



**Testimony of John Lewis Needham
Manager of Public Sector Content Partnerships, Google Inc.
Senate Committee on Homeland Security and Governmental Affairs
Hearing on “E-Government 2.0: Improving Innovation, Collaboration, and Access”
December 11, 2007**

Chairman Lieberman, Ranking Member Collins, and members of the committee.

It's a great pleasure to be with you this morning to discuss Google's role in making government more accessible to citizens. My name is John Lewis Needham, and I am the Manager of Public Sector Content Partnerships at Google. In that capacity, I lead our company's efforts to build public-private partnerships with government agencies in the U.S. and internationally.

Google's mission is to organize the world's information and make it universally accessible and useful. The work that I focus on at Google is critical to this mission because few bodies of information are as important to Internet users as the broad, deep, and authoritative data provided by government.

Making publicly available government information more accessible and useful to citizens not only helps deliver to Internet users the government information they need, but it also enables the government to provide services more efficiently and effectively to taxpayers, and it makes our democracy more transparent, accountable, and relevant to its citizens.

This committee has a long tradition of promoting government transparency, accountability, and effectiveness, which are values that Google shares and hopes to support through its products. We believe, for example, that Web-based platforms like Google Maps and Google Earth – which rely in part on government-provided geospatial data – can be used by the government to better serve its citizens. For example, the U.S. Geological Survey recently developed an overlay in Google Maps that shows real-time data on earthquakes all over the world, connecting citizens to its Earthquake Hazards Program. The National Park Service has also published an overlay in Google Earth that enables users to learn more about recreation opportunities throughout the country.

And as video sharing continues to grow in popularity among Internet users, governments are beginning to turn to platforms like YouTube to better engage with citizens. For example, the U.S. Coast Guard launched a YouTube channel last year through which it shares frequent updates about its rescue missions, humanitarian efforts, and law enforcement operations.

These are among the many innovative and useful ways for government to serve its citizens through the web. This morning, I'll focus my testimony on search engines – the most fundamental Web-based resource for millions of Americans – and share with you how people throughout our nation are using the power of Web search to find and interact with their government.

In my testimony, I will cover four main points:

- First, I'll share some trends in how Americans connect with government online;
- Second, I'll identify the challenges that citizens face in trying to find government information and services on the Internet;
- Third, I'll explain a technology known as the Sitemap Protocol or simply Sitemaps, an open standard that enables web site owners – including government agencies – to make their content more accessible to search engine users; and
- Finally, I will highlight a few of our successful partnerships with government, and outline steps that agencies across the federal government can take to make their web sites more accessible.

How citizens are connecting to government online

Let me start by describing how citizens today are connecting with government information online. According to recent research by the Pew Internet and American Life Project, 77 percent of U.S. Internet users go online to find some form of government information. We can assume that this already impressive number has risen farther since this survey was undertaken in 2004. These Internet users are looking for government services in health care and housing, browsing for tourism and recreation information, or doing research for work or school, to name just a few searches for which government provides important resources. And additional research from Pew reveals that Internet users – nearly two-thirds – expect to find their government online.

So we know that Americans are increasingly looking for and expecting to find online information and services from their government. We also see that Internet users are choosing search engines like Google as their preferred way to connect with the government.

Search engines work by sending a software program to “crawl” the pages on public web sites, adding that information to an index of all the content these “crawlers” find on the Internet. As a result, when a Google user types a query into the search box of our search engine, we very quickly access that index to return search results that we've previously crawled and indexed and that are relevant to that user's query.

Here's an example: The National Institutes of Health's web site, www.nih.gov, offers a rich collection of public health and medical information from the 27 institutes and centers that comprise NIH. Let's say you're trying to find out the status of a study on avian flu. You might not be aware of the relevant NIH service, which is located at www.clinicaltrials.gov, or how to get directly to the page that lists all current avian flu-related studies (which is located at www.clinicaltrials.gov/ct2/results?cond=%22Influenza+in+Birds%22), so you start your search at Google.com.

This is a likely scenario given that very few Internet users go directly to the nih.gov site. In fact, according to our analysis of Internet traffic to NIH web sites during July 2006, only four percent of unique visitors went directly to nih.gov sites. ComScore reports that, during that same month, 70 percent of unique visitors used search engines like Google to find NIH resources, rather than typing complicated, and often unknown, Web addresses into their browsers.

