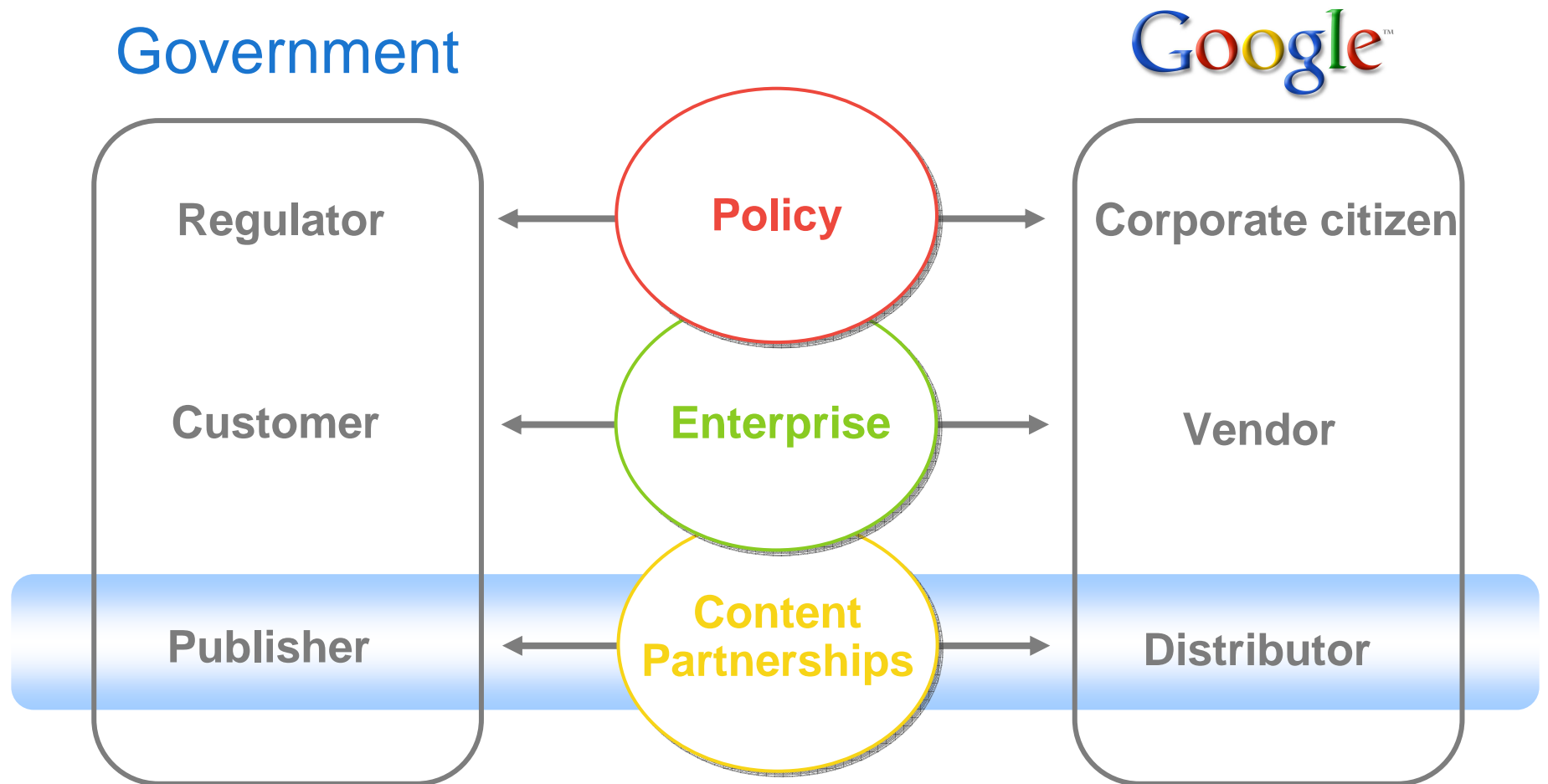


Ensuring government is only
one search away

Implementing the Sitemap protocol

Google's relationships with government



Sitemaps.org

An open, industry standard for search engine crawling



What are Sitemaps?

