Legislated URLs for Permanent Documents on the Internet

Legislated URLs for Permanent Documents on the Internet

The Problem

Access to public documents is often difficult due to unclear document arrangements on public websites. Access is also often difficult due to shifting website organization. Links to public documents on the Internet are often broken because the names of the webpages, on which the public documents reside, have been changed.

Permanent Hyperlinks

Hyperlinks are the clickable, underlined, (usually) blue labels that take Internet users from the page that the user is currently viewing to another webpages on the Internet. Under the hyperlinks (or sometimes spelled out explicitly in the hyperlinks) are the 'http://' formatted URLs (Universal Resource Locators) that uniquely identify webpages on the Internet. These URLs give each webpage on the Internet a logical location that is independent of the physical location (server computer location) of the webpage.

Unfortunately, the URLs of webpages are easily changed, leaving the hyperlinks that pointed to the changed webpages broken. When the name (domain) of a website is changed, the location (URL) of all of the documents on that website changes. For each document with a changed URL, all of the hyperlinks, throughout the Internet are then broken. Even the simple reorganization of a public website can break the hyperlinks to all of the public documents on that website.

Legislated Permanent URLs

To implement legislated permanent URLs, it is not necessary to change anything on the Internet.

Legislated permanent URLs would be created by first establishing the domain name of the website. Almost every public entity now has a public website domain name. Because the domain name of the website belongs to the governmental entity that purchased the domain name (subject to copyright and trademark limitations). And, because that governmental entity is governed by the legislative body, the legislative body controls the permanent location of the website (within the namespace of the Internet).

The enabling legislation would first establish a fixed domain name for permanent public documents. Then, the enabling legislation would required that, once a permanent public document is given a permanent URL, an act of the legislative body would be necessary to change the URL of that document.

Permanent URL Structure

Within the website (domain name) fixed by the legislative body, identification of the governmental entity, and the document management division within the entity, can be given. Within the document management organization, the record series and record identifier can be given. To keep the website (and document management) organization as flat (simple) as possible, it is suggested that the organization structure be kept to three levels, that the record series be kept to three levels, and that the document division be kept to six levels.

This system will accommodate any number of levels of document organization, but too many levels will lead to confusion. Also, as organizations change over time, complex structures intended to mirror the organization structure will become outdated. Because the URL of the documents cannot change, the differences will have to be explained, but not modified. A simple logical document management structure, designed for permanent use, is recommended for the website.

Example of a Permanent URL:

[http://www.ci.LA.CA.US/Clerk/Records/Permanent/Engineering/technical/plans/35231/25/6] (This example is not implemented on the Internet.) In this example, the domain is http://www.ci.LA.CA.US, the organization is Clerk/Records/Permanent, the record series identification is Engineering/technical/plans and the document identification is 35231/5/6 representing plan number 35371, sheet 25, frame 6.

Numbers

Numbers can be assumed to be right justified. If desired, a '1' followed by a fill-string of zeros can placed to the left of each number to ensure that numbers sort in the same order whether the numbers are left or right justified.

Temporary Documents

This system will not work well for temporary documents. It is designed for permanent documents. Temporary (active) (in-process) documents require sophisticated databases for implementation. Because of the complexity of these databases, the database software (and associated data structures) is likely to last less than one century.

Revisions

Permanent documents require an act of the Legislative body to be changed or destroyed. Permanent documents embody the history of a society and provide evidence for future fair acts, accolades, and retributions. For these reasons, changes to permanent public documents (or even diminutions of their accessibility) constitute an attack on the fabric of the society in which the documents were created.

Providing ready access to public documents is a fundamental function of society and is also a basic foundation for society. References to, and comments on, the contents of public documents are encouraged as a basic vivifying force in society. Making available references to changes in documents, described as updates and revisions, are an important part of preserving and protecting public documents.

The update information should be inserted as the first page (revision history page) of the updated document (or of the first image set of the first page if pages are stored as separate files). The updated document should remain available at its original location (URL).

Updates to the update history page should be appended to the end of the text of the existing update history, with no changes to the existing text of the update history. If desired, all documents can start with a revision history page. The standard revision history page can be added to each document when the documents are first placed on the Internet.

To limit complexity, complicated documents like legal codes (or GISs, Geographic Information Systems), can be stored in their entirety once per year. A detailed history of revisions and cross-references for these complex documents can be kept separately from the permanent copy of the documents. These complex histories require sophisticated databases for implementation. Because of the complexity of these databases, the database software (and associated data structures) is likely to last less than one century.

