Archiving Websites (Sessions 1, 2 & 3)

Questions	Various Answer(s)	Example(s) Cited (if any)
Purpose? [1]	• To capture content	
	• To capture the look and feel	
When?	• When major changes/upgrades occur	
	• Monthly	
What to save?	• Consider whether website has intrinsic value • Content more important than look and feel for many companies • Decide what is important based on priorities, funding, and needs • Use appraisal guidelines • Determine how many levels of site to preserve • Archives may preserve component parts and content and not entire site	• Press releases • Senior executive speeches • Annual meeting/investor presentations • New product/service introductions • Brand-related information and developments
Why?	• Documenting public-facing presence of company, information, etc.	
	• Capture of past styles/look and feel	
	• Legal reasons	
Who has responsibility for the capture?	• IT (in-house) facilitated	
	• Archivist	
	• Company website manager	
	• Shared responsibility of various departments	
	• Third party/vendor	• Hanzo Archives • ArchiveIT • Wayback Machine (no contracted responsibility)
How to do it options	• Print out archival pages (must be cataloged to facilitate searching) • IT-facilitated • Save on network drive with back-up • PDF used to capture text (but does not work with application-driven sites) • PDF is "safe" format (relatively stable) • Screen grabs • Adobe Acrobat Professional (can capture simple HTML sites) [2]	• Microsoft SharePoint • One IT department suggested using a "virtual machine" (??)

---

[1] See also 'Who uses archived website pages?'

	• "Archives Web Format" from Internet Explorer (unreliable results) • For complex or application-driven sites capture solution should be automated within IT/IS or flagged content should be preserved without Archives' manual intervention[3] • Ensure or contract for appropriate archival preservation and security of captured material	
Who decides what to capture?	• Archivist • Legal[4] • Marketing • IT • Web Team • HANZO (External) • Internet Archive (External) • Tessella ?? (External)	• Should be included on records management retention schedules
Who uses archived website pages?	• Legal • Communications (External & Internal) • Public Relations • Brand/Account Managers • Archivist	• Trademark defense • Telling company story • Employee publications/-communications history • Anniversaries • History of product/brand • Reference (e.g., history of website)

---

[2] Not a usable solution for data sites, SharePoint, etc.

[3] Advance flagging of specific website content intended for preservation by departments, records manager, and/or archivist is important as daily review is rarely practical for Archives.

[4] Legal implications need to be addressed. Determine what is the 'original' version which is especially important when archival copy captured is digital version of site used in partnership with a printed version.

Examples of Companies Currently Archiving Websites

Company	Archival Software/ Method	Functions	Issues
Coca-Cola	HANZO	- Delivers functionality and flash of site - Results in WAC file which is not proprietary - Captures site as it was at time of capture - Contract includes quality control - Will drill into the site as far as budget will cover - Files retained on hard drive / digital asset management system - Quarterly snapshot of 20 sites - Not global and not all brands, but many Coke brand sites worldwide - Mostly static	- Number of externally facing websites - US sites and international sites - Cost - Time/ staff time - Firewall issues - Appraisal – What do you want to capture? How far do you want to drill down? - Ownership - Social Media capture – How far back to capture? [5] - Access - What tools to use: screenshots, PDF/ PDF-A, Snagit, printing
Cargill	Hanzo	*See above*	*See above*
Nationwide	Web curator tool from Internet Archive ('Wayback Machine')	- Captures intranet site - Open Source tool proposed by Nationwide IT - Recently changed intranet to single sign-on so have to start over	Use of Open Source tools
Assoc. of Amer. Medical Colleges	Grant- funded project is pushing interest		
Target		Prints out 'interesting/archival pages'	- Time-consuming - Too many tools needed to access – wiki, SharePoint, etc.

---

[5] Coca-Cola captures the Archives' blog. Legal issues remain surrounding preservation and use of other people's comments posted on blogs.