



**Council Discussion, Questions and
Answers from the Fall Meeting 2004**

Future Digital System

Contents:

- I. Briefing Topic: The Future Digital System**.....2
 - I.1. Setting the Stage2
 - I.2. New Information.....3
 - I.3. Micro Recap4
 - I.4. Additional Information4
- II. Assumptions From Council**4
- III. Questions to Council, with Council Discussion**.....5
 - III.1 What do you think will be the benefits and challenges to GPO
of the future digital system?5
 - III.2 What additional challenges do you see?5
 - III.3 Discuss how different classes of users will need to assess content in the
future digital system. The first user being the expert user, a librarian acting
as an intermediary for the content and the user. Describe the ways in which
the expert user will access the content.....6
 - III.4 Given the ways in which the expert user accesses content what search
capabilities and other features will the future digital system need to offer?6
 - III.5 Describe the ways in which the general library user will access the content
in the future digital system.7
 - III.6 Describe the ways in which the general public user will access content
in the future digital system.8
 - III.7 Given the objectives and needs for accessing the future digital system what
does metadata need to accomplish to support these needs?.....8
- IV. Questions from Council Addressed at the Meeting**.....9
- V. Audience Questions Addressed at the Meeting**.....11
- VI. Audience Questions Addressed after the Meeting**.....16

I. BRIEFING TOPIC: The Future Digital System for the U.S. Government Printing Office

I.1. SETTING THE STAGE

GPO's mission includes both printing government documents and disseminating them to the public. Under the public printing and documents statutes of Title 44 of the U.S. Code, GPO's mission is to fulfill the printing needs of the federal government and to distribute those printed products to the public. All printing for the Congress, the executive branch, and the judiciary —except for the Supreme Court — is to be done or contracted by GPO except for authorized exemptions. The Superintendent of Documents, who heads GPO's Information Dissemination Department, disseminates these government products to the public through a system of nearly 1,300 depository libraries nationwide (the Federal Depository Library Program), GPO's Web site (*GPO Access*), telephone and fax ordering, an on-line ordering site, and its bookstore in Washington, D.C. The Superintendent of Documents is also responsible for classification and bibliographic control of tangible and electronic government publications.

Current printing industry trends show that the total volume of printed material has been declining for the past few years, and this trend is expected to continue. A major factor in this declining volume is the use of electronic media options. The move to electronic dissemination is the latest phase in the electronic publishing revolution that has transformed the printing industry in recent decades. This revolution was driven by the development of increasingly sophisticated electronic publishing ("desktop publishing") software, run on personal computers, that allows users to design documents including both images and text, and the parallel development of electronic laser printer/copier technology with capabilities that approach those of high-end presses. These tools allow users to produce documents that formerly would have required professional printing expertise and large printing systems.

These technologies have brought major economic and industrial changes to the printing industry. As electronic publishing software becomes increasingly sophisticated, user-friendly, and reliable, it approaches the ideal of the print customer being able to produce files that can be reproduced on the press with little or no intervention by printing professionals. As the printing process is simplified, the customer can take responsibility for more of the work. Thus, the technologies diminish the value that printing organizations such as GPO add to the printing process, particularly for simpler printing jobs. Nonetheless, professional expertise remains critical for many aspects of printing, and for many print jobs it is still not possible to bypass the printing professional altogether.

The advent of the Internet permits the instantaneous distribution of the electronic documents produced by the new publishing processes, breaking the link between printing and dissemination. With the increasing use of the Web, the electronic dissemination of information becomes not only practical, but also more economical than dissemination on paper.

As a result, many organizations are changing from a print to an electronic focus. In the early stages of the electronic publishing revolution, organizations tended to prepare a document for printing and then convert the print layout to electronic form—in other words, focusing on printing rather than dissemination. Increasingly, however, organizations are changing their focus to providing information—not necessarily on paper.

Today an organization may employ computers to generate both plates used for printing as well as electronic files for dissemination. Tomorrow, the organization may create only an electronic representation of the information, which can be disseminated through various media, such as Web sites. A printed version would be produced only upon request.

