

# Cloud Testing Whitepaper

Website Archiving

Digital Content Compliance

February 2011

Website Archiving

## Digital Content Compliance

It often takes a while for laws to catch up with new technology and best practice. This is certainly true in the area of digital content. This paper concentrates on Financial Services, however the subject is just as applicable to most areas of e-Commerce and e-Business.

### Contents

The Legal Position	2
How to Comply	3
Other Points to Consider	4
Conclusion	5

# Cloud Testing Whitepaper

Website Archiving

Digital Content Compliance

February 2011

## The Legal Position

### The United States

In the United States, **FINRA** (Financial Industry Regulatory Authority) in conjunction with the Securities and Exchange Commission (**SEC**) issue guidelines on the storage and archiving of digital content. To be compliant with the guidelines, storage, retrieval and surveillance solutions need to be in place. These state that copies of websites need to be archived and stored in an instantly retrievable format for six months, with an additional two years of off-line storage.

FINRA recently also issued Regulatory Notice 10-06 covering social networking. In overview:

- Micro-blogging sites like Twitter are regarded as advertisements.
- Email or 'instant messenger' that are sent to more than 25 prospects is considered as sales literature.
- Email or instant messenger sent to less than 25 prospects within 30 days, or a single destination, or an unlimited number of retail clients is considered correspondence.
- Social media sites like Facebook or LinkedIn that have password protection are regarded as sales literature.
- Discussions in chat rooms are considered as public appearances.

FINRA was created in July 2007 with the consolidation of the regulation enforcement and arbitration functions of the New York Stock Exchange with the National Association of Securities Dealers (**NASD**).

### The United Kingdom

In the UK, the Financial Services Authority has its COBS (Conducting Of Business Sourcebook), with COBS 4 covering "Communicating with clients, including financial promotions" and COBS 5 covering "Distance communications", both of which state the information that needs to be stored.

The UK is less clear about the specific requirements than the US but there remains an obvious need to be able to accurately recall all aspects of digital content on any given day in the past and to prove that specific content complied with any regulations relating to it.

It is expected that the FSA will issue more specific requirements in forthcoming COBS releases inline with FINRA regulations.

# Cloud Testing Whitepaper

Website Archiving

Digital Content Compliance

February 2011

## How to Comply

There are three main methods of archiving a website; taking screenshots of the pages, saving the content and images to be 'replayed' later, or simply backing up the content from the web server or Content Management System. All have their merits as described below:

### Screenshots:

As the expression goes, a picture says a thousand words, and this is exactly the case with the screenshot approach. Pages are visited using real browsers such as Internet Explorer or Firefox and a screenshot taken of exactly what a visitor would see at the time.

The screenshot is then stored and indexed, usually in an on-line service for later retrieval.

The advantage of this approach is that you end up with a simple easy to use archive which requires no special software to view. It is also future proof; as your website is captured in the current browser, rather than having to save browsers and operating systems to view the old website. Can you imagine in seven years time when who knows how we'll be accessing websites, trying to find a copy of Internet Explorer 6 and Windows XP?

### Capture and Replay:

With this methodology the website is visited by a web archiving agent and all content that makes up the page (HTML, CSS, JavaScript, Images etc.) is captured and stored. When a user later wants to view that page, it is rebuilt from the saved content using a browser.

This approach has the advantage of making the website 'live', i.e. you can click on the links to move between pages (assuming they have been correctly re-written).

There are a couple of negative points – as the content needs to be edited to make it 'clickable', it isn't an exact copy of what was served by the server and as seen by 'real users'. You will also need to keep old versions of browsers to view it in.

### Content Backups:

Content is backed up either directly as files from the web server or exported from the Content Management System (CMS) and stored. This needs to be done by someone with access to the content and so cannot easily be outsourced.

This method is the simplest, or should be, however the need to store the backup that can be re-read with future versions of CMS software, along with ensuring the media is stored safely can mean many headaches.

The main drawback of this method is that in today's world, web content comes from numerous sources and the only true place where the site content is properly represented is in the browser.

# Cloud Testing Whitepaper

Website Archiving

Digital Content Compliance

February 2011

## Other Points to Consider:

### Real Browsers vs. Archiving Agents

Some web servers return different pages to 'web archiver' or non standard browsers than they would in response to regular browser requests. It is important to capture and store what a 'real user' would have seen on that day.

### Login/Password areas

Many offerings only allow distinct pages to be captured, rather than allowing capture of end to end transactions, or pages that require the user to log in or select for example a post/zip code before continuing.

### Capturing all information streams

In addition to traditional 'web content' produced on the website, thought should be given to other content such as file downloads (PDF product brochures, Excel pricing files etc.) as well as RSS feeds of blogs and information.

### ESCROW

If an external service is used, it is good practice to store a second copy in an ESCROW facility.

# Cloud Testing Whitepaper

Website Archiving

Digital Content Compliance

February 2011

## Conclusion

If you are a Financial Services organisation with an online presence in the US then you will already be required to archive ALL digital content displayed across All Web Assets including Social Media. For the UK it isn't so clear but it is certainly best practice and is expected to become mandatory in the near future.

There are a number of approaches to archiving digital content but the only true place guaranteed to capture all content delivered from multiple sources and devices is in the client browser. The most robust methodology to reliably capture the actual content displayed to end-users is through the Screenshot approach with the essential requirement that data is well indexed and therefore easy to interrogate and securely stored.

### United Kingdom



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

United Kingdom: **(0844) 736 2602**

International: **+44 1962 736699**

Email: **info@cloudtesting.com**

Website: **www.cloudtesting.com**

### United States



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

Toll Free: **1 – 888 – 851 – 4988**

International: **+44 1962 736699**

Email: **info@cloudtesting.com**

Website: **www.cloudtesting.com**