Sitemaps are an easy way for webmasters to inform search engines about pages on their sites that are available for crawling. In its simplest form, a Sitemap is an XML file that lists URLs for a site along with additional metadata about each URL (when it was last updated, how often it usually changes, and how important it is, relative to other URLs in the site) so that search engines can more intelligently crawl the site.

Web crawlers usually discover pages from links within the site and from other sites. Sitemaps supplement this data to allow crawlers that support Sitemaps to pick up all URLs in the Sitemap and learn about those URLs using the associated metadata. Using the Sitemap [protocol](#) does not guarantee that web pages are included in search engines, but provides hints for web crawlers to do a better job of crawling your site.

Sitemap 0.90 is offered under the terms of the [Attribution-ShareAlike Creative Commons License](#) and has wide adoption, including support from Google, Yahoo!, and Microsoft.

Agenda

Barriers to search engine crawling

Scope of the issue

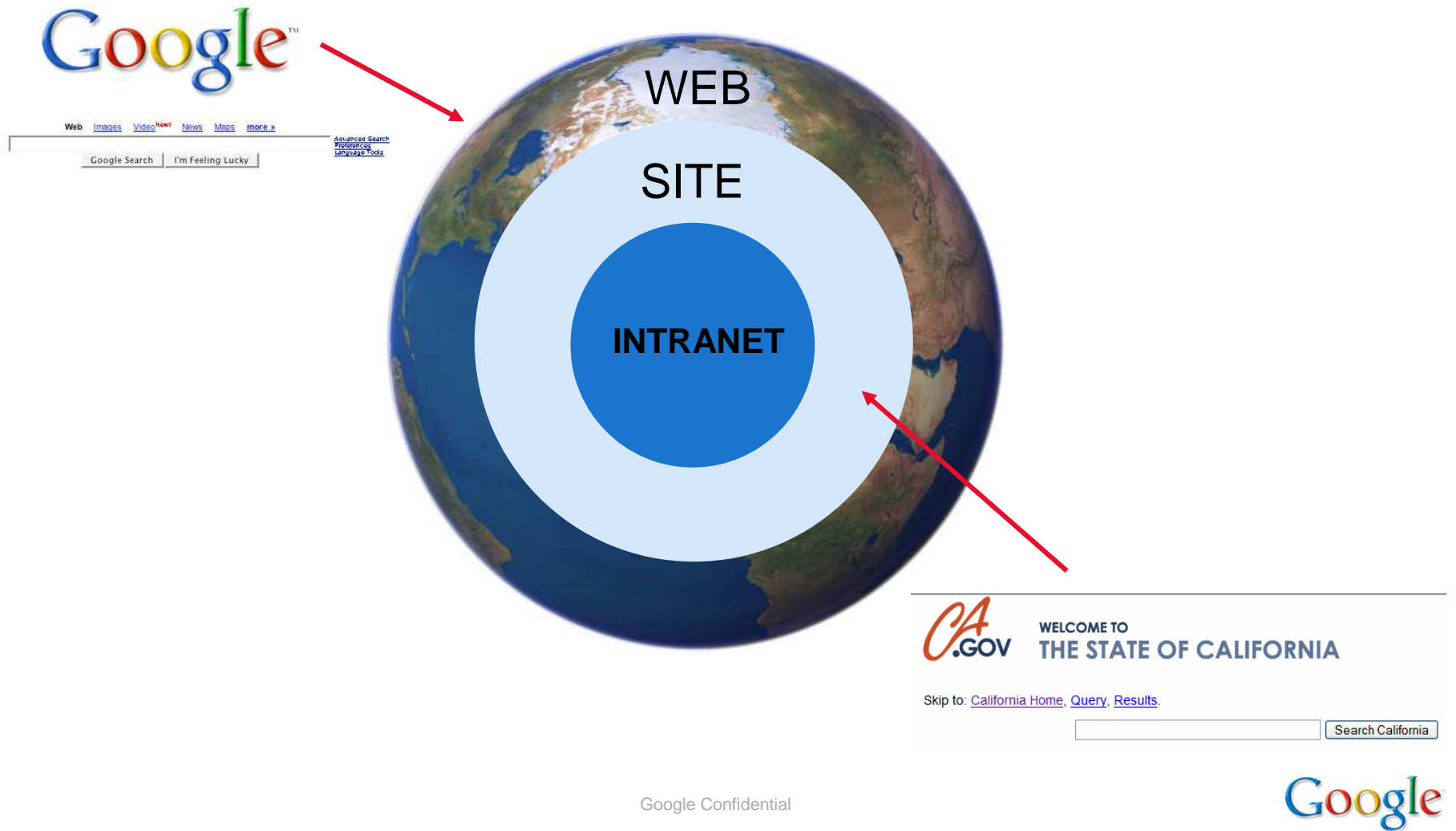
How citizens access government online

The Sitemap protocol

Success stories

Meeting citizens' expectations of government

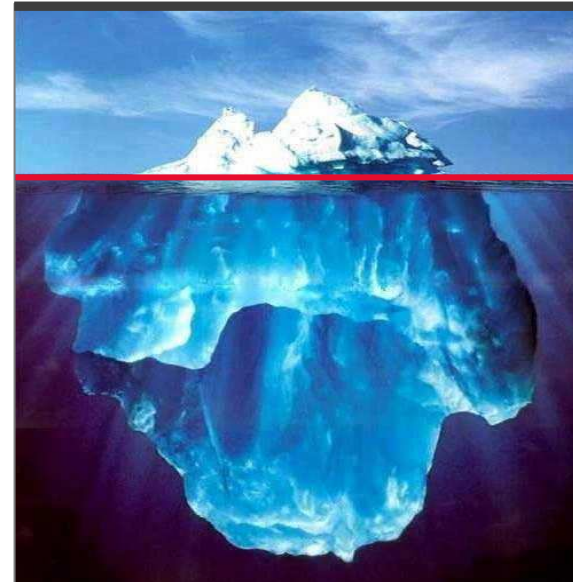
The three levels of search



Many government web services are effectively invisible to search engine users

Barriers to search engine crawling:

- Content “hidden” behind search forms
