but that publication is focused on authentication of the source of material, so that's why digital signatures do lend themselves to hash-based, not really mentioned in that particular publication. But those are one of the things that we look at, the GPO tries to take advantage of all the good standards work that comes out of this other agency (inaudible). And we'll talk about that a little bit later for XML too. Lots of other standards bodies, you know (inaudible) standards bodies across the international and the national space.

Is offline validation possible? It can be possible, in the second row there, with cryptographic digital signatures if one's clever about it. It's really not possible, we don't think, with hash-based methods. Now is that important? Well maybe it is, you know, maybe it isn't. If somebody's on an airplane, is it a big deal? Can they wait till they get down and maybe in an airplane that -- being on the Internet would be easy too. There are probably some other scenarios where is this helpful or you know, not.

So anyway, we thought that was kind of an interesting, to us these were some important factors and ways to kind of compare and contrast. So with that I think we'll just kind of open it up there. I'm sure there are lots of other factors and ways to think about, so that's really the purpose of today was to get some input from folks so that we're not just in an echo chamber here at GPO writing these.

I think, did you have a comment Jonathan?

MR. MAYER: I just had a clarifying question.

MR. HANNAN: Okay, sure. We'll go there and then we'll go to Harlan, I think Harlan you had one.

MR. MAYER: So when you're talking offline and validation --

MR. SPEAKER: Can you announce who you are as well, please.

MR. MAYER: This is Jonathan Mayer from Stanford. A clarifying question on offline validation. So are you mentioning here (inaudible) some sort of authentication on top of let's say HDTPS (ph)? Is that the idea? I'm not following how offline validation of a hash itself is not possible.

MR. HANNAN: Well I think -- that's a good point too. The concept would be that you have a way to offline valid the content itself. Now if you pre-stored all of that, which is essentially what happens in the techniques for cryptographic digital signature, at the time of signing a lot of information is put together to make it easy to validate without having to be connected to the Internet or any network. Yeah, you probably could do that with hashes if you did all of that ahead of time, for the purpose of saying, is that document the document that was on the URL, you know, two months ago. So I don't know if that answered your question. It looked like it didn't, but --

MR. MAYER: I guess I may be recalling -- mis- recalling my computer security class from undergrad. But some cryptographic digital signature techniques rely on taking a hash and then sign it.

MR. HANNAN: Almost every cryptographic digital signature technique does.

MR. MAYER: Right. So that's why --

MR. HANNAN: It builds upon that.

MR. MAYER: Right. So the same validation

technique you would need to check a hash is -- or I should say, the technique you need to check a hash is a subset of what you need to check a digital signature.

MR. HANNAN: Well, this kind of goes back to the concept of really knowing the source of the document, not -- you know, that's what we mean by validation is knowing both that the document wasn't altered but that you know the source of it. So that's really -- inherent in validation is the concept that you know the source. And with hash-based techniques you just -- you don't know the source. That's why it's a building block, of course of cryptographic digital signature techniques. It doesn't have to be, but for efficiencies purposes it almost always is (inaudible). That's just really for computational efficiency (inaudible).

MR. BOOTH: You may want to use the term "verification" instead.

MR. HANNAN: Okay.

MR. BOOTH: That's the more common used term.

MR. HANNAN: Okay. That's good. Okay. Yeah, no that's -- I think that's one of the

challenges. There really isn't a lot of -- you know, the standard terminology is tricky sometimes even for somebody in my space. So but that's great, we'll use that from now on. That's good. Thank you. And Harold -- I mean Harlan, I think you had one.

MR. YU: I had the exact same question as Jon. It seems like if cryptographic signature on the right hand column is yes then that should also be yes or an N/A if that's the wrong question to ask.

MR. HANNAN: We'll try to clarify that. And that's what the purpose of today was to get some input and not trying to say this is the be all and end all, but that's really good. That was the differentiating factor was the source of the (inaudible).

MR. SHULER: John Shuler, University of Illinois in Chicago. I want to ask it from another direction. How many users of GPO Access or potentially FDsys actually expect this level of authentication or verification from the stuff they draw down from your website?

MR. HANNAN: That's a great question. I think (inaudible).

(Multiple comments.)

MR. DAVIS: Rick Davis, GPO. I think when we began the authentication effort it was primarily at the request of the library community, particularly AALL. What's been interesting is I monitor our customer relationship management software on a daily basis to sort of get the pulse of the American citizen when they're coming in with questions. I don't think we would have gotten questions five years ago about whether or not this document is authentic, they just want the information. What we're finding now is not only from librarians, not only from researchers, not only from historians but citizens, people writing essays as 16 year olds, we're getting questions about what source should I cite as official. And I think it's an ongoing educational process to realize that when you're looking for something the first result that comes up in a search engine like Google is not necessarily your source that you want to put in your research paper.

So I think there was an early learning curve. I think it's getting better but I don't think it's --

you know, 40 million users of GPO Access but I think it's growing and I think where it is five years from now is going to be different from where it is today.

MR. SHULER: But isn't there a difference between being official and being authenticated or being verified?

MR. DAVIS: There is an -- several definitions of that term. I won't go into deep detail on this with the audience, but authentic from the standpoint that we want to ensure that the information is unchanged from the time it left the Government Printing Office, official from the standpoint of we have, I mentioned early on in my remarks that a lot of this content is born digital now with no print equivalent. And in some cases, like with the Office of the Federal Register, they're recognizing that the online content made available through GPO Access and now FDsys is the official publication.

I think we're going to be seeing more and more of this, particularly where you have no print equivalent.

MR. BENNETT: One thing that may sound

strange that isn't in the line is the idea of authentic meaning that you can find it at an official place. And so for example, you have a paper copy of something, you know, to print something out and you give it to someone then you -- digital stuff, the digital certificates and signatures don't show up, all you get is the ribbon or whatever. So I think one of the interesting aspects of authentication is a unique identifier, specifically a URL, that should be printed with everything and that way anyone who gets it -- the printed thing, so you can help authenticate the paper that way.

But the other aspect of this where you can start thinking about providing authentication not from encryptographic means, but through unique identifiers is a document can have multiple sections. You're going to get into a law of diminishing returns if you try and sign each portion of it. But if you have a unique identifier for each portion of it, say a ID attribute so it could have a fragment URL which becomes a unique identifier, as you do that for your documents and that they're human readable, then what you have is a method again of if people do want to have hashes or digital

signatures, they can add that externally, they can do hash trees so that they can, instead of having one document, quote "be authentic," you can have groups of documents be authentic because you've done an external hash that has all that information. And it could be portions of the document that you are authenticating for some purposes and not the entire document.

This is another way to I think draw out what authentication could mean. And the other aspect of it to think about is if you have a URL that points to something and you can go look for it, but it's really also used as a unique identifier, you could have, in your closet, you could warehouse a physical object that has that same unique identifier and you could always take the electronic one over to it and you know, have that citation or a URL and look at the two things.

So again, I think having unique identifiers is a real key to this because you can create fake bills that don't exist. How do you know a fake bill doesn't exist? Well you know, with this signature on it, how do you know who's signature is it? It's a long process, but the more you can think about things as

repositories and authenticating (inaudible).

repositories from using unique identifiers I think might be helpful as an added authentication.

MR. HANNAN: Yeah, it gets very challenging when you start thinking about tangible paper documents (inaudible). Yes?

MR. ANDERSON: Steve Anderson. One of the things that intrigues me about offline validation and maybe that can be encompassed within your comments I suppose, but is considering what the life cycle is, at least in the near term, of a particular document. And one of the things that these state courts are working on, I'm not necessarily going to speak for PACER or anything else like that, but one of the things that I think the judiciary envisions, at least in the next 10 or 15 years is significant e-filing.

One of the things that would go along with that would be the ability to capture something from the U.S. Code, Congressional Record, Federal Register, etcetera. Have that be authenticated for the life cycle of the use of that as an attachment in a case, so that if you are putting that in as part of a memo or a

pleading or a brief or something of that nature, you not only have your parties, you know, official thing being submitted to whatever court system there is, but you are going to be attaching your relevant case document from GPO or, you know, whatever else exists at the time.

You want to make sure that that is going to be authenticatable through the life cycle of that offline and online usage during the e-filing process. So when a judge gets it, you know, on the -- when he's finally taking a look at that file, the case file, that there's something that's going to be authenticated on his computer or her computer. And there are probably other significant uses like that in terms of academia, making sure that you can attach, as an appendix to some type of online publication, whatever report from whatever federal agency there is. So the offline possibility needs to be taken into consideration that there are going to be lots of, you know, attachings and detachings throughout the usage of that particular document, in my view. Thanks.

MR. GEE: Robert Gee, from the Law Library of

Congress. I wanted to note -- excuse me, can you hear me? To dovetail on what Steve was just saying, the authentication issue right now is that there's actually statutory provisions that would have to be amended as it relates to what is considered the legal evidence of the laws of the United States. Right now we have three separate publications that Congress has designated as the official text of the laws for the courts to -- or that constitute the legal evidence of the text of the law of the United States. And one relates to the statutes at large as being the printed version, the other is the United States Code as the (inaudible) law and through the United States Code, and then the treaties. And those three publications right now have only print as being the official source and therefore that is the authentic, official version that can be introduced.

So I think that as -- and I think the courts - - I shouldn't speak for them, but I would suggest that they would be very interested in seeing something that can be completely reliable in a digital format before they are going to be allowing those types of --

those particular publications to be introduced as evidence of what is the law of the land.

MR. HANNAN: Very, very good point. If we get in the GPO help what we do to further that, then I think it's kind of a win/win (inaudible).

MR. HORTON: Bob Horton from the Minnesota Historical Society. Could I ask if we could add one factor to this, as basically any implication for long term preservation? Just talking from someone who's a state archivist I mean what I could sort of guarantee of the state level possibly (inaudible) level as well, that anything that produces a barrier for long term preservation is problematic, extremely problematic. So you may want to separate the functions that authentication may have some short term need as opposed to some long term function, either the -- anything that you do for an authentication rationale may have -- become an impediment for long term preservation and authentication down the line.

MR. HANNAN: Well I think that's a great comment, thank you. It's analogous to, in the same spirit as the comment made earlier about citizen use

versus high speed, high volume automated. Those are, you know, preservation versus, you know, immediate verification, those are part of the different use cases or (inaudible) so they -- yeah, that's the art or the trick, right, is to find, hopefully, something that will help on all fronts or have a small number of things that get brought to bear. Boy, that's a great comment.

MR. HORTON: Plus it argues also for I think maybe the larger context or larger intellectual framework because there's no kind of one solution to this. There is one size fits all is just an impossible dream. So maybe, you know, a different approach for different constituencies, for different functions and over time as well and among different organizations. I mean I think there's a considerable difference between the federal and the state level, for example, the state level in terms of virtually all the functions we're talking about, is a much more decentralized and much more distributed set of responsibilities.

MR. HANNAN: Great. No, that's good. And we -- or that's one of the real purposes of today is to

really start to get at some of the thought processes, you know, like the question of is a second hash-based method something we should be (inaudible). So other questions, comments? I think at this point it's --

(Multiple comments.)

MS. DALECKY: Yeah, Selene Dalecky, GPO. I just wanted to add a follow up question on the approach that Daniel was talking about earlier. In terms of tying the unique identifier into the URL, is there a question of sustainability there with the transient nature of URLs? And how is that accomplished from a long term perspective?

MR. BENNETT: I think that that gets to a big issue which is content management systems and producing things with URLs. But one of the things that, you know, Europe is dealing with is, you know, how do we deal with this and have -- we're looking at URNs because they're more abstract, etcetera. But I really think that URLs are unique identifiers that have the special, it's like photons, they both are waves and particles at the same time, URLs are both unique identifiers and location references. And so you have

to understand that they have those two capabilities. And what we want is to recognize that over time they'll still be a unique identifier. If something changes you can still find them, say in other things, those references. So you can use it as a metadata then you can still find it.