As in private industry, printing and dissemination in the federal government are being heavily affected by the changing technological environment. This new environment presents both financial and management challenges to GPO. Just as the volume of material provided to private firms for printing has decreased over the past few years, so has the volume of material that federal agencies provide to GPO for printing. In addition, federal agencies are publishing more items directly to the Web—without creating paper documents at all—and are able to print and disseminate information without using GPO services. Similarly, individuals are downloading documents from government Web sites, such as GPO Access, rather than purchasing paper copies of government documents, thus reducing document sales.

These changes in federal printing and dissemination are also creating challenges for GPO's long-standing structure for centralized printing and dissemination. Agencies are to notify GPO of published documents (if they used other printing sources), which allows GPO to review agency documents to determine whether the documents should be disseminated to the depository libraries. If they should be, GPO can then add a rider to the agency's print contract to obtain the number of copies that it needs for dissemination. However, if agencies do not notify GPO of their intent to print, these documents may not be available to the public through the depository library program.

I.2. NEW INFORMATION

There is no existing single system, or group of systems, that adequately meets the needs and mission of the GPO. The Future Digital System is being developed to meet the future needs and will possess the following attributes:

- *Infrastructure independence*: An architecture that allows preservation of content- independent of any specific hardware and software that was used to produce them;
- *Modularity*: Ability to use plug-in components that can be replaced with minimal impact to remaining components as workload and technology change;
- *Scalability*: Capable of accommodating growth and managing differing sizes of repositories and ever increasing volumes of records;
- *Extensibility*: Be able to handle additional kinds of content over time, not limited to specific types that exist today;
- *Comprehensiveness*: Provide support for content management lifecycle processes for all types of records; and
- *Flexibility*: Enable the GPO to tailor content-based services to suit its customers' needs and enable the GPO to implement progressive improvements in its business process over time.

To meet strategic objectives, the GPO must integrate its solution for preservation and long-term access to content with the lifecycle management of that content throughout the Federal Government. To meet the challenges of today and the future, the proposed Future Digital System will provide the following capabilities:

- Capability to accept the transfer of content in a wide variety of formats as they were created or stored by their creators and the flexibility to easily adapt to future file formats;
- Capability to ingest, preserve, and provide access to that content;
- Capability to store content as electronic records in a manner that is independent of any particular hardware and software component over long periods of time;
- Capability to scale in order to store and preserve records based on the predicted digitizing of existing hard copy publications and the harvesting of volumes of fugitive content on Web sites;
- Capability to provide access to the content in electronic form for all users based on established user rights and privileges thus ensuring that Future Digital System users are able to access all of the electronic records that they are entitled to see;

- Capability to provide access to the content in a manner that is consistent with current technology and the changing expectations of its diverse user communities;
- Capability to adapt to changing technology in order to continue to provide access to and delivery of content desired by the user community; and
- Capability to identify the essential characteristics of the content that is being preserved for the purposes of authentication and certification.

Reference models for the proposed system are included in this section of the briefing book. These will be discussed in the presentation of Future Digital System.

I.3. MICRO RECAP

Current printing industry trends show that the total volume of printed material has been declining for the past few years, and this trend is expected to continue. A major factor in this declining volume is the use of electronic media options. The move to electronic dissemination is the latest phase in the electronic publishing revolution that has transformed the printing industry in recent decades.

The advent of the Internet permits the instantaneous distribution of the electronic documents produced by the new publishing processes, breaking the link between printing and dissemination. With the increasing use of the Web, the electronic dissemination of information becomes not only practical, but also more economical than dissemination on paper.

There is no existing single system, or group of systems, that adequately meets the needs and mission of the GPO. The Future Digital System is being developed to meet the future needs.

I.4. ADDITIONAL INFORMATION

GPO has completed the Concept of Operations for the Future Digital System, which will serve as the roadmap for development activities on this system over the next few years. The full document is on the FDLP Desktop at http://www.access.gpo.gov/su_docs/fdlp/tools/ConOps_1004.pdf, and a summary is at http://www.access.gpo.gov/su_docs/fdlp/tools/Conops_Summary.pdf. The presentation shown at the meeting is at http://www.gpo.gov/news/2004/04_10_18_dlc_presentation.pdf.