- Non-HTML links
- Outdated robots.txt crawling restrictions
- Server errors
- Orphaned URLs
- Rich media: audio, video
- Paid/subscription-only access



WEB
Searchable

DEEP WEB
Not searchable

Representative barriers

- 1: US Census Bureau Question & Answer Center (<https://ask.census.gov/>):
 - [census america's largest minority group]
 - Invisible page: <http://tinyurl.com/yqbjja7>
 - Barrier: Dynamic database

Representative barriers

- 2: ATF website (<http://www.atf.gov>):
 - Sample search: [atf small business program]
 - Invisible page: <http://tinyurl.com/2c2qam>
 - Barrier: [Robots.txt](#)

Representative barriers

- 3: USAJOBS.gov job search (<http://jobsearch.usajobs.opm.gov/>):
 - [statistician jobs federal government]
 - Invisible page: <http://tinyurl.com/2yja46>
 - Barriers: Agency information access policy and dynamic database

Scope of the issue

- US:
 - Federal government: an informal “Scope and Progress” report
 - EPA:
<http://spreadsheets.google.com/ccc?key=pUb62ZKHnzgqEoGF4LFf3Gw>
 - State of Texas:
<http://spreadsheets0.google.com/ccc?key=pUb62ZKHnzgpiqOMSqZj-AA>
- UK central government: initial analysis
- Organization of American States:
<http://spreadsheets.google.com/ccc?key=pUb62ZKHnzgo2eHJUS9bBRg&pli=1>

Public sector information is a pillar of the web

Generally authoritative, unbiased, free and trustworthy



The majority of citizens access government through search engines

National Institutes of Health (nih.gov):