So if you, as the federal government, decided that you would create a URL that was, for a bill for example, I think that's the easiest way to (inaudible) example, you know, 111th Congress, HR1, right? And you created a URL that contained that, can you ever imagine creating -- using that URL for anything else? And so the question is do you ever do it.

And in the future can -- aren't we at the point where if we have other things we can always use redirects or point to the older version or, you know, use a way back machine, getting it archived so even if the website goes down someone else will then be able to use that unique identifier to find it and you know, they are taking slices of the web at certain times. And so there are different ways to get at it but I think that it does mean that to avoid those types of

complications an organization, I think for the long term should really think about how they structure the URLs to try and make them as technology neutral as they can.

MR. HANNAN: Other questions? Yes, Harold.

MR. BOOTH: Yes, Harold Booth. I don't know if here in your space you have this need but if you have URLs you have the necessity of having confidential proprietary information within your space. URLs might be problematic in that thing, because you may not have things that are publically available to everybody in order to do the verification or validation process. So just to keep that in mind, I don't know if that's a problem in your space, but just to put that out there.

MR. HANNAN: No, thanks I think that's the purpose of today's workshop is to have a dialogue. The concept of using URLs, you know, as the unique identifier, interesting idea that we can see has some things to be worked out as well. But it is (inaudible).

MR. MAYER: I just wanted to add on this question of secondary techniques, something to kind of

keep in the back of the mind and we can discuss further as we go along. The threat model is particularly important here. If we're concerned about errors in transmission and storage and so on, hashes look a lot better (inaudible). But as we start to move maybe towards like the malicious klutz or towards the just purely malicious actor then of course we need signatures. So depending on the users needs and depending on kind of the application, you could imagine wanting the hash-based to protect against certain things or signature protect against more things.

MR. HANNAN: That's a great way of thinking about it. It looks like we're kind of at the point where we probably want to take a break at this point is what I was thinking. So if we can, it's about 10:35. Why don't we convene at 10:50. That will give us a 15 minute break. And then we'll keep going.

MR. BRIEBE: The clocks about five minutes off, so it's closer to 22 right now.

MR. HANNAN: Okay.

MR. BRIEBE: Just to be aware of it, if we're going to take 15 minutes.

MR. HANNAN: All right, I thought my watch was fast, but it's slow.

Okay, why don't we come back at five till. Thank you.

(Off record.)

MR. HANNAN: Okay, everybody we're going to start back up. We took a little bit longer break, hoping that that was useful.

I think at this point what we'd like to do is turn it over to Selene Dalecky of GPO to kind of lead us into the next discussion topic, trying to synthesize a little bit what we all discussed last hour or so and how it relates to two forms of authentication for PDF documents and whether that's something that communities need or just want. So Selene, do you want to start that out?

MS. DALECKY: This is Selene Dalecky from GPO. I just -- there was a lot of good discussion early on on techniques and when and how to apply them, in terms of the electronic content. And I think some of the things that came out that we would like to explore a little bit more are, you know, the technologies

(inaudible) but also when would be the right time to use them and what the needs are in terms of the use cases for this content that we see, at least from the GPO side.

And the two cases that we're looking at is direct to the end user. So there is a person who is looking for bill or looking for a law and needs to know that that is the authentic version, the official version, something that they can rely upon either in their research or actually in a court of law. The other would be data that's used for -- that's repurposed for other means, so system to system where we have, you know, a scraping or a bulk download from data.gov or from our bulk directory of XML or other content to be used either in total or to be mashed up with other content. What are the needs there in terms of conveying authenticity, conveying integrity to keep that integrity and authenticity chain going past the system to system delivery.

I think that another good area to think about in this part is I think this is direct to the end user but the use case that Mary Alice brought up in France,

where there's a system that's for information only that's more usable, user friendly, searchable, accessible. But then you always have the link back to the official version that can be used appropriately in serious research or in, you know, a court of law.

In other similar cases what we are doing with a project called the FR 2.0, it's being done in conjunction with the Office of the Federal Register and where we are taking the XML content that's been converted here at GPO from the SGML (ph) that's used to print the Federal Register and using that to power a search and a Web 2.0 interface to the Federal Register to make it more user friendly to draw out and spotlight the content that's pretty dense within a printed publication itself or within the PDF version of the printed publication.

But whenever you see the FR 2.0 results in a section or an article that's presented, you will have a link back to the signed PDF version since that has been deemed the official version of the Federal Register, by the UFR and NARA.

So we look at these different use cases.

Who's going to be using the content and for what purpose is this content going to be used? Then the question becomes what are the authentication needs for those -- for that content and for that user? How does it differ and is this something that we can decide upon or that we can kind of put a path forward on and develop needs around so that we can look at the technological solutions to these approaches.

And John, I didn't know if you had anything to add or if anybody kind of had some initial thoughts on this and you could start the discussion around getting use cases of who is getting the content and for what purposes and how does that drive the authentication and integrity needs to -- that information on the integrity and authentication that needs to be disseminated to the end user of this particular content.

MR. HANNAN: Thanks, Selene. I think that frames it pretty well. Let me just try to add one thing to help spark some dialogue and that is we've got -- already talked about, I think it's a very useful viewpoint, of single individual user versus automated

system to system processing. So that's one way to think about to breaking down the user/consumer cases. Probably the first way to do it.

I guess the question is for automated processes are there hurdles in the current authentication models with PDF that prevent -- present hurdles and problems? And if there are what are they? And then is there something in the automated system to system taking information from GPO processing it and making it useful to other parties, is there something in the current techniques that need to be improved upon as we go forward?

MS. DALECKY: And I had one other thing to add, as just kind of a nuance to the second case study of the system A to system B content transfer as opposed to the system A to end user content transfer. There's the system A to system B content transfer and then system B to end user. And the question there is, assuming that GPO would be system A in this case, what is GPO's responsibility to the end user who's using a system in between the originating system and the user of that content?

So again this would come in in case of mashups or in the case of pulling out content from, you know, a data set that GPO puts up on data.gov and repurposing it and presenting it out to users in a different format.

MR. HANNAN: So are there thoughts there or burning desires to help guide us at GPO in this regard? Or if not we'll just move on and go to the next one.

MR. SHULER: John Shuler, University of Illinois at Chicago. If I understand your remarks, I mean are you saying that GPO has an ongoing custodial responsibility for stuff that originates from your boxes, no matter where it goes and who uses it?

MS. DALECKY: That's one of the questions that I'm posing to get feedback on.

MR. SHULER: How does that exist within GPO's traditions of not having any copyright material that they produce and distribute?

MR. HANNAN: Mike, do you want to take that or do you want me to take it? Let me go ahead and give you my own personal view of things as we're -- I mean we're on the technology side, unless Rick or Selene,

you guys want to jump in.

MR. DAVIS: Feel free.

MR. HANNAN: I think the GPO approach is how can we facilitate the authentic use of information that we disseminate for the community and not provide negative consequences to that. So I think what we're really asking for -- and again, it's kind of abstract in some ways, and this is I'm sure the first of a few discussions potentially about this. But how can we, through the use of -- and through making good decisions about how to use some of the technical tools available, how can we facilitate authentic content being assumed downstream by other actors in the system.

MR. WASH: But let's dive into the data.gov example, because it's a good one (inaudible) and to your question, John, about the custodial or the stewardship responsibility. You know, we have the Federal Register and unfortunately Mike White from the Federal Register couldn't make it today, but we have the Federal Register content in FDsys and it's digitally signed. And they have deemed that to be the authentic and official electronic version of the

Federal Register.

MS. DALECKY: Not only electronic, but an official version of the Federal Register.

MR. WASH: Okay. And there's the XML version of that that we've created that is referenced on data.gov. So I think the question that Selene's use case was leading to is the fact that the XML is referenced on data.gov, system B I think in your example right, when it references back to us system A, do we have responsibility for the data.gov representation of that data if it were to be authentic? Is that something that you all would expect GPO to be responsible for in that chain from system A to system B where users may go to system B, data.gov, to get access to that data? Is it our responsibility?

MR. HORTON: Bob Horton, Minnesota Historical Society. Maybe not the right analogy, but I run an archive in the library and all sorts of people come in and use the material there. I take no responsibility for what they do with it. I mean is that similar to what you're saying? I mean you're making this content available for use, interpretation, re-use, mashup. I

don't see how you could hope to manage it any kind of reasonable way.

But it then becomes the responsibility of some, you know, tertiary user, somebody who's gone to that intermediary, used it and then wants to trace it back. But that's his or her own decision to whether, you know, to sue somebody for libel or do some more research or write another book. I just don't see how you could take that on, that responsibility.

MR. DAVIS: This is Rick Davis, GPO. From a policy perspective, I would see it at minimum that if we make that data set available in XML format and it's put on data.gov, it's mashed up by another user or another third party, at minimum there would be an interest in going back to that chain of custody issue, tracing back to where the original data came from. Was that information authentic that I worked with to mashup.

But going back to Mr. Shuler's question, traditionally when GPO has disseminated that information then it's up to the user to make any use of it as they see fit.

MR. HANNAN: This is John Hannan from GPO. I think that's a useful model. I think we kind of, you know, see that really as probably about the only practical way to keep furthering this. But it's interesting, it's an interesting discussion.

MR. SHULER: I just don't see anything inherent in the technique that we're discussing that suggests that GPO's relationship with its products has changed significantly enough to alter the no copyright traditions, if you will.

MR. HANNAN: There's no intent to change any of that. We're simply looking to facilitate things in an electronic world that are useful to the community.

MR. SHULER: When you're saying custodial, I mean when you enter the concept of custodial into the discussion you then enter that idea of control over secondary, tertiary fourth uses and so on.

MR. HANNAN: So, okay Selene did you introduce custodial? Only kidding, because (inaudible). I think it's a good -- again that's the purpose of today's dialogue is to explore different facets of this and to help inform our decision making

as we go along about how to (inaudible) systems to the best (inaudible).

MR. YU: I just want to get a clarification about the semantic meaning of what it means to digitally sign data. So is the assumption that GPO will only sign a data set if it considers it the official version? You know, official version of the Federal Register and only when it's official will it be signed? Or does it mean, you know, we're putting this out unofficially on data.gov, it's still being published by us and it hasn't been modified in transit, will you guys still sign those documents? So I mean an associated question is, you know, if in the data.gov workflow, if somehow instead of the data being hosted by each agency, it was hosted centrally and any time you uploaded a data set to data.gov it automatically got digitally signed in the workflow, would agencies potentially be less willing to put data on there because they would feel like putting something into data.gov would somehow make it more official because it was now signed?

MR. HANNAN: Well I think that's exactly the

purpose of today's dialogue. Those are great questions. And we certainly don't claim to have all the answers at GPO. We want to accept a lot of input and perspectives so that we can help to guide what we do to help make it the best for the community overall.

I think the concept of, you know, a digital signature obviously takes some effort and some time and so when that's applied at GPO it's usually for the purpose to provide something that's more reliable in some ways. But I'm not sure I would state it or go, you know, to phrase (inaudible) your question.

MR. YU: Yeah, because I mean to go to a specific example. I mean I've talked to Mike Wash about this before with the XML version of the Federal Register. And I mean it always seemed to me like because the XML version may or may not be an exact replica of the content in the PDF version, you guys aren't willing to sign it because you're not sure yet. But the signing actually provides, you know, a level of reliability, as you say, even if it's not, you know, the official that, you know, is legally representative of the Federal Register.