This example is consistent with research that Google has conducted and statistics from other parties on the flow of Internet traffic, which indicate that as many as four out of five Internet users reach government web sites by using Google and similar search engines.

But if the information on a particular government web site is not part of the index underlying a search engine, citizens are bound to miss out on information or services that the agency offers. Today, too much public government information is effectively unavailable to the average American. It can't even be found in the federal government's own search engine, USA.gov. It can't be found through Google's search engine. And it can't be found through other Web search services on which the public relies.

Obstacles to finding government information online

Search engines have made connecting to online government resources easier, but challenges remain. Even though the search engine indexing process I explained earlier usually works quite well, certain barriers can get in the way of search engines ability to provide users with exactly what they're looking for.

More specifically, we have found that many government agencies structure their web sites in ways that prevent search engines from including their information in search results – often inadvertently. These technical barriers include:

- **Dynamic databases.** The most common barrier is the search form for a database that asks users to input several fields of information to find what they're looking for. Our crawlers cannot effectively follow the links to reach behind the search form.
- **Robots.txt files.** A "robots.txt" file is a text file that can be placed on a web site and that lists pages that shouldn't be crawled by a search engine. When a robots.txt file's instructions inadvertently prevent search engines from crawling pages, this becomes a barrier to providing government information through search engines.
- **Inaccessible links.** Outdated links on web pages can disrupt the crawling process, making it impossible for search engines to comprehensively index and make available the information on a web site.

All of these barriers prevent the public information on government sites from being included in Google or other search engine results. As a result, information that is intended for public use is rendered effectively invisible to citizen users.

Let me offer a brief illustration: A citizen may be interested in locating the Environmental Protection Agency's enforcement actions regarding a particular company, so that user conducts a search on Google.com with the company's name and the keywords "epa enforcement." The results would include relevant information, including data from EPA. But what this researcher would not find is a link to the Enforcement & Compliance History Online database at www.epa-echo.gov/echo/, which offers a list of enforcement reports for specific companies. This is because the information in this database cannot be included in a search engine's index.

EPA.gov is certainly not the only government web site that search engines have difficulty indexing. And though defining the exact scope of this issue is challenging, we believe that a large segment of government web sites and databases cannot currently be indexed due to the technical barriers

I've mentioned. In fact, we estimate that the information in part or all of 2,000 federal government web sites is not included in search engines results.

Ensuring the accessibility of government information online

With all that said, the good news is that there's a simple technical solution to help make accessible and useful to search engine users this virtually invisible public government information.

In 2005, Google introduced a technical standard that helps to ensure the accessibility of information on a web site. This standard is called the Sitemap Protocol. It provides a mechanism for a web site owner to produce a list – or map – of all web pages on a site and systematically communicate this information or “Sitemap” to search engines.

When a federal agency places a Sitemap file on its web site, search engines can readily identify the location of all pages on the site, including database records lying behind a search form. Using this sitemap, search engines are more likely to index and make the information that the agency's web site provides visible to citizens.

The Sitemap Protocol has been widely embraced by the search engine industry. In November 2006, Google, Microsoft, and Yahoo introduced the Sitemaps.org web site, formally announcing their joint support for the standard for their respective Web search services. Ask.com and others also support the standard. What this means is that, in implementing Sitemaps, a government agency can be sure it's better serving the American people, no matter which search engine individual citizens are using.

Taking the step to implement the Sitemaps protocol is free and easy. It doesn't require site redesign, the purchase of new technology, or more than a few hours or days of a webmaster's time. Implementation involves creating a list of web pages, or URLs, in an acceptable format, and adding the file that contains this list to a web site. Google provides a variety of tools to help accomplish this task and we present them to public sector Web managers at www.google.com/publicsector.

It is important to note that I'm only talking about making the public information on a government web site more available to citizens. Information that is maintained on internal web sites, including personally identifiable information, should not be made accessible through any search engine and is not the type of information we're working to crawl and index through Sitemaps.

Successes in connecting citizens with government

We believe that it would be technically simple for federal government agencies to produce a Sitemap for the information on their web sites, and that doing so would bring significant benefits to individuals throughout the country. And we know that implementing the protocol is easy to do because we've worked with many government partners – at all levels – to take this step.

For example, the Department of Energy's Office of Scientific and Technology Information operates a large database that makes research and development findings available to the public. OSTI developed a Sitemap for its Energy Citations and Information Bridge services in just 12 hours, opening up 2.3 million bibliographic records and full-text documents to crawling by search engines. The process took such little time that OSTI's director, Walt Warnick, remarked that the agency had “spent more time talking about what we did regarding the Sitemap Protocol than we

did executing it.” After its implementation of Sitemaps, OSTI saw a dramatic increase in traffic to its services as more citizens discovered previously unknown resources through Web searches.