- More than 70% of unique users in July 2006 were referred by search engines (Google, Yahoo, MSN, AOL, Ask)

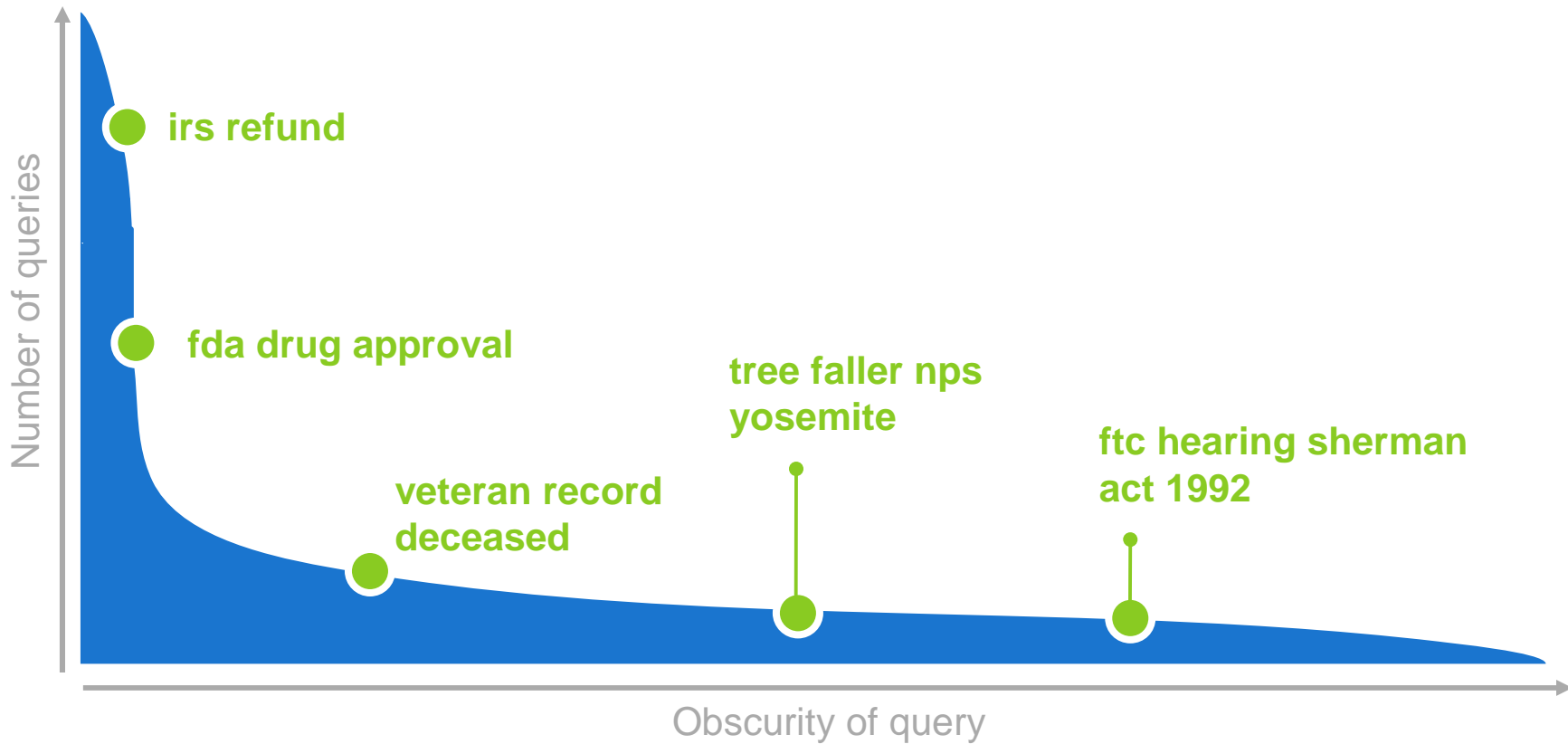


- Only 4% of unique users came directly to nih.gov sites

Source: ComScore, 2006

And citizens expect to find everything through search

The long tail of US federal government information



The good news

- The Sitemap protocol provides a solution:
 - Can make all pages, documents, and records in a web service accessible to search engine crawlers (and thus visible to search engine users)
 - Does not require website redevelopment or redesign, just a comparably modest time investment
 - Keeps the site's owner in control