MR. HANNAN: Yeah, so I think great comment and we're going to talk a little bit later about if we were to think about how to apply authentication in an XML arena, what are some of your feelings about some of the input, some of the criteria. And then what would that, you know, "mean." Those are policy oriented things that, for the specific case of the Federal Register, obviously the Federal Register has got a lot to say about what that really means. So, yes Harold.

MR. BOOTH: Harold Booth. Just to kind of add on to Harlan's comment. You may want to establish different types of signatures or certificates for different purposes, you know, to establish -- one source semantic meaning and then -- that which would be the official version and then you could have another one just to establish titles, so it doesn't have to be an either or choice on your part by any means. You can still use the same framework that you've been -- technical framework that you've established.

Another thing, as an outsider to your community, one thing I'd like to just mention, you're trying to establish digital means of doing something

you're already doing in the real world. Maybe it would be useful, as you talk about this issue, you talk through how you do it in the real world today in a printed manner and then just kind of say, "Okay, well what's the analogy, or what's the analogous process or step that we would need to do in the digital realm."

MR. HANNAN: Great. That's great input. That's the only way I can, you know, make any sense out of this myself. I totally agree, I think those analogies are really often times they're useful for (inaudible).

MS. RUSSELL: Lisa Russell. I don't think it's working.

MR. HANNAN: Maybe the battery's dead.

MS. RUSSELL: I was using the wrong switch. Lisa Russell, GPO. What we've been doing with authentication to date is that when there's something that we want to authenticate we go to the originating agency and talk to them and say, "Hey, is it okay if we originate your stuff?" And getting at that official issue we've had some agencies where, you know, like we got permission to sign the U.S. Code on FDsys however

in order to do that we had to agree that we're not saying that this is official, that they can use it a court of law. What we're saying with that is that we're putting a signature on it and it indicates to the user that it has not been changed. You know, it's come through the official channels and it has not been changed, but no this is not -- we're not saying this official, that you can take it to court.

So we worked through a lot of those issues. And we've also looked at -- you know, we've looked at do we use different -- you know, we use a blue one in this case and a pink one in this case, those types of things with the signatures and that gets a little bit tricky. And we've also looked at -- you know, people have talked about we want to put numbers on them and say, "This is -- this has official level one, this is level two, this is level three," that gets a little tricky when you start looking at some of the older stuff. Is something that's been on the shelf in a rare books room in a library more official or less official than something that was taken directly from an agency's website which may or may not have been hacked.

So you know, it gets tricky. I think there's an argument there, which is a one and which is a two and when we start putting those numbers on it people start getting a little bit confused.

MR. HANNAN: Yeah, I would just add, this is John from GPO, that I think that's the purpose of today's dialogue is to try to get some feeling about how it might be useful to interpret these kinds of questions going forward, because there isn't a lot to really guide us, quite honestly in this regard, to your question, Harold. That's why I think we want to get the widest range of input we can so that we can make good decisions and kind of use some of these electronic capabilities that are becoming more practical to use.

All right, so everyone's -- let me think about how to get at this again which is we tried this because I think the question that one of the things that we had as desired outcomes was to try to get a feel for was the concept of a hash-based, pick one method that might be lower weight, less processing intensive, is that a useful thing that's needed by parties in the system that we should think about

providing through our electronic dissemination channel. Any thoughts or -- I mean has anybody really got to have that or -- yes, Steven?

MR. ANDERSON: Microphone, please.

MR. HANNAN: Is it on?

MR. ANDERSON: It is on. Steve Anderson, Maryland State Law Library. During the break one of the questions that came up was in a locks type world where you have other duplicates of an original, it seems like having a hash-based method for official copies would be reasonable. What I don't know, I mean getting getting back philosophically is how many original authenticated copies one might want. If one only wants to have one authenticated original that's great and then you use the, you know, digital signature for that. But if the world that's envisioned includes official duplicates, then it seems like that would be very easily handled by a hash-based method.

Of course techies could probably correct me with all of this, and they're more than welcome to.

MR. HANNAN: Well I think -- this is one -- just to jump in, that's really the question isn't it?

It's like if you're going to provide a less robust means why are you doing anything with it? Why isn't it just available? So given that the other channel is a more, you know, robustly authenticated piece of material that could be used.

I mean I think that's really the interest question that we're trying to get a hand on, because clearly having multiple methods costs more, is more complex, more things can go wrong. So if the concept is that there's a light weight method that brings value, you know, why would it really be needed if the file is there already to be used and you could rely on a more robustly authenticated available electronic version if some question came up.

It's just a question, I don't -- I'm not saying right or wrong or anything, but that's maybe something that you have some input on.

MS. DALECKY: Selene Dalecky, from GPO. And I think one thing to look at in terms of what methods are needed, I guess how much security, how much comfort do you need with a data set. Does that determine the method as well? For example, if we're looking at a

data set that is unofficial, that is something that people are going to be reusing and mixing with other content or taking portions of for you know, informational purposes only, not for specific legal purposes or it's not going to be cited, is that something where just a dump where the hash has taken place and you can show that when it's transferred from this system to the next system? Or if you've pulled out the data set you know that it's been securely maintained in a repository that has security controls around it, is that enough as opposed to you need to take any file somewhere and cite to it and point to it as the official version that you need to use for some matter of recourse?

MR. YU: Yeah, I mean I think you know, hash-based methods obviously don't get you everything, but you know, as we said earlier, you can piggyback on existing security mechanisms. So you know, if you publish the hash on a secure website, on an SSL https website then you know, somebody could take that hash, know it's actually from the GPO and then do the integrity check with the downloaded data set that they

have.

I think the tradeoff is, you know, when you do cryptographic digital signatures you might have a system that's sort of all in one, possibly an easy user interface, whereas if you sort of compose these various pieces together to create the system you want, you just don't have easy tools for people to, you know, do the checks and to do the verification. It's possible, it's just not easy.

MR. HANNAN: Right. I think that is the challenge, that's right.

MR. MAYER: Jonathan Mayer from Stanford. I guess I would also add to the extent that with some of these algorithms, not all of these algorithms a hash-based method would be subsumed in a cryptographic digital signature, then you can facilitate some user choice there. So if the user doesn't really care too much to actually go to the computational trouble of checking the signature then they can only check the hash. And if the signature's important they can take that additional step.

MR. HANNAN: How do others feel in the room

about the usefulness of that second kind of a channel? Useful, make use of it, don't care?

MR. SHULER: John Shuler, University of Illinois at Chicago.

MR. HANNAN: Yes, John.

MR. SHULER: Speaking from the users that I've been dealing with for the last few years, I would say the authenticity comes from situational rather than technical. They're coming to an official depository, supposedly they're talking to an authentic librarian who's working within that depository. The items have associated with them all these bits and pieces that say, these quote/unquote are "official publications" both digital and paper. And for 80 percent of the people that I deal with, it's good enough. They don't need to go back and do the due diligence, they don't need to go back and assure custodial chains of responsibility.

So the question I would ask you, and this is a culinary to my earlier question, what's driving your question? Who are you trying to serve by asking the question? I'm reminded of the old phrase, "when you

hammer every problem's a nail," what's the problem that you're trying to get at? If you're talking about the users I deal with, hash marks, digital signatures not so much, really don't care. But if you're talking about the legal community, if you're talking about other people that have a much more intimate like with where this information came from, I would say it's a huge concern.

It's about where does the majority of your business lie which be the question.

MR. HANNAN: Right. Yeah, no that's great. Thanks. Can we get a microphone?

MS. SEARS: Suzanne Sears, University of North Texas. I would agree with John to a point that yes, most of the users that come into my library, they're not asking me, "Is this authentic," because they're making the assumption it is. When they come into the library they're assuming that as a librarian I've already vetted that source and that I'm giving them a reliable source.

So I think that, you know, yes we need to have it authenticated on your end so that when we're

purchasing -- I was talking to Steve earlier, you know, just when we purchased a book, if we purchase an Encyclopedia of Political Science we're going to make sure that its from a reliable source, a reliable publisher, somebody who has authority in the field. We're going to do the same thing with the depository material. If we know that chain of custody is really important to us, is it coming from the federal agency, is it authentic, before we present it to our users.

And I do think that we've built up the trust in the community that they're not asking that question because they're already assuming that if we're giving it to them it is official.

MR. HANNAN: Thank you. That's really good info.

MS. BAISH: Although in 2010 people aren't always -- you can turn off that one. Well this is 2010 and we're all struggling to keep our print library collections because more and more people are accessing information coming from FDsys and GPO Access, going to the Department of Labor website, going to the Senate website. And so I mean in terms of your authentication

and the mission of this agency, that's where AALL sees your importance and totally supportive of your description of what FDsys is.

In other words, you are as you did in the print world, the agency for informing the nation and insuring the permanent public access and the authentication and the preservation down the road. The challenges of course for you are we're now, in a completely decentralized world of federal government information and how do you redefine your role. And we're looking to you actually to do that and solve these problems in some way.

I think that the discussion about, you know, is there any authentication that is totally portable, verifiable that moves within that document is -- should be the ultimate goal of the agency. But obviously we're restricted to current technologies.

MR. HANNAN: Thanks, that's great. Well I think that's something that may be kind of loops back to what Harlan said earlier about -- and this is very much on our minds, is that the decisions we take, we don't want to take decisions that have incentives for

there to be less of the workflow involved, in terms of both agencies using our capabilities and services or other data.gov resources. We want to take decisions that help facilitate the movement in an authentic way of these electronic materials.

So that's an important concept, I think, for us at GPO. And so this workshop is a way to try to get at will certain techniques hinder, set up incentives that won't have as much information being facilitated with parties that we've always done business with. So I think that's a great -- that's something that's really important to us.

Other comments? Thank you.

MR. HORTON: Just briefly, Bob Horton from the Minnesota Historical Society. I wanted to echo what John was saying, maybe emphasize it a bit more. I'll give an example. Years ago Minnesota published or passed a law on establishing digital signatures and certificate authorities. And I happened to be sitting next to the ABA representative at the committee hearing and he muttered to me that, you know, this is too much for the little things and not enough for the big

things.

And I think that's something you should keep in mind is that maybe 80 percent of the people don't care all about this, and the people who really care will probably be able, you know, like O.J. Simpson's lawyers in the D.N.A. trial, blow a hole through whatever you do because it's not as simple as attaching a signature to a single document, you've got a whole variety of business rules and connections and partnerships and relationships that you're going to have to document and define if anybody asks any really probing and substantive questions. So, you know, maybe this is a -- maybe that makes it a little bit more complicated, but I do think whoever said -- I believe -- I would say it was Kate, I mean there's a metadata and there's a business rule approach to this that is as important, maybe even more important than the technological wrapper you're using for your transmission.

MR. HANNAN: Yeah, great comment.

MR. MAYER: Jonathan Mayer from Stanford. I wanted to respond on the point about whether users care

or not. I guess I want to offer a brief comment, kind of in defense of the user. And that is I don't -- can't speak to every user of course but I think there's an important role here in that there is a chain of trust. Right? So if you use, let's say Google Patents, you're trusting in Google and you're trusting Google got their material from the Patent Office. I think we're pretty comfortable with trusting those kind of intermediaries every day. So to the extent a consumer is looking at a mashup of GPO data I think they do have some measure of confidence, let's say whether looking at a mashup made by the (inaudible) Foundation instead of you know, something like ThisIsAMadeUpMashup.com. They have some means of judging, it's not perfect but I think the situation's a little bit rosier than had been suggested.

MR. HANNAN: Thanks. Anybody else from GPO? Any other thoughts? Thanks, this was really good discussions, great viewpoints. Thanks for sharing that.

So let's go ahead and -- we're probably right about lunchtime honestly. So I think what we --

MS. DALECKY: I did just want to kind of close up with one more --

MR. HANNAN: Oh, please. Yeah, sure.

