

Preservation In Practice - What Are The Options? A Case Study

Decoding the Digital
British Library Preservation Advisory Centre

27 July 2010



- o Joint archive service for the 5 West Yorkshire metropolitan authorities:
 - o Bradford City Council
 - o Calderdale Metropolitan Borough Council
 - o Kirklees Metropolitan Borough Council
 - o Leeds City Council
 - o Wakefield Metropolitan District Council
- o Largest local authority record office outside London:
 - o Over 4200 cubic metres of traditional archive material
 - o In 2008/09: more than 13000 personal visits, 25000 document productions, answered nearly 26000 enquiries
- o Public service points in each West Yorkshire district, plus:
 - o Commercial records management storage and administrative headquarters in Morley (near Leeds)
 - o Three archive outstores, including purpose-built electronic records room at Morley



“A decidedly analogue approach to a digital problem”



- Nobody has all the answers
- Doing something is better than doing nothing – start small
- Knowing what you’ve got (or are about to get) is a vital first step
- “If we try we may fail, if we don’t try we will certainly fail”

WYAS Digital Archives Working Group

Digital Archives Policy

available at <http://www.archives.wyjs.org.uk/wyjs-archives-policies.asp>

fostering awareness and support

creating a framework for practical work

Practical Digital Archives Work

Software testing, using sample objects

Survey of digital records already held by the Service (first documented digital accession received 1999)

Digital deposit form and guidelines

A digital media recognition guide for staff

Advocacy and training

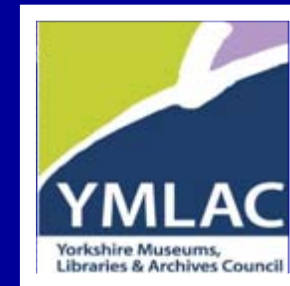
Small, non-specialist, staff working group

3 archivists, plus records manager

Flexible, but limited, support from West Yorkshire Joint Services' IT department

MLA Yorkshire: A Case Study

- Regional Museums, Libraries and Archives partnership organisation for Yorkshire
- Wound up December 2008
- Developed out of the Yorkshire Museums Council (YMC), with a continuous history back to 1963
- 1.4 m³ paper, 1963-2008
- 80GB+ digital, c2002-2008



Negotiating a Hybrid Deposit

- Very short timescale – no time to try out software tools
- Depositor had not considered digital records to be ‘archives’
- Digital content may not be held locally (website, e-bulletins)
- ‘Self curation’ of electronic records by staff - closer to personal archive than organisational
- Confidentiality
- Conflicting interests
- Positives:
 - No obsolete hardware
 - Predominantly current office file formats
 - Helpful IT support
 - ‘Dowry’ to pay for boxes, portable hard drive etc.



Collecting Digital Content

- New digital deposit documentation
- 'Grab & Run' copying - no time to complete appraisal prior to transfer
- Transfer using 1TB USB external hard drive & FTK Imager Lite (<http://www.accessdata.com/downloads.html>)
- Capture 'original order' & generate checksums
- Problems encountered:
 - Long file paths/file names
 - Need all users to be logged out of corporate network
 - Don't underestimate the time needed to transfer large amounts of data

Processing Digital Content: Procedure

- Transfer to standalone PC
- Authenticity verified using checksums
- Virus check – twice – using AVG Anti-Virus (Kaspersky on corporate network)
- Use spreadsheets to analyse file listings / check for duplicates
- Appraisal

Processing Digital Content: Tools



- Pre-accession appraisal tools
 - Karen's Directory Printer
<http://www.karenware.com/powertools/ptdirprn.asp>
 - Directory structures
<http://www.paradigm.ac.uk/workbook/record-creators/capturing-directory-structures.html>
- Forensics software
 - FTK Imager Lite
<http://www.accessdata.com/downloads.html>
- Checksums
 - MD5, SHA-1
- Virus checking
 - Free options: AVG Anti-Virus, Avast! etc.
 - Software and procedures
- File format identification
 - PRONOM
<http://www.nationalarchives.gov.uk/pronom>
 - DROID <http://droid.sourceforge.net>
- Paradigm Workbook
 - <http://www.paradigm.ac.uk/workbook>