II. ASSUMPTIONS FROM COUNCIL

- II.1. GPO needs to enable a world class information content system to meet the needs for permanent public access to official government content
- II.2. This Future Digital System will possess 4 key functions regarding content: Version Control, Authentication, Preservation and Access
- II.3. The Future Digital System will be rules-based and policy-neutral, allowing it to adapt to changes in the operational environment over time.

III. QUESTIONS TO COUNCIL, WITH COUNCIL DISCUSSION

III.1 QUESTION: What do you think will be the benefits and challenges to GPO of the future digital system?

DISCUSSION BY COUNCIL

One benefit is a more timely information base for what's coming in and what's going out. It still remains to be seen how long it would take for something to be sucked in and spit back out the other end, but it looks as though it would be a much briefer period of time than what we've experienced at GPO in the past and it would be easier to track information as a government information professional as it goes through the process. So two benefits would be timeliness and transparency.

Another benefit is that we would have the ability to easily push materials out into our existing system without having to create new mechanisms, that the rules would be written in such a way, and that the policies insure that there is no fee access to the information.

The benefits for this system are: verify and track versions, assure authenticity, observe content, and provide permanent public access. These would be enormous benefits assuming that this comes to fruition.

If GPO doesn't have to actually invent this and write the programs, if there are things they can buy off the shelf for parts of this, then that would be a benefit. It's a huge undertaking.

III.2 QUESTION: What additional challenges do you see?

DISCUSSION BY COUNCIL

The potential is there, depending on how this is implemented, for actually integrating government information into everyday lives and everyday research if it works. And that would be a nice sort of anti-silo thing to happen.

We do want to make sure that we aren't met with the challenge of wrapping this in a proprietary format at any time.

You have a lot of assumptions here that people are going to buy what you've got and part of that would come with the authentication part, that there's going to be a big market for this. And perhaps there is. You may have some numbers and data to back that up. But the challenge is, once the system is designed, the parts that are supposed to make money may not.

An ongoing challenge that faces not only GPO but other organizations in the government, is this business of balancing the requirements for metadata versus the tendency to have fugitive documents. There's ways the great bulk of fugitive documents could be captured pretty readily now but without much metadata. So there's a tradeoff that is going to be a challenge for GPO in the future balancing those two.

Another challenge is the procurement process within the federal government.

III.3 QUESTION: Discuss how different classes of users will need to assess content in the future digital system. The first user being the expert user, a librarian acting as an intermediary for the content and the user. Describe the ways in which the expert user will access the content.

DISCUSSION BY COUNCIL

This assumes that the expert user is a librarian, and it depends on the original content of the information and who the information is designed for. The expert user should be looked at more as a broader class. If it is economic data, the expert user is not a librarian. The expert is an economist or a research economist. So we need to define expert user.

Our expectations ride very much on the quality and quantity of the metadata attached to the material, how it's indexed. In the example of an integrated library system, how thoroughly can it be mined? What access points are available and what flexibility does the user have? While Google is great for some general things, frequently searches must identify things so specifically that it will require a strong integration with the metadata.

Is there a distinction between identifying content and accessing content? There's finding aids and then there's the ability to pull the content down or view the content. By accessing content does that basically mean identifying content?

The access that an expert user has, whoever that expert user may be, should be the same as other users would have. It's maybe tiered access with advance search functionality, but there shouldn't be special areas that certain experts can use but that the general public can't use.

This implies that you will get the librarian expert user group training and that that may be the difference right there, that we'll have specific training on how to do it and that we are to impart that to others, like economists.

III.4 QUESTION: Given the ways in which the expert user accesses content what search capabilities and other features will the future digital system need to offer?

DISCUSSION BY COUNCIL

This also crosses into general users. We have a large growing Hispanic population. Would there be a Hispanic/Spanish version to get into this?