XML Sitemap

- A comprehensive list of a site's URLs tagged with:
 - Location
 - Last modification date
 - Change frequency
 - Priority

XML sitemap

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.google.com/schemas/sitemap/0.84">
  <url>
    <loc>http://www.example.com/</loc>
    <lastmod>2005-01-01</lastmod>
    <changefreq>monthly</changefreq>
    <priority>0.8</priority>
  </url>
  <url>
    <loc>http://www.example.com/catalog?item=12&desc=vacation_hawaii</loc>
    <changefreq>weekly</changefreq>
  </url>
  <url>
    <loc>http://www.example.com/catalog?item=73&desc=vacation_new_zealand</loc>
    <lastmod>2004-12-23</lastmod>
    <changefreq>weekly</changefreq>
  </url>
  <url>
    <loc>http://www.example.com/catalog?item=74&desc=vacation_newfoundland</loc>
    <lastmod>2004-12-23T18:00:15+00:00</lastmod>
    <priority>0.3</priority>
  </url>
  <url>
    <loc>http://www.example.com/catalog?item=83&desc=vacation_usa</loc>
    <lastmod>2004-11-23</lastmod>
  </url>
</urlset>
```

One solution for all search engines

Once implemented, Sitemaps can be made available to all major search engines



“The launch of Sitemaps is significant because it allows for a single, easy way for websites to provide content and metadata to search engines”

— Tim Mayer, Senior Director of Product Management, Yahoo Search

“We are 100% behind this protocol -- this kind of collaboration will help improve the search experience for all of our customers”

— Ken Moss, General Manager, Live Search

Success stories

- US federal government:
 - PlainLanguage.gov
 - OSTI, Department of Energy
- State of California
- National Archives of the United Kingdom

PlainLanguage.gov success story

- Plain Language Information and Action Network (PLAIN), a federal inter-agency volunteer working group that encourages clarity in government communication to the public through PlainLanguage.gov
- Before-and-after examples of government documents served dynamically, thus uncrawlable



PlainLanguage.gov success story

Google Web Images Video News Maps more »
over-the-counter drug label Search Advanced Search Preferences

Web

[The New Over-the-Counter Medicine Label: Take a Look](#)
All nonprescription, **over-the-counter** (OTC) medicine **labels** have detailed usage and ... The new **Drug Facts labeling** requirements do not apply to dietary ...
www.fda.gov/cder/consumerinfo/OTClabel.htm - 34k - [Cached](#) - [Similar pages](#)

[Over-the-Counter Medicines: What's Right for You?](#)
Read the **label!** **Drug labels** change as new information becomes available. That's why it's important to read ... Consumer Education: **Over-the-Counter** Medicine ...
www.fda.gov/cder/consumerinfo/WhatsRightForYou.htm - 40k - [Cached](#) - [Similar pages](#)
[[More results from www.fda.gov](#)]

[Understand Over-the-Counter Drug Labels](#)
Reading the product **label** is the most important part of taking care of yourself or your family when using **over-the-counter** (OTC) medicines (those that are ...
www.webmd.com/content/article/61/67580.htm - 40k - [Cached](#) - [Similar pages](#)

Plain Language: Over-the-counter drug label DocID: 6
Examples>**Over-the-counter drug label**. **Over-the-counter drug label** Food and Drug Administration, No-Gobbledygook Award Winner. No None Version available ...
www.plainlanguage.gov/test/Examples/indexExample.cfm?record=6&search=BA - 13k - [Cached](#) - [Similar pages](#)



Plain Language.gov
Improving Communication from the Federal Government to the Public

Home What is PL? Why PL? Using PL

[Examples](#)>Over-the-counter drug label

Over-the-counter drug label
Food and Drug Administration

[Over-the-counter drug label](#) (pdf)

- Web manager successfully implemented sitemap in ~8 hours, using available resources and through trial and error