MS. DALECKY: -- just to kind of bring back what we were just talking about, in terms of overkill versus not enough. In the case of FDsys now or when we're bringing in more content from various sources, if we have a documented, secure content management system that follows best practice guidelines that can meet audit standards, the content that we have is secure. But if we can also show from an input process that when we are actually bringing content into the system, we've secured that channel as well and we know who's putting in the content and we know where that content is coming from originally, so it's a deposited content from a born digital content from an agency, or it's a converted file that somebody from the Library of Congress is providing us, so we've secured those channels and we can show that, we can demonstrate that through our audits and our documentation, then the question becomes what level does that need to be conveyed to users. And I think that's the 80 percent

don't care, it's enough to know it's coming from GPO. And the 20 percent, well it needs to be very clear all that you've done, you know, succinctly showing any signature or something that we can then present to somebody who needs to know this information.

And then how much needs to be beyond that? How much is knowing where it's coming from and how much is the content itself? How does it -- does it have to persist with the content in the form of a signature on a PDF or can it be enough to know that this content came from a source that can prove that when you got it from our source it has been maintained and the chain is there and it hasn't been broken.

And I think in terms of looking at system A, system B to the end user or some of these other use cases that that really boils it down. Is how much of the security is knowing where it came from and being able to present where it came from, and how much of it is the content itself persisting with that content being handed down from the system to person to person.

MR. HANNAN: So is that a question to the audience, Selene?

MS. DALECKY: Yeah, I think it's just kind of something to think about and say, "Well what is most important and in what case?" Again, is it enough and 80 percent of the time, just to know where it came from originally and that's enough, or do we have to persist the authentication of the content itself and make sure that everybody knows that that content is still good and hasn't been altered you know, time and time and time again as it passes on from system to system or person to person.

MR. HANNAN: I don't think we can answer our question (inaudible). Lisa, I'm just kidding.

MS. RUSSELL: Lisa Russell, GPO. Just sort of a related thing. I think we've been talking about different users needing -- wanting different levels of authentication. I think sometimes the same user wants different things in the same situation -- in different situations, because once we started signing the bills we started getting questions from users who were saying, "Hey, I used to be able to download your file and make notes on it for my own use. I can't do that anymore. What's going on here? Can you fix this for

me?"

And there is actually a way where they can save a copy that does not have the signature on that then they can make it, but in that case, you know, when they're getting from us they want that assurance that this is good, this is the correct version, but then they want to be able to use it for their own purposes and tweak it and so on. But then we also want to make sure that they're not going to be able to pass that off as an official version. So I think there are some other issues there maybe that the same user might want different things in different situations.

MR. HANNAN: That's a great point. I would tend to say it's pretty clear that, you know, if there was no cost and it was an ideal world, obviously we would have, you know these multiple channels. And I think that we're going to work hard to think about, you know, what is the marginal effort and complexity of doing that. Based on the discussion here it sounds like there might be some interesting value to come out of it.

I think the concept of, you know, mashups is

an interesting one that's probably beyond the scope of what we're doing today. But we'll continue that conversation on that once we learn how to walk.

So I don't want to cut anyone off. Does anyone else want to chime in or address what Selene talked about? If not I think this is a good time to break for lunch and then come back at one o'clock and we'll keep plowing through this material. And I hope this is useful. Does that sound good to everyone? I think this is a good time to hit the cafeteria before it gets really crowded.

So with that we'll break till one o'clock. Thanks.

(Off the record)

MR. HANNAN: All right. I think we're all back. Thanks. I hope everybody got something decent for lunch. And to kick off, to get back into this exciting topic -- so what we'll do at this point, Ted's making fun of my usual --

SPEAKER: I do like that. From lunch, exciting topic.

MR. HANNAN: Yes. Let's go into some

questions and some areas that we want to get your feedback on related to authentication of things beyond the files, which is what we'll do today. So on slide -- on 12 in your packet, you'll see a little bit of observations from our end. Thanks, Jessica. We went ahead and just got going. You haven't missed anything. So this slide is observations.

So and one of the things that has been brought up by those is -- well, for content types or BETA formats other than PDF, maybe we could use PDF as a carrier, which it certainly is technically capable of doing, to encapsulate other file types, like XML, text files, ASCII text, whatever it is.

The disadvantages for that, there are some disadvantages, which is that those embedded files don't have direct authentication, they kind of inherit the authentication of the PDF carrier. And how do you link all that standard information with the authentication of the overall file to the [inaudible] file as well? There's not a lot of standard ways of doing that yet. A lot of people have ideas, but they're just -- so my conclusion right now is probably the best option is

we're examining other options. That's what I'd like to get your input on.

So when we think about all the -- high volume, system-to-system processing applications, we inevitably, it seems like, start talking about XML data or, potentially, binary data. So there are a couple of just high-level options, just from the technology side about thinking about this. Both of these formats, the first one, which is XML, and the second one, which is the traditional PKCS7 wording and concept, a lot of you have unambiguous digital representation of some data.

And so I guess we'd like to see if you guys have any feedback either way, realizing this is a very technical question. It may be something that you just say, let's go to the next slide. But it seems, to us I guess it all makes sense because we're in such an XML-oriented world in many respects, and there's been things in GPO with respect to the Federal Register where XML is already a type of final that we're disseminating.

So it seems like XML's the right thing, but we wanted to get your feedback too, so that's really

how we're going to open it up to the audience at this point. Unless, Ric, did you want to jump in?

MR. DAVIS: Yeah. Ric Davis, GPO. I don't think this one's working right now. Ric Davis, GPO. One thing to consider on that is, in addition to the signing process of native files, is also the net effect on the end user. Because I think we talked about this at the beginning of the discussion. One of the reasons we chose the process that we did was not to require for the end user difficult third-party plug-ins to be able to make use of our authentication processes.

And as John mentioned, right now, you know, we can take native file formats and put a PDF wrapper around them and make those available. But, ideally, what we've heard through use cases is users want the ability -- us to have the ability to sign those native files. But on the other end, they want the same ease of use that they have with the PDF to be able to do validation.

And, again, I think this is where, inevitably, we talk about like the two basic ways to segment the users who may be interested in this data

for GPO, individual users, citizens and then large intermediaries or system-to-system processors. So any feedback at this point? Mike --

MR. GALLUCCHIO: This is Kevin Gallucchio from Department of Defense. So you've asked the paranoid guy again. And I would caution you against using PDFs because they have a lot of kind of nasty things in them, like hyperlinks, and you could run -- include executables and all sorts of attachments that you really don't want on your computer.

And in a standards -- from a standard developing point of view, the PDF kind of attachments is really kind of an afterthought functionality that they added after the specification was built. So what is XML, is kind of a much better way of supporting extra data and files and things like that. So I would say XML or even the MIME attachments or things like that would be much better. You can write applications that handle these -- it's trivial because you have these packages that can extract -- go to different parts of an XML document, you know, and say, you know what?

Give me this path, give me the element, you know, that's located at this path, and it's very easy to handle for [inaudible] whereas with PDF, you're going to be tied to Adobe a lot, or you know, you've got to develop something specific to PDF, and that's going to be a bit harder, I think, in my opinion. So --

MR. HANNAN: Thanks, that's good. Others at this point? Or we can just keep going along. Okay. Let's go to the next slide and see if this -- so the next set of slides talk about native XML authentication, assuming that that's kind of of interest, because as I say, we're already disseminating some XML, so we are thinking about does it make sense, and if it's valuable, and then how would we think about doing authenticated XMLs, so these next few slides talk a little bit about different ways to do signing and digital signatures with XML.

And there are, basically, three different types, envelope signatures, enveloping and detached signatures. The pros and cons we'll talk a little bit about in the following slide so if we can just go to that.

I did have a graphic of pictures here which I wish I'd had, but maybe we can post something on a web page after this. So an envelope signature is the concept where the signature is embedded inside the file, sot he signature is a child, essentially, of the content or the file itself. It's inside the file. The document or the file needs a placeholder to hold one signature or multiple signatures.

So the advantages are everything's right there in one place. This is, essentially, the PDF concept. Offline validation facilitates that verification, kind of more readily facilitates that or it does facilitate it. The signed and the unsigned content, they end up being kind of in the same format as it turns out. That could have some useful features to the practicality, ease of use downstream.

Enveloping signatures, this is the concept where the signature is really kind of a parent of the content. In other words, the signed content is kind of underneath the signature itself. So when you first go to open the file, your software has to be smart enough to know how to deal with the signature right off the

bat because that's kind of the overarching -- it envelopes the whole thing.

That has pros and cons relative to existing software tools and how cheap they are, how readily available they are, that sort of thing.

Again, the signature and the [inaudible] are very much coupled together. It's just kind of a different way to orient them. But the signed and the unsigned content have different formats. Sometimes that's a little more complicated. Potentially, it means you have to have two viewers, not just one. So the unsigned raw content has got the different format than the signed stuff. Not that that's right or wrong.

Detached signatures, the signature is completely separate from the content. The processing of the signature and the document are separated. There's no difference at all between the signed and unsigned content. They're exactly the same. It's this second step you're doing with another piece of digital data somewhere that will determine what you're going to think about that content.

You know, when you start thinking about mash-

ups of -- separate documents, in this content, I think it actually starts out with some advantages to it in some ways. Disadvantages are that you've got a link. You've got to have a process to link the content and signature together. There really aren't any standard protocols for this yet. I'm sure there will be at some point. Therefore, they tend to be higher costs, it's a lot more coding experiences.

Some of the same disadvantages as you have with the embedded files in PDF come up in this context. And lastly, kind of -- some of our observations in thinking about this was since the signed and the unsigned content are exactly the same, it can be quite easy for users or parties involved in the process to become confused about what really is or isn't actually the content that is important.

MR. WASH: Is there an example of the attached signatures available today?

MR. GALLUCCHIO: That's a great question. I don't really know of any practical systems that are used for anything but I may not just be aware of it.

MR. WASH: So where does surety fit in of

these three examples?

MR. GALLUCCHIO: They're in -- they're not a signature but they're in that class --

MR. WASH: Where the --

MR. GALLUCCHIO: Yeah, it's a detached methodology --

MR. WASH: For the encrypted signatures, this is detached --

MR. GALLUCCHIO: Yeah. Yes, sir.?

MR. YU: This is Harlan for [inaudible]. I don't really think it matters all that much whether it's enveloped or enveloping or separate. I mean, the people who are going to be using X mobile data, they're going to be developers who know how to use command line tools. They can find that signature and apply it to the data as long as there's a specification for what that data is and how it's attached. The W3C has a standard for XML signatures, and maybe we should just use that.

MR. HANNAN: All right. That's great input. Thank you. And that's, I think, an interesting viewpoint. Let's try to go, if we can, let's go

through this which tries to summarize some of our conclusions to this point, and that will be a great jumping off point for just what you've laid out.

So, you know, our current approach uses envelope signatures. In some ways, it seems practical, might be [inaudible] XML world. For the cases in which there are multiple signatures, then this is, by far, the simplest method, but again, we're going to talk later about -- on the other hand, there are things that are applied by -- signatures or, in the case where you do have multiple content signers.

The next item really speaks to that. So when you do consider the case where you have multiple signers, you want to try to use a way to do chain of custody using these kinds of authentication techniques. There are other ways of doing that too, as in meta data.

But when you're doing it just with these techniques alone, trying to convey chains of custody, an envelope signature approach requires a common specification. You absolutely must have a common specification for the document and try to find some

reasonable upper limit about how many hands it's going to go through.

The -- XML requires a lot of coordination between GPO and content originators anyway as it is. We already coordinate quite a bit. Maybe that's not such a big issue, but it certainly is applied by the envelope signature process.

When you start thinking about chain of custody, it seems like it's more complex to use enveloping or the detached signature methods, more complex for the protocols and the software involved. So maybe we should start with envelope. If we're going to do this and it's of value, we can start with envelope and see how it goes.