We do not want to have impediments placed in the way of retrieving this material easily; we do not want to use what we call the net library model of obstruction. Because people have made it as hard to get the material out as possible. We want it to be intuitive to the end user, not to the expert, as to how to get this out. We do not want difficult downloads and complex processes. We also don't want it built around proprietary software that they have to have to get the format out.

It should be granular so that pieces can be found as well as entire volumes. For example, it is already there with the Federal Register, where you can find individual pieces without having to dig through the whole document.

There's the standard things that people want. They want to be able to do full text searches, they want to be able to do Boolean searches, they want to be able to do fielded searches, and if they mess up and make their search terms too broad, they want relevance ranking and results.

The expert user as well as the end user wants a very reasonable guarantee of their privacy, that their information will not be mined out.

One other possible feature an intermediary might like would be the ability to browse as opposed to search.

III.5 QUESTION: Describe the ways in which the general library user will access the content in the future digital system.

DISCUSSION BY COUNCIL

It would be nice if the general user didn't have to figure out that what they were looking for was government information and go to a separate place to find it. We have that model now.

A lot of places are looking at integrating this as fully as possible into the library's regular catalogs. So whatever this design is it should be easily integrated so it could be searched directly from the library catalogs and accessed as though it were just any other piece of information in the library system.

Taking that a step further, that integration should go beyond the library to be fully integratable into other systems out there on the network as well as library systems.

It ought to be integrated with Google and other things. Also the interface, the usability should be really studied and designed in the best possible way for ease of use.

If we're talking about a broad spectrum of libraries, the system needs to be able to deliver the information to systems that may not be on the cutting edge. One of the problems we see a lot of is the library, the public library, or the user in their home, doesn't have the absolute positive latest version of IE, therefore they can't see particular functionality. So that's something that needs to be addressed in some form.

If there were some way that you could mark private sector versus public-produced information, so that one could easily see that one is the public and authoritative and legal version, versus one that is reprinted or some other type of distribution, that would be a benefit.

The depository system should push this material to libraries, so that we are not having to go and pull this material into our systems. We want to have it coming to us. That will

ease it into the library systems, as well as pushing it into other areas. If you push it out to Google and others, they will see how they can use it to their advantage, too.

III.6 QUESTION: Describe the ways in which the general public user will access content in the future digital system.

DISCUSSION BY COUNCIL

You should be also looking at some kinds of indexing tools that are automated semantic indexing, the kinds of things that don't rely on human manual intervention, and that's related to scale and volume content in the future if the system really works well. Usually these systems involve some manual interaction but it's a scaling potential that the future general public should find useful and are already using in some situations.

The general user getting this information from home or work or wherever should basically have the same level of access through library catalogs that are online, through Google and other utilities on the net. And presumably through something GPO operates along the lines of GPO Access sort of indexing.

The main difference would probably be that they might not have the trained intermediary librarian to help them; however, it would be a goal to have the services of a librarian wherever they are. So even from home, in effect end users could be using the assistance of librarians.

III.7 QUESTION : Given the objectives and needs for accessing the future digital system what does metadata need to accomplish to support these needs?

DISCUSSION BY COUNCIL

We need to see the report from the experts on metadata before we can really address this issue. That is a very important group to get the input from.

The more metadata the better when it comes to being able to retrieve information. But the problem is it's expensive and back in the old days if you didn't have metadata you were sunk. You couldn't retrieve anything really. And that's changed, so that now metadata isn't absolutely essential, as Google has demonstrated for web pages, and some other people have demonstrated for doing cross-database searching. So it's a moving target, and there's pluses and minuses to doing whatever you choose to do.

One of the things that metadata needs to accomplish in a system like this is describe the unique piece of information that you're getting. Because this system has the potential of bringing in data sets and geospatial data sets and maps and music and posters, so that the metadata schema that one ends up using needs to both mesh with the standard and be flexible enough to incorporate all the various formats that will be delivered.

IV. QUESTIONS FROM COUNCIL ADDRESSED AT THE MEETING

IV.1 QUESTION: How is it that we're going to insure that perpetual access and what underlying business model is there to do that?