- As new examples are added to the database, the sitemap is regenerated and submitted

OSTI success story

- Department of Energy agency that “makes R&D findings available and useful, so that science and technological creativity can advance”
- Web manager submitted sitemaps for Energy Citations and Information Bridge services, opening 2.3M bibliographic records and full-text documents to crawling
- Sitemap standard assures search engines have “a complete picture” of information in OSTI services



OSTI success story

Google [Web](#) [Images](#) [Video](#) [News](#) [Maps](#) [more »](#)

nuclear physics giant resonances osti [Advanced Search](#) [Preferences](#)

Web

[Energy Citations Database \(ECD\) - Energy and Energy-Related ...](#)
 Subject, N68351 --Physics (Nuclear, Experimental)--Nuclear Properties ... The **giant dipole resonance** is not appreciably excited for any of the targets.(AIP) ...
www.osti.gov/energycitations/product.biblio.jsp?osti_id=4013309 - 15k -
[Cached](#) - [Similar pages](#)

[Energy Citations Database \(ECD\) - Energy and Energy-Related ...](#)
 Nuclear Physics Lab. Sponsoring Org, DOE/ER. Subject, 73 **NUCLEAR PHYSICS** AND ...
 tapes: **giant resonances**; nucleus-nucleus reactions; **nuclear** astrophysics; ...
www.osti.gov/energycitations/product.biblio.jsp?osti_id=6686188 - 13k -
[Cached](#) - [Similar pages](#)

[PDF](#) [Yields of Radionuclides Created by Photonuclear Reactions on Be, C ...](#)
 File Format: PDF/Adobe Acrobat - [View as HTML](#)
 Division of **Nuclear Physics**. Prepared by the OAK RIDGE NATIONAL LABORATORY ...
 where the (iy,n) values are averages over the **giant resonances** of ...
www.ornl.gov/~webworks/cppri/2002/rpt/112299.pdf - [Similar pages](#)



Home About What's New Basic Search Advanced Search

Energy Citations Database 1948 - Present

Help Full-Text Availability Security/Disclaimers Comments

DOE Information Bridge Energy Files OSTI Home energy.gov

Availability information may be found in the Availability, Publisher, Research Organization, Resource Relation and/or Author (affiliation information) fields and/or via the "Full-text Availability" link. For a journal article, please see the Resource Relation field.

Title Giant resonances observed in the scattering of 96- and 115-MeV alpha particles

Creator/Author [Youngblood, D.H.](#) ; [Moss, J.M.](#) ; [Rozsa, C.M.](#) ; [Bronson, J.D.](#) ; [Bacher, A.D.](#) ; [Brown, D.R.](#)

Publication Date 1976 Mar 01

OSTI Identifier OSTI ID: 4013309

Other Number(s) CODEN: PRVCA

Resource Type Journal Article

Resource Relation Phys. Rev., C, v. 13, no. 3, pp. 994-1008

Research Org Cyclotron Institute and Physics Department, Texas A and M University, College Station, Texas 77843

Subject N68351 --Physics (Nuclear, Experimental)--Nuclear Properties & Reactions, 6 <= A <= 19--Nuclear Reactions & Scattering;N68451 --Physics (Nuclear, Experimental)--Nuclear Properties & Reactions, 20 <= A <= 38--Nuclear Reactions & Scattering;N68551 --Physics (Nuclear, Experimental)--Nuclear Properties & Reactions, 39 <= A <= 58--Nuclear Reactions & Scattering;N68651 --Physics (Nuclear, Experimental)--Nuclear Properties & Reactions, 59 <= A <= 89--Nuclear Reactions & Scattering;N68751 --Physics (Nuclear, Experimental)--Nuclear Properties & Reactions, 90 <= A <=

- Benefits include better representation in search results and reduced load on servers (by limiting duplicate crawling)
- First implementation completed in 16 staff hours -- can now be easily replicated across search engines

Google partners with states of Arizona, California, Utah and Virginia

Google and Four US States Improve Public Access to Government Websites