That's kind of conclusions to this point. So we want to open it up to you. I realize this is pretty technical -- but to see what you're thinking is, do you care one way or the other, do we have -- for those, especially those that are really interested in, you know, straight-through XML processing, for instance, and you know, high-volume, hands-off, no human involved in it orientation. So let's start with you. Do we

need the mike?

MR. ANDERSON: That's fine. I think everybody can hear me.

MR. HANNAN: Okay, yeah.

MR. ANDERSON: Steve Anderson from the Maryland State Law Library. Your comment earlier that GPO is not a standards body would seem to preclude the usage of the -- which one was it -- the detached one where there was no formal standard to govern the detachment or reattachment process.

Simply because in that scenario, then, you would have to be developing -- you would have to develop a standard to do that. And if you don't want to be a standards-creating body, then don't use that method.

MR. HANNAN: I think that's a great comment, I think, and we appreciate you voicing that.

MR. ANDERSON: Just trying to make it easier on you.

MR. HANNAN: Yeah. I think that's some of what has become apparent to us, but again, we want to try to validate that. So thanks, that's great,

Stephen. Well, John, are you next? Did you want to go? We'll let Harold go then.

MR. BOOTH: Harold Booth. Just to add on -- maybe to re-complicate your life again rather than trying to simplify, but the fact that you're creating XML documents is, in effect, you are creating specifications or standards for those documents. At the end of the day, you are doing that. So when you embed that signature as part of that format, however you do that, you are, in effect, creating that specification standard, however -- the terminology you wish to use.

I will add on that when we're talking about detached and the definition within W3C, having a sibling element signature block is equivalent to detached. It doesn't, necessarily, need to be a separate document, per se. It just needs to be -- so if I have -- and I know I'm getting technical, and I apologize to anyone that's not familiar with XML, that if you have a signature block, it's a sibling to the thing that's being signed, that's considered detached within the vocabulary of the W3C signature standard,

just to be clear.

It doesn't, necessarily, mean that they're separate documents. So if someone was asking for an example, often times with open-source software, they will have a separate cache file that's separate from the executable that you're downloading. That would be another example of detached; that would be two different files that you have to manage and maintain separately.

MR. HANNAN: Okay. Thanks. And we'll update our materials and try to be more precise with that. That's a great comment. So --

MR. MAYER: I want to build upon Harlan's comments. I guess this morning I kind of made the case against the apathetic user. I now kind of want to make the case in favor of the apathetic developer. At base, you need to store some signatures and what they're signing. You can store them at the beginning of a document, you can store them in the document, you can store them in another file.

But from the developer's perspective, they're going to load these in and check them. I think the

real considerations that come to mind are if there's a software library or several good ones available for some of these, they seem to have some widespread support that would tilt in favor. Similarly, if there are particularly bad libraries where it's really hard to manipulate, one of these PDFs come to mind, it's notoriously quite hard to programmatically interface with, then that would be kind of thumbs down on it.

MR. HANNAN: Thanks. That's good info. Others have anything at this point? I realize this is a pretty technical area -- oh, thank you. Kevin?

KEVIN GALLUCCHIO: Kevin Gallucchio from DOD. It's just that the -- I just want to warn you that the enveloped approach is kind of the most radical for -- you won't be able to ramp up, whereas the detached is like the least amount of work. You'll be able to have, like, files that have detached signatures next to files that don't. This is not -- that might be a plus or a minus.

But hardly anything in government starts, you know, as one set. Everything's compliant. Usually we ramp up, at least in the DOD. I don't know, maybe in

your part of the government, it works smooth. But we ramp up, so detached would be the easiest, enveloping second easiest, and enveloped the hardest.

MR. HANNAN: The hardest, because of that having to have the common specification for --

MR. GALLUCCHIO: Yeah.

MR. HANNAN: -- things go. That's good, that's great input, thanks. That's really good. Others at this point, or ready to keep moving on and get through this? So, Harlan, if I -- just one quick comment. If I was going to summarize your input, it doesn't matter as much, just any one of them might be okay.

MR. YU: Yeah. I don't think there are huge technical advantages or disadvantages. It just depends, as John said, the available tools and what people are used to. As long as people know what to expect, I think it's --

MR. HANNAN: Right. Harold, in this, do you have any -- you've already given us some good input on this. Do you have anything that we ought to think about again, those three, which ones we're going to

choose? Given that we know we have lots of straight-through system-to-system users, but also --

MR. BOOTH: I have nothing to add with respect to which one you should choose. I think, ultimately, that's -- again, from the feedback you got, there are pros and cons to each of them again. So it's more, I think from your end, how difficult it is for you to manage and maintain that, I think, is really the decision that you, as an agency, need to make.

The one thing I would caution is, you know, as I'm starting to dig through and implementing this within my own space, there are numerous issues with XML signing. And, again, I'm going to get very technical, and I apologize to anyone that doesn't -- isn't terribly interested in this.

MR. HANNAN: That's okay.

MR. BOOTH: There are -- the W3C has recognized this. There are a series of best practices that -- technical papers they're trying to produce, and there are security issues that, as people need to validate or sign data, you need to be very much aware of because of the very nature of XML signing. And

these issues don't, necessarily, exist in some of the more -- the RFC that you referenced before, the PKCS7s type standards. And I'm not going to enumerate them all. They do exist. I just want to, I think, raise your awareness that they do exist.

MR. HANNAN: Thanks.

MR. BOOTH: And I'm more than happy to -- I can give you a link to all those various resources.

MR. HANNAN: That would be great. I think we -- I would probably take you up on that, so that sounds -- that's great information, so thank you. Okay. Before we leave this one, real quick, I think to summarize this, it seems to us that, at GPO so far, that if we decide to try to offer authentication in a native format, besides what we're doing today, it probably would be either native XML, or in the MIME oriented, you know, PKCS7 world.

Again, those are both definitive, binary objects. And WC3 seems to be the set of standards on that. I don't think there are really any others in native XML at this point, but that was something I did want to ask Harold, are there any other standards

besides --

MR. BOOTH: Not widely used. There are some that have been used by some niche communities, but for the most part, WC3 standard is considered the XML standard. But there are others.

MR. HANNAN: Okay, right. That's the great thing about standards, right? There are so many to choose from, as somebody once said. But thank you. Because that kind of what our summation too. That seemed like the standard for native XML. I appreciate your feedback.

MR. BOOTH: I'll also add that there are many libraries -- not many -- there are libraries out there that already have implemented the WC3 standard as well.

MR. HANNAN: Got you. Super. So, to summarize, we're not sure what to do yet, obviously. That was one of the purposes of, as we try to figure that out, to have this conference and workshop, get some feedback. It's been very helpful. I think this will help us formulate how best to think about moving forward.

So now, we're going to move and switch gears

a little bit into chain of custody. Use cases, we wanted to use this as an opportunity to gather some feedback from you on this concept. So, Ric, if you have specific things you want to jump in and add into this, please do so. That would be great.

Basically, this is about an indicator of the originator of the content, in other words, getting to the provenance of the material. There are a lot of different ways to think about doing that, some of which these technical techniques can facilitate for you. But we talked earlier today, I think Kate mentioned it from GPO, about meta data being just a source of this information that's part of the larger signed or authenticated content.

But with that, one of the things we wanted to note is that this, from a policy perspective, we can't dictate this at GPO. It requires the content originators to be on board. It's really an education process for those agencies, seeing the value in it. We're communicating on this at GPO with all of our stakeholder agencies and partners, but we just wanted to make sure you all realize it's not -- we can't

dictate in this arena, but we'd like to facilitate this because, I think, a lot of the parties would like to have this.

MR. DAVIS: Ric Davis, GPO. And I think part of that, building upon what Kate mentioned this morning on FDsys is making it easy on content originators or creators of information to digitally sign content and establish chain of custody from the beginning of the information life cycle.

It's interesting, when you look at what has been digitally signed, fist on GPO access and now on FDsys in beta format, one of the first things that we signed were federal budget files from the Office of Management and Budget. And I think as Reynold mentioned this morning, those had full chain of custody. We issued a digital cert to them, and they made use of it at the beginning of the process.

There is an educational process that we're doing with other content originators, but the piece of feedback that we most often receive is, you know, if they want to understand first why is it important, why does anyone care, and secondly, make it easy to do.

MR. HANNAN: So I think -- this is John in GPO. I think one of the things that might be helpful if this is really an important facet of this topic for you is to kind of assist in that education process as you're chatting with people, that may reinforce that concept. I think a lot of agencies are receptive but, you know, everyone's got a lot to do, extra things and changes to process are always something that people don't take lightly, they have to be very careful about. So let's go on to the next one.

Here are some factors that we just wanted to throw out to stimulate some discussion, maybe. Ability to view the entire chain of custody. Having that with the document, that seems important, we just wanted to validate that with you. Open standards, again, seem like the way to go if they exist. They do for some of the more fundamental mathematical techniques.

The chain of custody being completely present, preserved with the document, that's a little bit -- the first comment, but is that a valuable thing to you. And for XML and PDF files, both of these other methods can allow the full chain of custody to travel

with the document. Having embedded the meta data, clearly, would allow it to -- if the meta data is, travels with the file in question, then it, obviously, provides that holistic capability too, which seems simple in the best case.

Were there other-use cases or other aspects of this that you all might have that we can think about beyond these items here, or do you have some feedback? Maybe some of these things aren't important as -- to worry about that. So we thought we'd ask, open it up and try to stimulate a little discussion on this topic, on this end if folks have thoughts. Yes, Suzanne. Can we get the microphone back? Thanks.

MS. SEARS: Suzanne Sears, University of North Texas. Chain of custody is important to us. And it being in the meta data would be fine with us as well. But it is definitely something that we want to see with the document.

MR. HANNAN: Yes. Harold?

MR. BOOTH: Harold -- I think I can probably streamline [inaudible]. Someone did say that the chain of custody is important to them, and you said -- the

reason why I wanted to speak up is you said meta data. By that, do you mean that you just wanted to know that it went from one person, to the next, to the next and just I'm asserting that, I'm attesting that? Or do you want a non-repeat -- such that I've signed it, I hand it to somebody else, they sign it, and then I hand it to somebody else and they sign it?

It's very different than just putting meta document into the document and saying it went from this person to this person, to saying I've signed it, and someone else signs it, and then party C signs it, and no one can say, I never signed it. So I just wanted to make sure that you understand that there is different technologies that come into play depending on which way you want to go.

MR. HANNAN: That's a very interesting point. And so, yeah, how do people feel about that?

MR. DAVIS: Ric Davis, GPO. I can tell you, from a policy standpoint, the easy answer is we want it all. Because associated with authentication, you know, we want confidentiality, authenticity, integrity, non-repudiation, all four elements.

MR. BOOTH: Right.

MR. HANNAN: And so that starts to drive you into the digital signature, only because it is harder to deny a part in the process, if you will, not that folks want to do that, but again, pros and cons. There's not really a right answer. Selene?

MS. DALECKY: Selene Dalecky, GPO. Where this also gets a little tricky and comes into play is earlier when we were talking about from a signing, you know, affixing a digital signature to a government publication that we need to distribute, does it go to -- - you know, are there a hundred signers from an agency that's a person's name who is signing this document that's going to the public, or is it a role, like superintendent of documents?

Or is it, you know, one person who is -- assumes the authority of signing, but really, it's just kind of in name only, and what happens when that person leaves? Does it -- you know, does the weight of the signature get carried with you indefinitely because you affixed it to a publication that's now into the public? So from a non-repudiation standpoint, is that really

important in distributing the electronic government documents, or is that, again, a little bit of an overkill?