Exposing the Document and Document Management Structures on the Internet

The hierarchy of the URL structure should be exposed to users on the Internet. Moving up a level (deleting the last element in a URL string)

should produce a list (visible to Internet users) of all of the elements at the level just viewed. For example, deleting the page number and moving up a level should produce a list of the pages of the document. (Pages need not be individually accessible. For some documents, such as engineering drawings, pages are commonly made individually accessible.) Individual pages can be served by a document management system and in this case the list of available pages should be synthesized when the query response is generated.

Metadata

At each level of the document structure, a metadata page could be created. This page would be displayed, in the document structure and in the document management structure on the Internet, as a file with the name 'metadata'. The metadata file would give the history and use of the organization structure, document structure, the document(s), or the parts of documents at that level of the structure.

The revision history document page is not part of the metadata page because the revision history must be kept tightly coupled to the actual permanent public document.

Private

This plan is for permanent public documents. It will also work for permanent private documents on a private intranet.

Summary

Legislating permanent URLs for each public document is simple to legislate and simple to implement. Legislating permanent URLs for each public document is a new and extremely effective means of providing an essential public function that is fundamental to the continued existence of society.

Appendix

References for the Internet namespace:

W3C: "The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential as a forum for information, commerce, communication, and collective understanding." [http://www.w3.org] (W**3, World Wide Web) [http://www.ICANN.org] (Internet Corporation for Assigned Names and Numbers) [http://www.InterNIC.net] (Inter Network Information Center) See [http://www.internic.net/faqs/authoritative-

dns.html] for a non-technical explanation of the Internet URL namespace.

Note to Readers

Updates and More Detailed Descriptions

When using the information in this article, please check the website [http://www.ArchiveBuilders.com] for updates. The version number of this article is just before the page number below. The website also has articles that provide more details on some of the terms and concepts in this article.

Comments

Please let us know how you like this paper, or if you had any questions. What would you like to see in the future? For more, and the most recent version of this article, please visit our web site at www.ArchiveBuilders.com.

Please send your comments via email to SteveGilheany@ArchiveBuilders.com. Tel: +1 310-937-7000. Fax: +1 310-937-7001. Also, please let us know where you saw this article.

Acknowledgements

Reprinted from Archive Planning, Volume 6, number 4, 2002, Archive Builders' analysis newsletter for document management.

 See www.ArchiveBuilders.com.

All trademarks are the property of their respective holders.

Note to Editors

Paper 22054v002

We will continue to update these articles as we get comments. Please contact us for the most current version before you publish. Also, please request permission to publish the article. Permission will be given freely for most purposes.

Steve Gilheany
Archive Builders
1209 Manhattan Ave
Manhattan Beach, CA 90266
Tel: +1 (310) 937-7000 Fax: +1 (310) 937-7001
[SteveGilheany@ArchiveBuilders.com]

Bio

Steve Gilheany, BA in Computer Science, MBA, MLS Specialization in Information Science, CDIA (Certified Document Imaging System Architect), AIIM Maser, and AIIM Laureate, of Information Technologies, CRM (Certified Records Manager, ARMA) has twenty years experience in document imaging and is a Sr. Systems Engineer at Archive Builders.

Author

Steve Gilheany is a Sr. Systems Engineer at Archive Builders. He has worked in digital document management and document imaging for twenty years.

His experience in the application of document management and document imaging in industry includes: aerospace, banking, manufacturing, natural resources, petroleum refining, transportation, energy, federal, state, and local government, civil engineering, utilities, entertainment, commercial records centers, archives, non-profit development, education, and administrative, engineering, production, legal, and medical records management. At the same time, he has worked in product management for hypertext, for windows based user interface systems, for computer displays, for engineering drawing, letter size, microform, and color scanning, and for xerographic, photographic, newspaper, engineering drawing, and color printing.

In addition, he has nine years of experience in data center operations and database and computer communications systems design, programming, testing, and software configuration management. He has an MLS Specialization in Information Science and an MBA with a concentration in Computer and Information Systems from UCLA, a California Adult Education teaching credential, and a BA in Computer Science from the University of Wisconsin at Madison. His industry certifications include: the CDIA (Certified Document Imaging System Architect) and the AIIM Master (MIT), and AIIM Laureate (LIT), of Information Technologies (from AIIM International, the Association of Information and Image Management, www.AIIM.org), and the CRM (Certified Records Manager) (from the ICRM, the Institute of Certified Records Managers, the official certifying body for ARMA International, the Association of Records Managers and Administrators, [www.ARMA.org]).

Contact:

SteveGilheany@ArchiveBuilders.com
Tel: +1 (310) 937-7000 Fax: +1 (310) 937-7001

For more information, courses, and papers:

 [http://www.ArchiveBuilders.com]