Mountain View, Calif. - April 30, 2007 - Google Inc. (NASDAQ: GOOG) today announced partnerships with the states of Arizona, California, Utah and Virginia to make it easier to search for hard-to-find public information on state government websites.

These partnerships developed as both Google and officials with the four state governments recognized that the public is increasingly turning to search engines like Google to access government services, but that a significant share of the information available on the web is difficult to find. In Google's index of information sourced on the web, as a result, many online government services can be difficult for the public to find.

Connecting citizens with their government by offering the public easy access to information and services is consistent with Google's mission - to organize the world's information and make it universally accessible and useful," said Eric Schmidt, Google's Chairman and Chief Executive Officer. "These partnerships are a strong sign that Google is pursuing partnerships with agencies to better serve the public."

California joins Arizona, Utah and Virginia as one of the first states to partner with Google to improve accessibility of their online services.

"California's website is a portal for the public to learn about our state and our services," said California Governor Arnold Schwarzenegger. "Our partnership with Google empowers Californians to know more about our government and access to services and helpful information."

To improve access to state agency websites, Google has had its technology managers with the four states remove barriers to adding information to its index, so that more pages on the sites can be found through a Google search.

This means job seekers in Utah can now search on Google for the state and find job postings provided by the state's Department of Workforce Services. In Virginia, students interested in the region's colonial history can use Google to find a greater variety of online resources provided by the Commonwealth's archive, the Library of Virginia.

“Transparent digital government is a top priority in Virginia. Our goal is to provide easy, quick and intuitive citizen access to every government resource. Google is helping us to achieve this.”

-- Tim Kaine, Governor of Virginia

California Department of Health Services implemented Sitemaps



Web [Images](#) [Video](#) [News](#) [Maps](#) [Desktop](#) [Moma](#) [more »](#)

california richwood ground beef

[Advanced Search](#)
[Preferences](#)
[Language Tools](#)

[Advertising Programs](#) - [Business Solutions](#) - [About Google](#)

©2007 Google

Now citizens can easily find more of the information on the agency's public websites

Web Images Video News Maps Gmail more ▼



california richwood ground beef

Search

[Advanced Search](#)
[Preferences](#)

Web

Results 1 - 10 of about 981 for [california richwood ground beef](#).

[PDF\] CA Firm Recalls Beef Products for Possible E. coli O157:H7](#)

File Format: PDF/Adobe Acrobat - [View as HTML](#)

FSIS-RC-020-2007. CALIFORNIA FIRM RECALLS GROUND BEEF FOR POSSIBLE E. COLI O157:H7 CONTAMINATION. WASHINGTON, April 20, 2007 – Richwood Meat Co, Inc., ...

www.fsis.usda.gov/PDF/Recall_020_2007_Release.pdf - [Similar pages](#)

[California Firm Recalls Ground Beef Products For Possible E. coli ...](#)

24, 2004 — Richwood Meat Co., Inc., a Merced, Calif., firm, is voluntarily recalling approximately 90000 pounds of frozen ground beef and beef patties that ...

www.fsis.usda.gov/OA/recalls/prelease/pr007-2004.htm - 12k - [Cached](#) - [Similar pages](#)

[Food Safety > California Based Richwood Meat Recalls Ground Beef ...](#)

Richwood Meat Company has voluntarily recalled approximately 107000 lbs. of ground beef products for possible E. Coli contamination.

www.emaxhealth.com/75/11242.html - 17k - [Cached](#) - [Similar pages](#)