Other federal agencies that have recently embraced Sitemaps are the Government Accountability Office, which used the standard to make a database of 30 years of GAO reports and Comptroller General decisions visible to search engine users; the Library of Congress, which has made the text and images of its American Memory collections easier for Internet users to find; the National Agricultural Library, which opened its database of research in the field of nutritional supplements to medical researchers; the National Archives and Records Administration, which is now in the process of sitemapping the federal government’s largest public database; and GovBenefits.gov, which now makes its profiles of over 1,000 government benefit programs just one search away.

At the state level, we’ve launched Sitemaps partnerships with Arizona, California, Florida, Michigan, Utah, Virginia, and the District of Columbia. These partnerships are making it easier for residents to uncover the postings for jobs available in their state, reports on school performance, and the professional license record of contractors.

The private sector long ago recognized the increasing importance of Web search. Consequently, private sector webmasters have widely embraced the Sitemap Protocol, enabling search engines to create the rich index of content you can access today through Google.com. Still, the federal government lags behind.

Last month, this committee took an important step in elevating the profile of efforts like Sitemaps to better connect citizens with government by voting in favor the E-Government Reauthorization Act of 2007, S. 2321. The act would direct the Office of Management and Budget to create guidance and best practices for federal agencies to make their web sites more accessible to commercial search engine crawlers. It also requires federal agencies to ensure their compliance with that guidance and directs OMB to report annually to Congress on agencies’ progress.

We commend Chairman Lieberman and Ranking Member Collins for their leadership on this issue, and we look forward to working with you to have this important legislation passed into law.

Continuing to help government make its information available to citizens

Mr. Chairman, while my remarks today have focused on web sites and search engines, it’s clear that in the years ahead government agencies will need to make information in other formats more accessible.

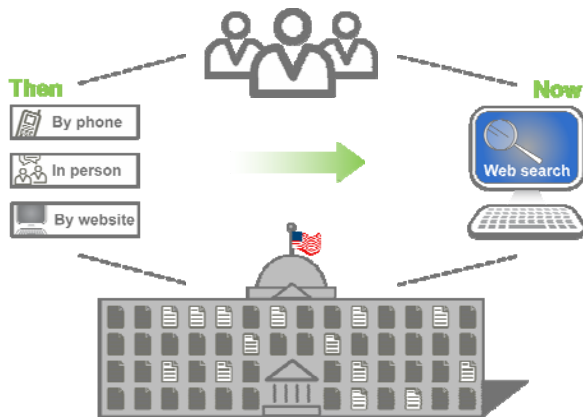
In the Web 2.0 world, where more and more citizens are using blogs, wikis, online mapping, video sharing services, and social networking sites to communicate and collaborate with each other, there will be even more demand for government to bring information to citizens where they are through these new platforms. This information will also help serve as a core component of the user-generated content that is driving the deeper engagement of Americans with each other, and with our democracy, through the Web.

We are excited by the promise of this trend, and we’re committed to continuing to better connect government to citizens.

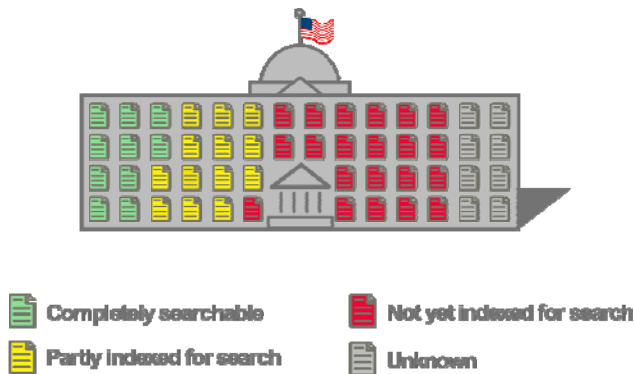
Thank you.

Addendum: The Sitemaps Story in Three Illustrations

1. **The Context:** Citizens are connecting with government information through search engines.



2. **The Problem:** Information in part or all of an estimated 2,000 federal government web sites is not included in search engines because of technical barriers.



3. **The Solution:** Sitemap Protocol provides a free and easy way to ensure that government information is made easily accessible to citizens through search engines.