Portrait of a Gentleman Techie, after Thomas de Keyser

<http://www.flickr.com/photos/notionscapital/2496369741>

Appraisal in a Digital World

Microsoft Excel - G_and_H_complete_file_List.xls

File Edit View Insert Format Tools Data Window Help

Save As... 75% Arial

Type a question for help

A6824 FILE

	A	B	C	D	E	F	G	H	I	J
		Filename/Filepath	Extension	Size	Date Created	Date Last Modified	Date Last Access	File Vers	MD5 Hash	SHA-1 Hash
6003	FILE	Digital Strategic Leadership Paper.doc	.doc	46,592	12/11/2007 16:30	12/11/2007 16:30	17/12/2008 09:40		22295C730198C7E5020DE138D4F90CB	6A4DF29DB4
6004	FILE	Big Day De-brief.doc	.doc	35,329	13/11/2007 15:20	20/11/2007 12:55	17/12/2008 10:06		222A318D35906ECF8BA921767B9AE330	F3F6F5DCE
6005	FILE	Draft M&G Educ Strat PAC 0204 JW.doc	.doc	59,392	23/09/2008 10:19	23/09/2008 14:19	17/12/2008 10:00		222B1D208F3D77BF8B7019E8727E70	922A4419AE
6006	FILE	Learning Managers Meeting 15 June 06.doc	.doc	87,040	14/08/2007 15:05	26/06/2006 17:02	17/12/2008 09:37		22323D27A6B2E80462C36D6364A5E3E	6EFD982B8
6007	FILE	Feeling the V'sth.doc	.doc	138,752	19/08/2007 19:03	22/12/2004 10:55	17/12/2008 09:41		2232F4FC8A19ECC3C9060042455F2C6A	104B611765F
6008	FILE	Accreditations 2007.doc	.doc	56,320	21/11/2007 12:10	23/11/2007 12:21	17/12/2008 10:23		2237C7434DCC64044EBDCC8E2A50DBA5C	5641776D3F1
6009	FILE	ymiaedatreport-part3(2mar).doc	.doc	295,696	19/08/2007 18:58	13/03/2006 08:12	17/12/2008 09:36		223BD7B4055733B27928558D4EBDF44F	1D03F790AF6
6010	FILE	Health policy Brief.doc	.doc	186,880	19/08/2007 19:02	17/05/2006 15:36	17/12/2008 09:41		223E59DBF37DDDB06D476C853D000494	D0B964B0E9
6011	FILE	Transfer to MLA Control description 10.doc	.doc	31,744	19/11/2008 09:27	19/11/2008 09:27	17/12/2008 09:35		2240905E83BA40B638E2A76F05E5BA03	3094B402E38
6012	FILE	SVC Agenda 25Nov04.doc	.doc	38,400	17/11/2004 00:03	19/11/2004 09:29	17/12/2008 10:23		2240462C3044E996B633DEF1C59B91	40F AF 484E2I
6013	FILE	SVC Agenda 25Nov04.doc	.doc	38,400	17/11/2004 00:03	17/11/2004 00:03	17/12/2008 10:23		2240462C3044E996B633DEF1C59B91	40F AF 484E2I
6014	FILE	SVC Agenda 25Nov04.doc	.doc	38,400	24/09/2007 11:46	19/11/2004 09:29	17/12/2008 09:41		2240462C3044E996B633DEF1C59B91	40F AF 484E2I
6015	FILE	SVC Agenda 25Nov04.doc	.doc	38,400	24/09/2007 11:46	17/11/2004 00:03	17/12/2008 09:41		2240462C3044E996B633DEF1C59B91	40F AF 484E2I
6016	FILE	Framework 4 Future VMLACrespoense F.doc	.doc	31,744	19/08/2007 18:56	24/03/2004 10:57	17/12/2008 09:35		224462A98F2EAF8A8A82B9B5C3AF2	B0CB204EDF1
6017	FILE	PDNet Gen booking form.doc	.doc	25,600	19/08/2007 18:50	05/02/2006 10:47	17/12/2008 09:34		2246086CDFDA850A5A7122BC3E95019	F2C50F78B81
6018	FILE	PDNet Gen booking form.doc	.doc	25,600	19/08/2007 18:50	05/02/2006 10:47	17/12/2008 09:34		2246086CDFDA850A5A7122BC3E95019	F2C50F78B81
6019	FILE	Finance assistant.doc	.doc	59,392	02/11/2004 13:45	20/05/2008 13:20	17/12/2008 10:02		224AB624DC73EF033087190140B1C0D5	6827477A18F
6020	FILE	Teenagers and stories project Yorkshire.doc	.doc	26,112	14/08/2007 15:04	19/09/2006 07:48	17/12/2008 09:37		225E8D75D1300C5822974BFBEDADD39A	D6D3DAAB83I
6021	FILE	Strategic Commissioning 2004-6 MAP1.doc	.doc	31,232	03/05/2005 09:27	29/07/2004 08:50	17/12/2008 10:02		2258D7F6402AADF7C5D12F367D695F7	8C358775D5I
6022	FILE	Strategic Commissioning 2004-6 MAP1.doc	.doc	31,232	14/08/2007 15:00	29/07/2004 08:50	17/12/2008 09:39		2258D7F6402AADF7C5D12F367D695F7	8C358775D5I