[STATE HEALTH OFFICER ADVISES CONSUMERS NOT TO EATRICHWOOD MEAT ...](#)

California Pacific Associates Ground Beef Patties 80/20 – 2 Round – Code 4115 ...

Richwood Meat Company Fresh Ground Beef Patties 36/8 oz patties – Code: ...

www.applications.dhs.ca.gov/pressreleases/store/PressReleases/07-32%20richwood%20e.html - 34k - [Cached](#) - [Similar pages](#)

National Archives of the UK

- ~20 million records across multiple services inaccessible to search engines
- Sitemaps successfully implemented to open several million records
- [needham national archives uk]

The screenshot shows the National Archives website interface. At the top left is the logo, a large red letter 'A' with 'the national archives' written below it. To the right of the logo is a navigation menu with links: '> Contact us > Help > A to Z index > Site search > Accessibility'. Below this is a red horizontal bar with white text links: 'Home', 'About us', 'Visit us', 'Research, education & online exhibitions', and 'Search the archives'. Underneath the red bar is a breadcrumb trail: 'You are here: [Home](#) > [Search the archives](#) > [National Register of Archives](#) > Details'. The main heading is 'National Register of Archives' in large red font. To the left of the heading is a collage of historical images including a man in a cap, a woman, and a child. Below the heading is a search section titled 'Search the register by:' with a magnifying glass icon. There are four search filters: 'Corporate name', 'Personal name', 'Family name', and 'Place name', each in a light-colored button.

Meeting citizens' expectations of government

Google Benefits

- **Expands** the reach of Google's search services
- Incorporates more **authoritative, trustworthy** information
- Increases crawling **efficiency**

Agency Benefits

- Ensures **all** public information is visible to **all** potential users
- **Maximizes** the investment e-gov services and the search engine channel

User Benefits

- **More and better** information from a trusted source
- **Quicker** access to and navigation of government information and services
- **Serendipitous discovery**

Toward more transparent government

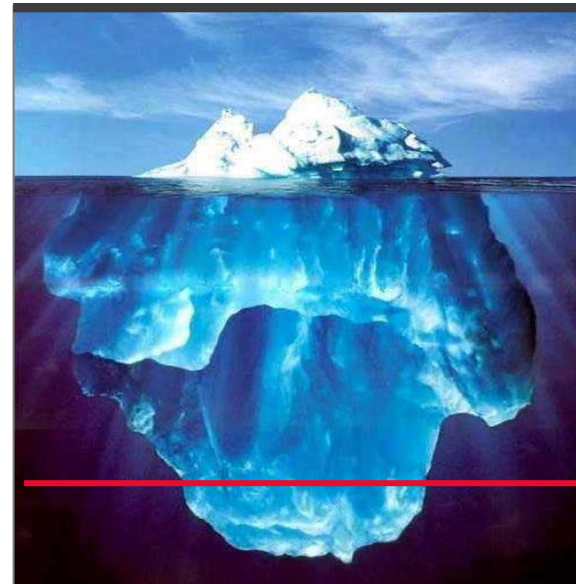
- What are the long-range information policy goals that remain unmet by current eGovernment strategies?
 - Effective, rather than just functional, access to all public information, through citizens' chosen channels (search engines)
- What new dynamics of information services and web publishing marketplace should governments consider in setting eGovernment strategy?
 - Search engines have become citizens' preferred gateway to all levels of public information

Toward more transparent government

- How can governments leverage open standards (Web protocols, data formats, etc.) to achieve public sector information goals and better interoperability?
 - By fully adopting the Sitemap protocol, a government can ensure that all public documents are only one search away
- Are there any significant differences between Web services created in commercial or other non-governmental environments, and those designed specifically for government application? If so, what can we learn from these differences about the unique requirements of eGovernment services?
 - To enable navigation over its corpus of information, and unleash its value, a government must fully embrace web search and structured data distribution

Join in

- Get informed:
www.google.com/publicsector
- Track progress by US federal government agencies at Federal Sitemaps wiki:
<http://tinyurl.com/3byhy7>
- Advocate for the Sitemap protocol
- Query: John Lewis (JL) Needham, Manager, Public Sector Content Partnerships, jlneedham@google.com



WEB
Searchable

DEEP WEB
Searchable