And then in what cases is it important? I mean, I'm sure that when we look at, actually, the first receipt of the content into FDsys, for example, from agency customers, that would be the point where we would want to know that the person who has the authority to give us the final version of this document for distribution has signed off on it, be it the printing officer or whoever it's assigned to.

I know that's the case over with the Office of the Federal Register with their eDOCS system. Agencies are now sending their public -- their notices and their articles that go into the Federal Register to OFR by email that's been signed so that they have that non-repudiation. In that case, it's important.

So I think the strong distinction of where in the chain of custody those various levels are needed is important.

MR. HANNAN: Yeah, I think -- this is John Hannan, just to add a little bit because this is a

nuanced point that most of you folks probably haven't really been thinking about too much, or maybe you have, but just hopefully to be helpful.

It is interesting about the concept of a person's name versus the role that they fulfill at an agency, especially with respect to perpetuity and preservation. And there are different ways of coming at this. When you have an individual signing a paper contract or paper agreement, often times, it's not only their name but their role showing they have the authority to bind the agency or the organization to this agreement that's being fulfilled.

So -- and you can kind of see how we can come at it at the GPO already in terms of it's not a person's name, it's the role of the superintendent of documents, because that's such a longstanding identified rule. So that's something that is important to think about a little bit as a community as we go through this, and there are pros and cons both ways.

Some of the drivers that would say, looking at role-based certificates, which require, you know, again, more handling that goes around them at the

beginning before they're substantiated, it can have some payoff down the road, as opposed to 30 years later looking at something and saying, oh, you know, John Hannan signed something. Which if, as long as there's a lot of binding of my title in my personal certificate, that can be -- that can overcome that.

But oftentimes, you don't want to churn through if I change, you know, jobs within an organization, or have to churn through different certificates, so just interesting. I thought maybe that would be helpful to you, so probably you have --

MR. YU: Yeah, I was hoping to step back just a little bit. I don't feel like I have a good sense of the current state of government's [inaudible] -- so who's the master signer, or who's the central trust authority, and -- who currently has keys, who has ability to sign? How is that sort of managed? I don't really have --

MR. HANNAN: Well, that's a great question, and I have all that information if you'd like to know. So -- and you probably might not be surprised at the end, Harlan, that there is no central authority for the

U.S. Government because nobody could agree on giving -- seeing that to someone. So what it is is it's a hub, it's a hub and -- where there's the federal bridge certification authority, which is essentially a hub, to allow other issuers to meet a common set of policy and procedural requirements, and that's how it's been substantiated at the Federal Government, thanks to NIST largely taking a leadership role in that many years ago.

So that's how it's substantiated today. That's been borne out mostly through the HSPD-12 PIV card programs within the Federal Government Executive Branch side, which is like an initiative that the Federal Government undertook several years ago to provide a standard form of ID card. It also had digital certificates on it as well setting up the potential to be used for various topics like this sort of supply chain orientation potentially.

And there are -- some agencies issue their own from their own PKI. Some do choose to use a provider that GSA, on behalf of the federal PKI, certifies and is linked into that hub. So that's kind

of the high-level answer to your question.

MR. YU: Is there a lot of document -- I'm sure there's a lot of documentation. Is it possible to sort of --

MR. HANNAN: Yes.

MR. YU: -- send it around or put it up somewhere where I could find it?

MR. HANNAN: Yeah, sure. I'll give it to you offline because there's a lot of links that I can give you. And we do, also at GPO, have our own public key infrastructure. We're one of two federal agencies that are authorized to issue to other agencies. Treasury Department is the other one.

So we've gone through all those processes which are not trivial to be able to issue digital certificates, which is one of the reasons that the agency undertook this several years ago was to position ourselves to be ahead of the curve for suppliers at the agencies if we want to start to take more steps with this, but it's also, you know, our requirements are much like -- I think Ric mentioned, or Selene actually mentioned with the Federal Register, which our

viewpoints are looking for issuers that are cross-certified to the federal PKI.

MR. YU: Thanks.

MR. DAVIS: Ric Davis, GPO. As John mentioned early on, you know, we are cross-certified with the federal branch. I think if you go to that website, there is a link to our certificate authority statement, at least right off of there as well.

MR. HANNAN: Great question. Anything else? Yes, Jonathan?

MR. MAYER: I guess I'd like to do a little probing on what exactly the threat model is in these chain-of-custody cases. I mean, certainly [inaudible] and if there are kind of external considerations, whether it's good policy or so on, certainly.

But in terms of threat model, I guess, the vulnerabilities I see are either there is not an actual signature before something got to the GPO, and that somehow kind of slipped through the cracks, or the GPO's key was somehow compromised or the algorhythm they're using to sign was broken, or someone got really, really lucky, like in the cryptological sense

of lucky. So, I guess, from a pure security standpoint, I'm finding it hard to see how it adds that much, if the GPOs already have this terminal signature.

MR. HANNAN: I think that's a great question. That's one of the things we want to try to get some feedback on here is, you know, again, in an ideal world, if adding strong signatures for chain of custody cost zero and was easy to do, of course, you would do it. Well, the reality is it doesn't cost zero, and it may or may not be easy to do.

It will be easy to do based on the decisions that we end up making, as easy as it can be, but that is, actually, the fundamental root of some of the things that we're struggling with is how best to provide this, at least initially, so that it's on a strong note. But you kind of, I think, really crystallized it, Jonathan, so thanks, that's -- that's a good one.

So do others have thoughts about that, like oh, you know, I think Suzanne's already commented on this, and I think that?s great, is that, you know, chain of custody is good enough, and Harold, rightly,

pointed out some of the other -- the classic on-the-other-hand point of view which is, well, it has its limitations too. So I think that's really what we're looking at at GPO is, you know, what's the sense of people -- is a meta data orientation good enough? Don't worry about it for now, look at bringing in other stuff down the road.

And in that regard, if we start off with meta data that proves not to be good enough, are we really fined. So Suzanne, you have -- thanks, Ted.

MS. SEARS: Suzanne Sears, University of North Texas. Going back to what -- going back to what John said earlier, 80 percent of my users, you know, they're going to trust that what's in the library is valid, and they're not going to care about chain of custody. For me, it's important --

MR. HANNAN: Sorry about that. It's --

MS. SEARS: I'll just talk loud. For me, it's just really, really important to know that it is an official agency publication and not just something that John Smith who used to work at, say, the OTA and it's an OTA document, but his own version of this

document. So for me, the meta data is okay, but I'm not a technical person, I'm just the end user who doesn't understand anything that you said. So --

MR. HANNAN: That's okay.

MS. SEARS: I mean, I know my catalog record, which I'm sure you could not mark catalog, but I know which field it goes in my mark catalog record, and when I'm creating meta data for our digital collections, they're based off those catalog records.

And for us to just know that, you know, this is an official publication from the Department of Energy, or this is an official publication from the Government Printing Office, that's the important part that I need to know. So the tech part behind that, I can't really give you comments on that.

I can just tell you as an end user, we do need to know that it is an official publication. And in the printed form, we have that because it says, printed by the Government Printing Office.

MR. HANNAN: That's great. That's super helpful. Thank you. I think that's absolutely important. Mary Alice?

MS. BAISH: I think a distinction needs to be made with what the role of the GPO is in the digital world when we're talking about chain of custody. For example, speaking of the courts, it is the judge or the reporter that verifies, whatever term you want to use, what is going to become the official version, say the Court Opinion, and so much of that chain of custody is going to happen at the publishing agency or the senate or the congress or the individual court, not at GPO, because you're not that official publisher.

So then it's -- at what point in that chain of custody do you all have a role in kicking in and tracking it so that you can then be that third party authenticator?

MR. HANNAN: Thank you. That's, I think, a really great way to personalize it in many respects. Others, before we move on from the chain of custody topic? Hearing silence -- oh, Harold.

MR. BOOTH: Sorry.

MR. HANNAN: That's okay. No, it's great.

MR. BOOTH: She mentioned that you know that it's been printed by the Government Printing Office, so

putting on -- I'm about as paranoid, I think, as Kevin over there. How do you know that it was printed by the Government Printing Office? I could just stand up my own printing press and print off anything I want. That's issue number one.

Issue number two, I think one of the things that's kind of left unsaid a little bit here is that when we're talking about digital documents, publishing, you know, myriads of multiple different types of versions of copies is, actually, relatively cheap and inexpensive relative to how it used to be where I had to get a printing press and, you know, a binding book and do all this stuff. It was a huge investment in money, so only large organizations could do it.

Whereas now, anyone with a computer could just basically publish something and, you know, they could make it look like a formatted GPO document and then send it off to the world and instantly have, you know, literally millions of copies floating around.

MR. HANNAN: Well, I think that's why we're having this workshop.

MR. BOOTH: Right, well, I'm --

MR. HANNAN: But you're right. You hit it right on the head in terms of -- and again, one size fits all? Maybe it isn't -- you know, we're trying to figure it out, I think, as an integrated community. So -- yes, Steve?

MR. ANDERSON: The more I think about this the more I think that it's really an important -- a more important discussion to have than I had originally envisioned, simply because on the one hand, and looking at this through, you know, both the history of the book as well as, you know, what other state agencies' needs are if anything's ever going to be replicated, one of the things is that the chain of custody, essentially, between the first -- between saving the document on somebody's PC and having it be "printed or published by GPO" is probably a much more cumbersome and convoluted process than we think.

It goes through editorial processes or whatever. At what point in time does everybody feel comfortable saying we, you guys, are going to sign off on this document at this point in time? And you're right about threat models and, you know, what's the

risk. And, you know, quite frankly, we could probably afford a little room to be flexible.

But realistically, we need to do two things. One is making sure that we have this stuff locked down so that posterity and judges and legislators are going to say this is the official thing and has been for the past 70 years. That's number one. Number two is you want to still incentivize very much the creation of digital media. And it is not like the print world in that way. I mean, you're right.

Anybody with, you know, a Blackberry can go ahead and publish whatever they want, essentially. And, you know, that's a very different model, and we want to still incentivize people to produce content and not necessarily lock them down into some type of very regimented process, but you want to -- you want to, at the same time, leverage that ability to produce information, but at the same time, provide a significant assurance that, at some point in time, it's right.

So my suggestion would be if there's any way to kind of punt that whole issue somehow to some type

of working committee for further public input might be beneficial where transparency or -- I don't -- I think there's room to be a little more exploratory there.

MR. HANNAN: Thank you. That's great input, I think. Yes, sir.

MR. SHULER: John Shuler, University of Illinois, Chicago. Spinning off of that, I think the great advantage the GPO has over any other agency is what I call nested authenticity. You not only have technical tools, such as chain of custody, PGI, whatever; you also have associated with you socially and politically, economically, a set of depository libraries that provide another level of authentication in a social setting.

And your relationship with the other agencies, the dealings that you have with them exchanging, for instance, the Federal Register, is a further form of authentication.

I think where you spin off the idea of the further working group, if you move it from just a technological solution to -- more into Ric's world, the policy solution, it's going to become multifaceted, and

I think you move away from that conundrum that everything has to be solved one way, every chain of custody problem has to be solved with the same tool.

But I think what you're experiencing in this room a little bit is we have a multiple chain of custody issues that can be solved in different ways, and not all of us have to be resolved by what would be a strict, orthodox, legalistic point of view of what chain of custody is.

MS. DALECKY: Selene Dalecky from GPO. Just to kind of piggyback on that conversation and what was going on with Suzanne and Harold in the discussion of how do you know it was printed from GPO, and I think one of the things that you can refer back to is kind of what John just said where Suzanne really put herself into the chain of custody, if you will, by being a depository library, by receiving the printed book from GPO directly or from a trusted source with the inter-library loan or in some way like that and, therefore, can vouch for the authenticity of this particular publication in hard copy.