RESPONSE: The perpetual access to the data would be accommodated by the preservation process, where preservation includes the information storage process, managing that storage in case there is a need to refresh the data or change the format. That is the preservation process, which includes the process of actually pushing into the collection of last resort and pulling out if necessary. That activity is core to perpetual access, making sure that there is a copy of the content available for access at any point in time.

We would expect to continue to use FDLP funding for this function.

IV.2 QUESTION: Will education question be included as part of user support? There's an underlying assumption to your assumption; that is that users can educate themselves on how to utilize digital. But users don't necessarily educate themselves on how to mine our system for information. So that's a part of digital delivery systems that needs to be in the mix.

RESPONSE: There are other processes like training classes and education that are not done by the system but are externally managed in order to facilitate use of the system. You would not necessarily be expecting to build that kind of educational training within the system, although obviously to the extent it can, you always want a system to be simple, intuitive and self-instructing. But even when you start out with that goal sometimes external training is almost always required. We're going to have to begin to think as we build this about what the implications will be on FDLP librarian training.

IV.3 QUESTION: Have you had a chance yet to think about how you can change this vision into reality? For example, you envision a step-wise incremental approach where you deliver something that works and then build on that base or are you going to try to design this system as one unit?

RESPONSE: We have given a lot of thought to it, and the approach that we're taking is a phases approach where this document or concept of operations characterizing how the system is intended to be used and the functions that it will deliver was one of the deliverables. It serves as the basis for us to create the next stage of more detailed planning of how to actually get this done. We're in the process of that now.

There are elements within GPO today that are completely aligned with this system. And there are some things that are not entirely aligned. So actually the stage that we're going through right now at GPO is the mapping process, the taking activities that are going on within GPO today and seeing how they map to this concept. And we'll then start to manage those projects in a consistent way to this future digital system. So activity essentially is already under way.

Relative to do we see an incremental type of approach or the big bang type of approach, it will be incremental. There will be elements of this that will start to fall into place probably within the next year. There are some activities relative to the conversion of documents that we want to get some traction on right away. But any activity that we do from this point forward needs to be tested against the concept of operations so that we can make sure that we do this in a very coordinated way. So elements of it will start to get turned on over the next year or so.

IV.4 QUESTION: One of the statements under access is the subcategory of ordering, and the word payment always brings up a little flag to people. So when it is talking about the capability of users to place orders and submit payment electronically, what type of payment are we referring to there? Is it more the agencies paying for things, or is it the customer, as in the Federal Depository Library Program?

RESPONSE:

An ordering process that understands transactions from content originators certainly needs to be in place to support our agency customers.

The GPO is committed to free public access of the information maintained on GPO Access. We're not going to go backwards and say this class of information has a price tag on it. Everything we put in the system will be accessible at no cost. And it is only when we do extra work to order for a customer that we would expect to make a charge for it. And that's strictly if the customer or a prospective customer feels there's value to something we're doing is willing to pay for it.

IV.5 QUESTION: Is there an assumption that an automated system that handles the push and the pull, that makes depositing easier by creating a style tool box, but at the same time does the thorough harvesting, that a system that can do both things will somehow solve the problem of fugitive documents?

RESPONSE:

It isn't going to be sufficient for GPO to sit back and wait for everything to come to us, we are going to have to be proactive in going out and searching for information. And we definitely will have tools for doing that.

It's always in our best interest to work cooperatively with an agency so that we receive the material in a form and fashion that we like it. But if we can't do that, in the cases where we can't, we want to go out and look for it. Right now we may be getting about 50 percent, and we can see that number rising to north of 80 percent, maybe even as close to 90 percent. But the goal needs to be a hundred percent always.

V. AUDIENCE QUESTIONS ADDRESSED AT THE MEETING

The facilitator of the Council sessions accepted questions from the audience written on GPO-supplied cards. Seventeen of the twenty questions were answered during the Council session. Those questions and their answers are summarized below. Questions held to answer at a later date, either because of time constraints or the need for a subject matter specialist to provide more detailed answers, follow the questions answered during the session.

