

# Website Archiving Whitepaper

March 2011

Website Archiving

## CMS Version Control & Website Archiving

Regulatory bodies from Government to Advertising Standards and Industry watchdogs have been increasingly focusing their attention on the digital communications displayed across corporate websites.

As a result there is now a growing set of regulations governing the secure capture, storage and retrieval of the historic content displayed to website users. In addition to the regulatory requirements to archive website content there is also a very compelling best practice case.

Ensuring that website content is securely captured, archived and readily retrievable in the event of a dispute or litigation has become an imperative. However, while the website has become the primary channel for customer communication for many organizations, the technology and process of managing website content exposes significant vulnerabilities.

As a consequence, company Risk Managers are increasingly focusing their attention on measures to make website content compliance essential corporate practice.

### Contents

Risk & Content Complexity	2
The Original Rendered Page	3
Archiving Best Practice	4
Is Your Business Affected?	5

# Website Archiving Whitepaper

March 2011

## Risk & Content Complexity

Website content is multi-faceted E.g. Flash videos, Ajax, Silverlight and other rich media in addition to text and PDF downloads and is often as dynamic as it is ephemeral. With the advent of Multi-Variant behavioral targeting, personalised for specific users or user types the static web page is almost becoming an anachronism.

Furthermore, content is supplied from multiple data sources both internal and external, often including 3<sup>rd</sup> party advertisers. Adding to the potential variance is that different content can be displayed to different user types, accessing a website in different geographies and different browsers.

The problem for Risk Managers is how to reliably capture all of the content displayed across a website should it be necessary to prove what was actually displayed at any given point in the past. Information Security is an extremely important concern for all organisations and the question of what was 'displayed' is a key point in determining an appropriate data collection strategy.

It is almost irrelevant to prove what content flowed through a content management system because this internal capture ignores the nature of the internet and the myriad number of issues that could have impacted the content that was actually delivered to a web users screen.

To properly mitigate the risk of non-compliance and potential litigation, an appropriate **Information Governance Policy** must focus on the capture, archive and retrieval of the actual content displayed to web users.

The user's browser has become the only true place where all content converges and the web page is properly represented.

# Website Archiving Whitepaper

March 2011

## The Original Rendered Page

Restoring earlier web versions is notoriously difficult.

The rebuild of previous website versions also requires legacy browser versions to be archived in addition to operating systems including all associated web application data such as product databases and code. It is fraught with complications and page elements are often found to be missing or otherwise unavailable. It can take a considerable amount of time for IT to complete what is often an unreliable process.

Recorded website content is stored according to corporate and regulatory data retention policy and it should be simple, fast and accurate to search for specific web pages and specific textual content across an entire website archive for a stated period of time.

Most Content Management Systems only provide a reconstruction of a page that was delivered in the past. They do not show the actual page viewed by a user. Sole reliance on CMS version control cannot therefore be relied upon as an effective measure to ensure web content compliance.

The content displayed and available in the original rendered page in a user's browser is the only page that counts for legal purposes.

### Unstructured Content

Social Media such as Blogs, Facebook and Twitter all present new challenges for compliance teams.

Missives in such places tend to be of an unstructured nature but are still considered to be customer communication and are still subject to the same regulatory controls and carry exactly the same litigation risk as corporate communication. The fact that they lie outside an organizations physical domain and are not subject to the same authoring controls increases the risk of non-compliance.

The need to capture, archive and retrieve Social Media content is therefore great given its rapidly changing nature and the inherent limitations on internal authoring control.

# Website Archiving Whitepaper

March 2011

## Archiving Best Practice

Information Security and the protection of personal information is a business critical and legislative issue.

Any archiving policy for web content must ensure the following:

- a) Secure data retention
- b) Immutable Data
- c) Authenticated access control

### Other points to consider:

It is essential that at any point in time an **immutable copy** of the web content can be **retrieved**.

This content should be **certified** to ensure its authenticity with the time and date of capture and security hashes.

There should be **multiple copies** stored in multiple locations with additional measures to verify that original records are unchanged.

Placing the responsibility for content capture, archive and retrieval with an **independent 3rd party** eliminates the potential conflict of interest and risk of data being deliberately modified.

# Website Archiving Whitepaper

March 2011

## Is Your Business Affected?

Website Archive has customers across multiple sectors and they all archive websites in addition to maintaining records through Content Management Systems. All of our customers care about the soaring cost of litigation and need for regulatory compliance:

- Finance, Banking & Insurance
- Health & Pharmaceuticals
- Gaming & Gambling
- eCommerce, Travel & Retail
- Media

Where there is a regulatory need to demonstrate exactly what historic content was displayed in its original form there is a need to externally archive web content. In addition, where there is the potential for litigation over disputed web content there is a need to maintain an indisputable record of exactly what was displayed to users. The cost of not maintaining proper records can lead to fines, reputational damage and costly legal challenges.

External web site capture and archiving offers a cost-effective solution to prove what content was displayed to users at an exact point in time.

Website: [www.website-archive.com](http://www.website-archive.com)

Email: [info@cloudtesting.com](mailto:info@cloudtesting.com)

### United Kingdom

The logo for Cloud Testing, featuring the words "Cloud" in orange and "Testing" in blue, with a stylized cloud graphic behind the text.

Cloud Testing Limited  
PO Box 143, Alesford, SO24 4AF  
United Kingdom

(0844) 736 2602

### United States

The logo for Cloud Testing, featuring the words "Cloud" in orange and "Testing" in blue, with a stylized cloud graphic behind the text.

Cloud Testing Limited  
2885 Sanford Ave SW #13570  
Grandville, MI 49418, U.S.A.

1 - 888 - 851 - 4988