And then the same could, I think, be modeled

in the electronic world where how do you know this file -- as Suzanne points a patron to an electronic version of a document, well, it was retrieved from a repository maintained by GPO, and we have the same assurance from the originator and from the system that it was maintained in that this is reliable.

If you extend that even further out, when we talk about system-to-system content and how can we be assured that from system -- system B has content from system A that's the same and can be assured. If you've got that type of trust between the two systems in, say, an approved -- an approved depository repository where the transmission of the content has been assured and the maintenance is assured at the depository end, you again can have assurance of that particular data set. So I think it can keep extending and follow and model the print world, at least through that process.

MR. SHULER: Another way to look at it, the chain of custody is not protecting the document itself, although it's a vital part of it; it's protecting the relationship that the document establishes. And on a very legal -- that relationship is extremely tied

together to that particular document. But in relationships fostered by GPO through its multiplicities, that authenticity can take place in different ways in congruences rather than in specific non-overlapping applications, if you will.

MR. HORTON: This is Bob Horton from the Minnesota Historical Site. Is that another way of kind of saying you're developing a routine?

MR. SHULER: Yes.

MR. HORTON: Yeah. And then the routine course of business is a legal term and does have evidentiary value. And maybe that's why we're kind of struggling with this because no one has developed a routine. You are inventing a lot of these things. And with the recognition that the technology will change, you'll have to reinvent, but the routine is the goal.

And you can -- maybe there's a way of prioritizing this. Some of these routines will be easier to establish than others, and that's probably the place to focus, I guess. But I think you guys are right on target with that approach.

MR. MAYER: I guess I just briefly want to

flag from a technical perspective this idea that decentralized authentication, if you will, working with depository libraries and so on, this has been tried before in other computer security contexts. The one that comes to mind is authenticated websites, making sure that they are actually who they say they are by kind of using almost certificate depositories throughout the web.

It's proven to be pretty difficult getting everything synced up, a lot of problems getting adoption, users tend not to want to check against a bunch of different certificates, especially if it requires a lot of network access. Some folks might just want to use offline validation. So from other contexts, at least, I do want to flag that this has not been the most successful approach.

MR. HANNAN: Okay. Other viewpoints or thoughts? This has all been really helpful. If not, let's try moving to this next one because this is very quick, I think. Re-authentication over time, this is really more just to make -- provide some awareness for you about what can -- I think Kevin brought this up

earlier today about the fact that you may need to re-authenticate content based on events that would occur. Challenge number one, algorhythms and techniques change over time -- start talking ten years, and sometimes it's five years or less, potentially. And some examples of that are shown on your handout.

Challenge number two is the thing that keeps the computer security people up at night sometimes and that is sort of issue with the mathematics involved, that ends up meaning that quick action needs to be taken. So this is just a challenge for us. Not so much -- we just wanted to give you a little bit of awareness that we are thinking about this.

So our current planning is maintaining awareness about the requirements of the changing nature of authentication standards. We are continuously -- Selene and myself [inaudible]. Periodically assessing the requirements of re-authenticating content more frequently than annually, if we need to. We kind of look at that at least once a year.

We need to start, plan to start re-authentication along [inaudible] requirement changes,

rehab of automotive [inaudible] authentication [inaudible] utilize, fortunately, for that, but it still provides some challenges to the problem.

Lastly, what we plan to do with the industry data we mentioned is to canvas industry and suppliers of technical systems for innovative solutions that they may have which would take advantage of -- take advantage to help us in this regard.

Hard problem, especially based on just the size of the repository. If you thought you wanted to re-authenticate that, in a short period of time, that gets more challenging than if you want to do it well ahead of when you think you might be wanting to do this.

So this is our plan, and hopefully, we'll have some interesting things that come out of the industry that we can communicate, innovative solutions for this.

At this point, I think we're at a decision point. This is our last topic for today. We can either take a short break and come back, or just plow right through it and get done. Most people say plow

through it and get done. Is that what folks want to do for the most part? That sounds good.

So with that in mind, I want to make sure we give you a choice. The problem statement for us, it's been a challenge. You know, what is granular authentication? It's not easy to say, obviously, either for me. So the way that we tried to formulate this was from a couple of points of view. One, to authenticate an arbitrary portion of the document.

One page out of 500 were sent in. And the other perspective is locating, you know, tasks and relating it to a GPO authentic document. Maybe that's a useful thing, with the goal being, you know, getting an authenticated answer in a reasonable amount of time.

As we thought about this, and I'm going to go to a concept we have in the next slide just to get your feedback on it, the concept of trying to pre-authenticate, you know, sub-slices of content almost seems like a fool's errand to some of us. But, you know, it kind of cuts both ways because as soon as you spend all that time pre-authenticating things, it's really not useful to the next person that comes into

queue for how they want to slice down out of a 500-page document they're interested in.

So we have a concept on the next slide. It really goes to the second point above, which is this concept that you're trying to -- you have some text, and you really think it relates to some document, but you'd like to be able to have a definitive link back to the authenticated publication for GPO.

So let's just step through this real quick. I know it's probably hard to read. It's probably hard to read in your handout. I apologize. We can try to -- when we post this on the website later -- increase the font so it's not too much of an eye chart.

But this concept is just a concept. Nothing -- we just wanted to get your feedback to see if it was useful or not. The first step is a user's got some information they think -- that they have, but they're curious if it's really what the statute says, let's say. There may be bad examples. But not quite sure which one it was.

So the second step is that they would ask our system, I've got this text, I think it comes from a

statute. Can you help me? Our system would provide back a box on the screen that says, go ahead and answer the text. If you think you know something about the document that this is from, enter that too.

Then our system, in step four, will try to do a search through our authenticated material, may provide more than one hit and give you a sense of if there is more than one hit. And in step five, the user can say, oh, that. I thought that that was the statute it came out of. Great. Click on that and we could, potentially, provide a signed and real-time statement that this text is in this file. Does that sound like it would be something if we could do that that would be useful?

MR. SCHWEICKHARDT: John, Reynold Schweickhardt. Let me just add one of the concepts here is also from a burgeoning and effectivity point of view, so if you simply search for text and you find it in a repealed statute or appropriations bill from four years ago instead of current law, you haven't found -- haven't found the hit you were looking for, we're thinking that part of the value-add here would be able

to talk about burgeoning and effectivity of the hits that potentially would come back.

MR. HANNAN: I mean, clearly, this is a little bit like orienting a Google search. If you do a Google search on one word, you might get 10 million hits. If you orient the Google search -- or any kind of search, a little more carefully, a more reasonable answer can come back within a reasonable amount of time.

So there's a lot of that kind of aspect that we play into this, but this was something that a lot of us here have just kind of kicked around thinking, you know, if we had a capability to do something like this, would it be of value?

It kind of felt to us like it might be of value to people in your arena, so I think we might could just see if that makes sense. And if it does, then maybe we can look at what's involved in that, which it could be really hard, and you can end up doing it.

MS. BAISH: I think it's a very exciting concept. And in previous discussions, I know Mike and

Selene when we talked about the granularity, this is not the kind of value add delivering back to the customer that I thought you were talking about. The limitation I see is it's only going -- would only apply to the content of FDsys.

But it also, what it gets to is the issue when a user goes into a court opinion on Google Scholar or on West or Lexis, that they could take that granular piece up against, hopefully, what is in content of FDsys and verify it with your authenticated version. So from that component, I mean it's really kind of an exciting idea.

The limitations being it's your content only, but it gets back to one of my earlier comments that in the ideal world, someone could look at electronic court opinion and be able to find out whether it's I a commercial database or on a website or on GPO and third party be able to check the authentication. So this is great. I hope there are ways for you to do it, actually, or explore that further.

MR. HANNAN: Great. Thank you very much. That's what we were looking to see. Yes?

MS. CASE: Are you thinking that you'd only go for an exact match?

MR. HANNAN: I think we would have to -- fuzzy logic kind of principles, get too tricky. This is probably going to be hard enough. It would have to be spaces. It probably would have to be an exact match, exactly.

MS. CASE: So any system that repurposed the data and took out some text in the middle, your search would fail.

MR. HANNAN: In this concept because we're not trying to evaluate somebody else who's got the legitimate thing, we're trying to do what we can.

MS. CASE: So it would be an exact match.

MR. HANNAN: Yeah.

MR. WALSH: And this is just one scenario, right? It's one thing considered as an option. There are others that you would explore.

MR. HANNAN: Based on, yeah, what --

MR. DAVIS: Ric David, GPO. Another point to add to this issue, you note in the example John referenced one page as an example, which would be hard

enough. I can tell you that in some of the use case discussions that we've had, people have said why limit it to one page? Why not make it a paragraph, a sentence, a word? And the thing to consider in all of that is context.

You need to be able to map it back to the original document so that the authenticated granular content is considered within the context of the actual document, so that's a consideration. I think that if we could do it with a page, we would go from there, but I think that would be a logical start as opposed to even getting down to a smaller more precise segment of content.

MR. HANNAN: Well, I think that's a good point, Ric. This concept here really deals with that second point on the slide before. And in this case, what we're putting more on the user to exactly type in a paragraph they're interested in.

If they're willing to do that, I'm willing to wade through all the hits that might come back, and there probably are ways we could help make that a more meaningful display of the hits. This kind of gets it,

a kind of general purpose way of thinking about that, but there's a lot -- it's going to end up being a pretty good bid on the user involved in terms of how they structure their search. Yes, sir?

MR. MAYER: So I guess my reaction to the proposal is it looks a lot like search to me.

MR. HANNAN: Well, yeah.

MR. MAYER: I guess you could look for an exact match, you could kind of do a little bit more fuzziness. And then the crypto piece at the end, I guess, you could certainly, on the fly, sign specific elephants.

You could also run this entire thing over https, you get the same security guarantees. The only difference is it's kind of not portable, you can't take that signature away with you. So if that's -- fair enough. But beyond that, I guess my reaction is it's search.

MR. HANNAN: Thanks, that's good.

MS. CASE: It's also search on a large segment of text, and it's tough to get a -- well, first of all, most full-text search engines don't do exact

matches well. It's going to be a tough search to build.

MR. HANNAN: It might be. I think that's why we wanted to see, first of all, if it was possible and -- is it of value? That's really the question we'd like to get some feedback from you on. We can work through some of the practicalities. We don't know that this is really feasible at all as a concept to gain some feedback from -- if it was feasible to do, would it be of use? Because if you said no, it wasn't any use, I wouldn't go any further with it. But that's a great point.

You're absolutely right, and thank you for bringing it up. You're right. It's a hard problem. Just about like every topic we've talked about today is a hard problem, and that's why it hasn't been done yet. Yeah? You have it, Jessica.

MS. MCGILVRAY: I think that there definitely is value here. And it reminds me of what they're doing with electronic management -- electronic record management at NARA. They're creating a search field and working. So if you do decide to go to this, that

might be something -- I don't know.

MR. HANNAN: We might be able to collaborate.

MS. MCGILVRAY: Yeah.

MR. HANNAN: Oh, that's good. Thanks. Yes, sir?

MR. ANDERSON: I agree with Pat in terms of the caution about scalability. And I agree. I think a lot of the execution of this -- I mean, number one, I think it's pretty cool. But number two, I mean, given whatever resources that you might need to bring to bear on this, I mean I think off the top of my head, there's a scalability issue because you might have an expectation to find the sentence. And that is going to be very tough to do, I would think.

The second thing is that this presupposes that the end user is already using not an authenticated version, it would seem to me. Somebody's getting, you know, an unauthenticated copy of the Federal Register from Lexis or West Law, and then they want to make sure that that's the real deal.

So it's kind of -- if you've already created it before, I mean, I guess my question is -- I mean,

obviously, there's that scalability issue going the other way. If you're going through the entire issue of the Federal Register, you do want to find what it is you're looking for. But, you know, why aren't you using the original thing to begin with?