V.1 QUESTION: You spoke of the use of style tools to standardize ways that documents [are formatted]. What is your thinking on making agencies comply with this?

RESPONSE: It's a matter of working closely with OMB and of finding incentives rather than hammers to get cooperation from the agencies. One incentive might be to offer "one-stop shopping," where GPO could accept the content if it's compliant with our style tools and pass that on through to the archives, relieving the agency of that duty.

V.2 QUESTION: We are looking to the GPO for models for managing permanent public access to our state government documents. Part 1, will the developing system modules be scalable to state operations? And part 2, will the code be an open source versus proprietary?

RESPONSE: On part 1, we would like to get a sense of what the library community's feelings are, and get your advice in this area. On part 2, it's unlikely that we would use any kind of closed or proprietary software for any public purpose.

V.3 QUESTION: Reference was made to systems characteristics. In particular, that system will be policy neutral. Can you please further explain what this means, how a policy neutral system would work?

RESPONSE: A policy neutral system is one where you have enough flexibility in the design of the system that if a definition, for example, changes, the entire system does not have to be redesigned. One policy example would be the definition of what constitutes a publication or what constitutes aspects of the types of content that we deal with. We don't want to paint ourselves into a corner by saying that this is the way the system will always be, such that if there is a change either through law or the way that we want to run this system from a business perspective, for example, that we don't have to redesign.

V.4 QUESTION: Users should not have to identify themselves to the system to search or retrieve information; no logons or profiles.

RESPONSE: Part of the process we're going through now is understanding what the requirements and expectations are for user needs. We haven't defined what the access criteria would be for gaining access to content, and part of the process will be making those definitions.

V.5 QUESTION: Under the subject title of access, what degree of customized data access can be provided before changes occur? The concern is that easily anticipated uses will only be available to those who can afford it rather than charges being limited to requests that actually require unique/special effort for response.

RESPONSE: If we put the information into our content database, we will continue to make access to that information free. However, if we add value and customize the information, then we could charge for that service. We could build a subscription business from doing that.

V.6 QUESTION: How will the future system address corrected versions and errata and will uncorrected versions be retained online and if so be labeled as such?

RESPONSE: That goes to understanding what the rules are for versioning and version control. We haven't defined that yet. We need to make sure that we have the capability to maintain various versions. So if a rule, for example, comes into play where a correction to a document does indeed generate a version, then it will be an official version and both those versions will be available. So that's an example of what a rules-based system will be able to support.

V.7 QUESTION: Mention was made of cooperation with NARA and LC. The Library of Congress has been one of the main agencies that invokes a cooperative publication (cost recovery) and contributes less of their publishing output to the depository library program than any of the other agencies. Would this process encourage more disposition of LC publications to the system for more free public access even for those agencies reluctant to and how about quasi-government agencies like the Smithsonian?

RESPONSE: Many of these agencies invoke an exemption in Title 44 for publications which must be sold to be self-sustaining. The more we collaborate with them on a variety of things the more ability we have to influence their decisions. But it is ultimately not something we can coerce. Only something we can solicit and try to influence. So the more we're engaged with them, the more cooperation, the more dialog, the better chance there is. But there are no guarantees.

V.8 QUESTION: What about redundancy? Where is the backup?

RESPONSE: The system assumes that no good IT system exists without redundancy. No question but what there are backup systems, a variety of backup systems.

V.9 QUESTION: Please explain who developed the plan and what the process was?

RESPONSE: The Public Printer wanted someone with a very strong technological background. So he brought in Mike Wash, who as an engineer has gone through many of these major kind of projects. He put together a task force of GPO employees. There were four people from Documents: George Barnum, who is part of the new innovation and technology staff, Gil Baldwin, and two other people, Lisa LaPlant and Matt

Landgraf, who worked full-time for several months on this along with some people from other parts of GPO. They took a whole variety of existing documents, including all the documents we're talking about today, as input. Mike imposed the discipline of keeping them from asking the how questions and saying we can't answer the how question yet, put that aside, go on to the what, fighting through a lot of the definitional issues, and produced this document. And then he wrote this summary which was presented to you all yesterday. He did a presentation of it to a significant number of the GPO senior managers to get our response and see where it needed clarification. He translated it into English, so to speak. He did a phenomenal job in grasping the intricacies of all the different things we do and had the ability to see how they all fit together and are integrated with one another.