6023	FILE	CMSC005Mins.doc	.doc	34,816	14/08/2007 15:01	05/01/2005 17:12	17/12/2008 09:39		225CA8D355ABCE8EB8D75912F2E17482	2E7F16583A5I
6024	FILE	Digital Film Archive for Yorkshire CD A1.doc	.doc	45,568	03/12/2008 11:12	19/12/2005 11:34	17/12/2008 09:59		227003FDC45E95E44A3CA281261C1937	560EAE38B5
6025	FILE	Digital Film Archive for Yorkshire CD A1.doc	.doc	45,568	03/12/2008 11:14	19/12/2005 11:34	17/12/2008 09:59		227003FDC45E95E44A3CA281261C1937	560EAE38B5
6026	FILE	application ChangeUp.doc	.doc	122,880	19/08/2007 19:02	07/01/2005 14:50	17/12/2008 09:39		227B327567950D1A00C3617EEABCB8	FBAE88CCB
6027	FILE	Membership Info & Criteria May03.doc	.doc	214,528	04/06/2003 10:03	14/05/2003 15:07	17/12/2008 10:24		228251926CB3045FDE6C13387EA0E202	8F550689CFI
6028	FILE	07.12.05 notes.doc	.doc	24,576	30/05/2008 10:42	07/12/2007 09:29	17/12/2008 09:42		2287BC1F6104FE7DB179C4674DB78284	9462CFD28C
6029	FILE	07.12.05 notes.doc	.doc	24,576	07/12/2007 09:24	07/12/2007 09:29	17/12/2008 09:42		2287BC1F6104FE7DB179C4674DB78284	9462CFD28C
6030	FILE	funding and sustainability PPSG paper Sept.doc	.doc	142,848	19/08/2007 18:57	09/09/2006 09:57	17/12/2008 09:36		22894AC579EBAAD72A026A211F5ECADD	320458CE283
6031	FILE	Charities Commission Direct initial lette.doc	.doc	26,112	07/06/2008 12:09	07/06/2008 13:39	17/12/2008 10:02		228DBAB63705E2AF76A6E01D7F6439	588F1926F41
6032	FILE	Hubbed notes.doc	.doc	33,280	19/08/2007 19:02	05/07/2003 10:32	17/12/2008 09:42		229B4C76B48AD7438537D635FFD39F63	319F3A1AF61
6033	FILE	Agenda 21May 02.doc	.doc	20,992	02/12/2008 17:54	14/05/2002 16:04	17/12/2008 10:00		229D5A4DC24723593231289C5EDC761E	3C8B08A709
6034	FILE	Minutes 08.05.doc	.doc	59,904	16/06/2008 09:42	17/06/2008 14:40	17/12/2008 09:43		22A06B4AD65530C0537005CECE09B47	6D71706C58C
6035	FILE	Logistics.doc	.doc	24,064	06/03/2006 15:19	06/03/2006 15:19	17/12/2008 10:03		22A19E759C8CF2E54D922B6E1AFC A52	C3E83517F2F
6036	FILE	Accred Introduction letter.doc	.doc	796,160	19/08/2007 19:11	17/01/2006 09:21	17/12/2008 09:42		22A509F4A44F A333297AE6566B23493	0C34E163A
6037	FILE	Treasurers report re subscriptions.doc	.doc	22,528	01/12/2008 16:36	15/11/2001 16:53	17/12/2008 09:59		22A67F8E77CC01362F845E0A39A36629	10D89E596F1
6038	FILE	Treasurers report re subscriptions.doc	.doc	22,528	01/12/2008 16:33	15/11/2001 16:53	17/12/2008 09:59		22A67F8E77CC01362F845E0A39A36629	10D89E596F1
6039	FILE	26.05.06 LdPS Consultation Letter.doc	.doc	218,624	19/08/2007 19:13	31/05/2006 09:35	17/12/2008 09:46		22A9F16D9D0329CE2919E6F4240D1AA	5F50E376B17
6040	FILE	Item 15 Dates of future meetings.doc	.doc	24,576	23/02/2008 17:31	04/03/2004 11:27	17/12/2008 09:59		22B992F6FE1F5F02544D08F564B87F	37D18EAFABE
6041	FILE	Raw minutes 10.3.06.doc	.doc	29,696	14/09/2007 14:11	14/03/2006 16:24	17/12/2008 09:42		22BE683E2952C3DC5857269AD97FC50	722EACCB8
6042	FILE	A tasonomy of MDO activity and interve.doc	.doc	35,840	19/08/2007 19:13	14/05/2007 11:52	17/12/2008 09:44		22C4CB49BE8E7EA9F7F330F0FBF8A2	82AA2F95A5I
6043	FILE	jmevents.doc	.doc	20,480	04/11/2008 13:03	02/08/1999 10:43	17/12/2008 09:59		22C8C915007F797CEC63143DF1ABE0DF	1E48B87E6E
6044	FILE	jmevents.doc	.doc	20,480	30/10/2008 13:39	02/08/1999 10:43	17/12/2008 10:22		22C8C915007F797CEC63143DF1ABE0DF	1E48B87E6E
6045	FILE	Performance Review Documentation1.doc	.doc	63,488	19/08/2007 19:13	21/12/2005 10:41	17/12/2008 09:46		22DA1C0B4E77B27E420294DEAD27069A	D803F770FC2