MR. HANNAN: Well, I think that's an interesting question. Let me give you my take on it, which is this actually does support that use case, which is I've got the 500-page document, and I'd like something signed on this paragraph. I've got to type in the paragraph, or cut and paste it. Then you can get an actual signed granting of that at your arbitrary convenience, even though you've got the 500-page authenticated thing.

Now I don't know if that's a bit of use. That was kind of the concept here is you actually could do - - even if you already had the authenticated piece, you could supply the file name which would make this search probably very scalable.

MR. ANDERSON: Right.

MR. HANNAN: It's GPO's site -- to figure out --

MR. ANDERSON: And what's the end result to the user?

MR. HANNAN: We don't know yet except it would be a signed data structure in some format that you could, you know, PDF -- probably PDF is the most, honestly, practical thing for us to do given the challenges with signed XML. It would probably be a signed PDF of that saying this exact text is found in this file, maybe the URL. I mean, we haven't really gone very far with this.

We really wanted to just kind of preview it, see if it sounded like it would be of value, and we could start to go back and sharpen our pencils a little bit and see if it's something that might actually show some progress here. So -- I mean, those are the kind of things, really, we're thinking about.

This text -- all we can assert is, oh yeah, that text is in this document, and we would be willing to do that, I think, although I always have to talk more to lawyers, honestly, if it's valuable.

MR. ANDERSON: I mean, we can quickly, following up on that comment, I mean one of the things

that somebody is going to have to do in the legal world anyway is to see what the context of the quotation is anyhow. So the second -- the other user down the road from the information provider, who is user number one, is going to have to synthesize whatever snippet, whatever official snippet that is by going back to the original document anyway and taking a look at the context in which that's located, I would think.

MR. HANNAN: Well, I think, you know, that's a good point. So, having thought through this a little bit, the better you structure the signed thing you get back, probably the easier the interaction between user one and two is. With regard to that, certainly, the goal here is to -- the user two in your concept, where user one gets this signed data structure, you know, if user two wants to go back to the official source, they can clearly do that. That's great.

And maybe over time, that data structure will obviate the need for time to be spent by users two, three and four, potentially, maybe if it's -- so that would be the -- that's the thought process about why this might be a useful thing to, you know, put some GPO

resources towards perhaps -- we just really wanted to get -- does that make sense to you?

MR. ANDERSON: Yeah.

MR. HANNAN: So we wanted to really get some feedback from you all about did it sound like something that might be worth looking into. It sounds like, generally, it might be. Yes, sir?

MR. MAYER: This occurred to me earlier. Yeah, I think there actually also could be some really serious security concerns doing this. If someone can submit something, let's say comments to an agency or something like that that gets published by the GPO and you're willing to sign it, then you've kind of given away the ability to arbitrarily sign anything with the GPO key, which could be bad. So let's suppose --

MR. HANNAN: I'm sorry, yeah, could you explain that?

MR. MAYER: So the idea being, let's suppose some bad guy wants to get something signed with the GPO key. As long as they could get it into any document if you're willing to sign at any level of granularity, they just say sign that passage, that will --

MR. HANNAN: Didn't we already put that into a file that's signed and published on the Internet?

MR. MAYER: So you're signing the entire file instead of just the little snippet. And so the idea would be, you could imagine a tax where they want just a snippet signed. Because that can be bad, for example, if you're using a key.

MR. HANNAN: Right. So this --

MR. MAYER: Something to think about.

MR. HANNAN: It is. That's a great point, Jonathan. I really appreciate it. This gets down to what -- okay, so this just says that you made a comment in a document that was published. That's what would come out of that signed data structure. And so it does speak a little bit to, you know, that's all it would say, you know, to the person that you would show this to. And so, again, it gets back to the context that Rick mentioned, and that is an interesting -- we're trying to sort through that to about, it really dovetails with Steven's comment about it's easy to like -- it's easy to find the word "the" in all kinds of documents. That's not --

MR. ANDERSON: Or it's hard.

MR. HANNAN: Well, yeah, right. It's certainly not very useful.

MR. ANDERSON: Yeah.

MR. HANNAN: Yeah, it's not -- so you raise a good point about what can you really say. And that's why the wording about all this says is this text is in this file. It doesn't say anything more than that. And that may or may not be as useful. Certainly, if it's a big long string of text with a lot of context to it and a file that's pretty definitive and descriptive in its name, then you can start to use this more, you know, as human beings for useful business or whatever.

So that's all I -- I think it's a good point, and it's again, I think we'd have to think through this too so that we provided something that was really useful and wasn't subject to being misused. You're absolutely right about that. So we have to be really careful about, you know, the little asterisks on the fine print and what came out of this. So I appreciate you saying this. Thanks. Harold?

MR. BOOTH: So, just to kind of add on to, I

think, some of the various comments from what I heard, basically, perhaps just turning back, you have everything already up on the web, turning back a URL that provides what -- it sounds like people like the search capability. So you provide the search capability, and the response could be, you know, here's the URL, and the fragment of the document from which that particular text was found, or fragments where it was found.

And that fragment or URL can be shared amongst people, and people can then go to GPO themselves and then go and get that source document. But they get to go right to the location or citation that's of interest to whoever originally got that particular fragment or piece of document.

MR. HANNAN: Right, I think that's kind of what we had in the back of our minds here.

MR. BOOTH: The point being is you wouldn't need to have a signature over everything. You'd just provide a URL. They always have to come back to GPO servers to get that information.

MR. HANNAN: Well, I think that's the --

that's an interesting point about this concept is because you're trying to get out of a 500-page file that you're carrying around and showing to somebody. And so it's tricky, everything that's been pointed out, it's kind of tricky to think about how to bring value in that regard, but I think your comments are right along and good.

So this is helpful to us to think about, again, it's all predicated -- and as Steven and Pat pointed out -- it's really predicated on being able to get an answer back in a reasonable timeframe. And that's going to take some work on our part to think about and see if that's really achievable. There might be some limits on, you know, it's not just a sentence. It's maybe a paragraph at the smallest segment, something like that. So we have to think about that too. Yes, Steve?

MR. ANDERSON: Well, it also is going to depend a bit on the historic development of what a piece of information content is going to look like, what the package itself is. Because my hunch is that at the same time, you're going to have more 500-page

documents coming. You also might have a lot fewer 500-page documents coming, and you're looking at information content being sentence-by-sentence or paragraph-by-paragraph anyway, and looking at granularity in terms of authenticating individual content packages that are tiny rather than very, very large.

So, you know, I guess what I'm saying is we could go both ways from here on out. I mean, you might have a lot more volume in your file to authenticate. But you also might have a lot fewer pieces and it really goes to what approach information design is going to be like, you know, in the next 20 years, or longer.

MR. HANNAN: That's a good point.

MR. GALLUCCHIO: I think what we're talking about, there's a good argument for your guys to break up the data into reasonable chunks. I mean, you could, theoretically, sign your whole repository with one key, right? And then the user, to authenticate, would have to download that whole repository, and that doesn't make sense, right? So you could do it by book, you

could do it in chapter, subsection. You know, it's kind of -- it's up to you. You should try to make it practical. It shouldn't be arbitrary. It shouldn't be an oracle where you can just ask it to sign anything you want. So --

MR. HANNAN: Well, yeah. That's not really what that -- you don't get to sign whatever you want. It has to be part of a document that's already been signed. But, yeah, that's a great -- that's a great point. I think we struggled with that a little bit too, you know. Which way do you approach it from? You know, pre-signing all the granules, that's really arbitrary small levels. I think that's tough and time consuming.

But I think your points are really good, and comments -- I appreciate it. Other thoughts? Everybody tired? Wish there was more coffee out there? Hearing no more input at this point, I think we'll go to the next couple slides, which is wrapping up.

And the first thing on this is thank you so much for your time and your participation. It was really, really good. We really appreciate you taking

time out of your busy day to come down to GPO and be here. We're going to post information about the -- shop at the URL that we have listed there. We'll have the slides from today, the transcript of what we went over in several days, and then we're going to be working on a summary report to try to summarize things for us that we can always go back to.

That's going to take us a while to go through that. But if you have additional comments, this email address is set up to accept them, so please go ahead and send us thoughts. We realize you didn't see these slides before you got here, so if you go home or you're on the airplane and you think there's something, whatever, please feel free to email them in, and we'll factor that into the -- report. Mary Alice?

MS. BAISH: I just was curious, have you set a date yet for the industry day?

MR. HANNAN: No, we haven't yet. We're thinking probably August, September. We would like to do it sooner rather than later, but we want to assimilate all this good input and put some thought to all the things that folks have said, and that helped us

make that better. I think -- we haven't really talked about this in a great level of detail with either Mike or Ric, but I think the thought process is that for industry day, you know, if you all want to come, I think that's probably going to be [inaudible]. Exactly. Yes?

MS. DULABAHN: Mike, did you want to talk about the federal agencies group?

MR. WASH: Yes. Thank you for the reminder. There's a -- part of the INDIF (ph) initiatives, there was the creation of a Federal Agency Digitization Guidance --

MS. DULABAHN: Guidelines Initiative.

MR. WASH: Guidelines Initiative that is a group of federal agencies working together with some outside support as advisors, I think, is the way it's structured, right, Beth?

MS. DULABAHN: Right.

MR. WASH: For Xerox and Kodak and others.

MS. DULABAHN: Right. There's an outside advisory board, and we've also engaged consultants.

MR. WASH: Right. And that's been something

that we have used to help develop a digitization specification for preservation initiatives for retrospective documents. We're in the process of forming a similar type of group for authentication that will be, again, a similar type of organization, but it's a federal agency type of organization. It's just in the formation stages. GPO is going to lead that one with participation with other agencies.

But I think it would be really helpful to form an advisory group in a similar way with folks from here, if you are interested, to kind of follow what's going on within this working group of federal agencies on authentication and provide advisory input from time to time. So we should probably just kind of reach out and see who would be interested in working in an advisory role on that sort of thing as we get the working group up and running. Anything to add to that, Beth?

MS. DULABAHN: I'd just say the focus of that initiative, initially, had to do with digitization of documents. And I think that's one of the things, looking at the topics that have been discussed to date,

like chain of custody and so forth, we talked about, I think, primarily born digital documents, and then we had some reference to the print chain of custody, and I think there's a space in the middle that is the print documents that get digitized and move to digital form.

And so what I would see as one of the main topics for discussion of that group, among other things, would be what does it mean to authenticate digitized documents, especially when, for example, we've been asked to redact Social Security Numbers in other documents. I think there are a line of nuances related to digitizing documents that, collectively, the federal agencies need to come to grips with.

MR. WASH: Okay. So if there's interest in that, we'll be pulling more information together on the formation of it, but I think -- I would say for now, just be aware that that group will be forming. I think there should be some sort of advisory role, and if you're interested, we could form an advisory group for that group.

MR. HANNAN: I think with that, we really appreciate it. Thanks for coming, and we're adjourned.

(Whereupon, at 2:38 p.m., the Document Authentication Workshop was adjourned.)

CERTIFICATE OF COURT REPORTER

I, NATALIA KORNILOVA, the officer before whom the foregoing meeting was taken, do hereby certify that the witness whose testimony appears in the foregoing meeting was duly sworn; that the testimony was taken by me in stenotypy and thereafter reduced to typewriting by me; that said meeting is a true record; that I am neither counsel for, related to, nor employed by any of parties to the action in which this meeting was taken; and, further, that I am not a relative or employee of any counsel or attorney employed by the parties hereto, nor financially or otherwise interested in the outcome of this action.

_____

NATALIA KORNILOVA

NOTARY/COURT REPORTER

IN AND FOR THE DISTRICT OF COLUMBIA

My Commission Expires: April 14, 2012

H

Q

T