V.10 QUESTION: What will the relationship of this new system and the ILS be? This is a concern because the ILS is going to be implemented soon but this new system is still in conceptual stage and there is potential for incompatibility and conflict if they are developed as two separate entities.

RESPONSE: We are in the process of mapping current IT systems and current IT projects against the design document. We're looking at the longer-range system in terms of the systems that we're acquiring now, and planning how they will integrate into it.

The ILS is a major component of the design that we are working on. It's an Oracle based system which means it's very easy to write APIs to move data in and out of it to other systems.

So we feel pretty confident that the system will accommodate the ILS, and we don't need to stop moving forward on that to wait for this other design.

V.11 QUESTION: How will this system address the existence of other large integrated systems of Federal Government information? For example, American Fact Finder, AGRICOLA, Star ERIC, STAT-USA, Medline? What sort of proprietary or jurisdictional issues may have to be negotiated in order to guarantee permanent public access to this information through this new system? What incentives can GPO offer these federal agencies and thus persuade them to transfer administration of those pre-existing products to GPO?

The next part to that is: Will the new system attempt to capture agency databases, for example, American Fact Finder, CBC Wonder? Will the system migrate data to new more usable formats, for example, DOS based CDs to Window interfaces?

RESPONSE: There's a lot of parts to that. We have American Fact Finder marketing material already up. We've discussed some aspects of this, about our relationship with Department of Energy. There are obviously a lot of things that have to be negotiated and worked out.

And we're looking at a whole range of incentives, which include the ability to incorporate data from other agencies and historical data from our own agencies with the data they're currently housing, the ability to share resources for migrating this material, the ability to generate and share metadata through cataloging.

We do want the system to be comprehensive, and we know that there are lots of those jurisdictional and territorial and authorship issues that are going to have to be addressed. We obviously have an embedded relationship with the Congress and the Office of the Federal Register, and so some of those things are easier to deal with, but the others are going to have to be addressed as we go forward.

The information that now resides in these agencies ought to be somehow made available to GPO and to the depository libraries. There is also federated searching as a bridge. There are so many different data sets and so many different formats operating on so many infrastructure systems, that they're going to have to be tackled one by one in terms of interagency relationships and existing technology.

V.12 QUESTION: We assume that GPO is charged with managing all Government information. How does this match with other Government information managers of the judiciary, other Government initiatives like regulations.gov? Will the system capture this information or just in scope FDLP items per the current Title 44?

RESPONSE: The Public Printer has been having regular meetings with the Library of Congress, and the current and also the future archivists to make sure that our agencies are working in a complimentary and collaborative fashion, rather than in a competitive fashion.

We're already a partner in regulations.gov. We're already working with a lot of agencies on various aspects of their own publishing and dissemination. We're very sensitive to not duplicating resources. So as we go forward seeking funding for doing these things, it's going to be very important that we be able to document the fact that these things are being done collaboratively, and not competitively.

V.13 QUESTION: In planning for the new system is there going to be one treatment of an item versus many outputs or uses? For example, bibliographic record SuDoc access. The new system should be broader than GPO Access. What is the anticipated time frame for development?

RESPONSE: The goal is to have the new system fully functional by the end of calendar year 2007.

For the first part of the question, we recognize that there's both horizontal and vertical issues related to versioning. And we want to be able to create one parent, mitigated record that tells us that this print copy and this microform and this file on GPO Access and this print-on-demand file and the file the agency originally submitted to us for typesetting all represent the exact same document. So if someone is holding a paper

document in their hand and I'm looking at it on my Blackberry, and somebody else is looking at it on the GPO Access website, we have some way to know that we are looking at the same version of the content. So we are going to have multiple outputs that we have to manage and be able to identify as being the same item.

And then we're going to have to be able to identify the different editions, or, in cases where things aren't as clean as that, this is the snapshot that we took at three months, at six months, at nine months and at 12 months, so you can see how that changed over time.

So we are definitely expecting multiple outputs in different formats for different users, and we are expecting the necessity to manage multiple versions of the same content over time.

V.14 QUESTION: When authentication of publications begins in the new system, the assumption is that all new publications coming into the system will be authenticated. Is this correct?

RESPONSE: Yes, that is correct.

V.15 QUESTION: Is there a way to go back to the past ten years of digital publishing and authenticate those materials published electronically by Government agencies? In summary, is authentication viewed as a forward only activity or are there plans to go back and deal with authentication of previously published documents? If so, how would this be done or could it be done?

RESPONSE: We are definitely planning to authenticate older documents. As things are digitized, we're planning to authenticate them so that we can make statements about their provenance and status at the time of digitization.

And as we go forward from there, at least to be able to say this is unchanged from this time. As we refresh the data on GPO Access, to put it on the next platform, we expect to be digitally signing all of the GPO Access content.

Right now, the digital signatures work really well on the whole document, and so we're talking about starting with things like Congressional Bills where what's on GPO Access is a single file that represents the entire document.

But we also know that we have to be able to authenticate fragments, and we create fragments right now for the Federal Register, Congressional Record and the CFR. We're looking at how we can digitally sign those, but we're also looking ahead to the idea that somebody wants to extract a quotation from a document or an excerpt from a Presidential speech, or whatever it might be.

So we'll be able to get down to that micro level, take a quotation out and carry with it the indication that that quotation is authentic or accurate.

V.16 QUESTION: Will the database be addressable by search engines or will it be hidden Internet material in terms of individual titles within the database? Will the individual titles within the database be permanently addressable with a variant URL or based on standard controlled numbers?

RESPONSE: We do expect to have all of our content accessible at Google and Yahoo and whatever the common search engines are. The Public Printer is committed to supporting whatever way the American people want to find their information.

One of our requirements, and one of the existing features of GPO Access, is that all of the content is directly addressable. We did that specifically so there would be a PURL and libraries could catalog it, if they wanted to link from their website to a document on GPO Access.

But anything that we don't actually bring into our own system, we cannot guaranty that we can make directly addressable through a PURL or a Handle or a successor technology.

V.17 QUESTION: Is GPO planning on digitizing the older issues of the Federal Register or Congressional Record or both?

RESPONSE: Yes, both.

VI. AUDIENCE QUESTIONS ADDRESSED AFTER THE MEETING

VI.1 QUESTION: Requiring executive agencies to link to FirstGov does not result in integration of government information in response to people's information needs. So FirstGov is not the model. On the other hand, FirstGov is great at directing categories of users to areas of interest to them (such as business owners). The description of three categories of users in the Future Digital System document is not sufficiently detailed to facilitate understanding users' needs. There are intermediaries; experts going through library portals from their offices; novices accessing the internet directly, etc.

RESPONSE: GPO's research will further illuminate this point and allow us to refine our research to best capture these needs. We are investigating voice of user/voice of customer (VOU/VOC) techniques to refine our understanding of our customers' needs.

VI.2 QUESTION: Will the Future Digital System replace the FDLP, and will FDLP funds be used for it?

RESPONSE: GPO is not assuming that the FDLP would be replaced by the Future Digital System. Instead, the new system should be developed to meet the needs of the depository libraries, now and in the future. The funding stream for this new system needs to be determined. We have not accurately estimated the cost or the source of funds yet.

VI.3 QUESTION: For the "at home" user there should be a pointer to the expert user (librarian). The expert user should have access to a push technology or a co-browsing technology to use in assisting the "at home" user, and files available for downloading, e.g., PDFs, should be small enough that systems will not timeout.

RESPONSE: GPO appreciates these comments. We will consider these comments in our Voice of Customer (VOC) analysis, which is part of the effort to identify the requirements for the Future Digital System.