Ready

Start

Inbox - Microsoft Outlook

Microsoft PowerPoint - [...]

Final Directory Readings

Microsoft Excel - G_and_H...

NUM

16:05

MLA Yorkshire Website

The screenshot shows a web browser window displaying a crawl report from Archive-It. The browser is Internet Explorer 7. The URL is <https://archive-it.org/archiveit/partner/crawl/report/mimeType.seam?accountId=3408&crawlJobId=7700>. The page header includes the Archive-It logo and the text "ARCHIVING THE INTERNET FOR FUTURE GENERATIONS". The main content area shows the "File Types" tab selected, displaying a table of file types and their corresponding URIs and byte sizes. The table has three columns: "File Type", "Uris", and "Bytes". The data rows are: text/html (107101 URIs, 1.4 GB), application/pdf (47 URIs, 39.74 MB), application/msword (38 URIs, 8.88 MB), image/jpeg (23 URIs, 401 KB), image/gif (17 URIs, 177 KB), and text/css (6 URIs, 0.1 KB). A "Download" link is visible next to the table. The footer of the browser window shows the taskbar with icons for Start, Internet Explorer, Microsoft Outlook, Microsoft PowerPoint, and the Archive-It Crawl Report window.

Archive-It: Crawl Report [20081202031532930] - Windows Internet Explorer provided by IE7 WYJS

File Edit View Favorites Tools Help

Google C Go Bookmarks 5 blocked Check AutoLink AutoFill Send to Settings

https://archive-it.org/archiveit/partner/crawl/report/mimeType.seam?accountId=3408&crawlJobId=7700

Archive-It: Crawl Report [20081202031532930]

West Yorkshire Archive Service

ARCHIVING THE INTERNET FOR FUTURE GENERATIONS
COLLECT IT, MANAGE IT, SEARCH IT...ARCHIVE-IT

Welcome yorkshire Help Admin Settings Archive This! Templates Logout
Reset Password

Home Collections Crawls Reports Access Help

MLA Yorkshire

One-Time [ID #20081202031532930]

Crawl Report
Started: December 2, 2008 3:15:32 AM
Completed: December 5, 2008 3:34:09 AM

<< Back

Summary Hosts Seed Status Seed Source File Types

File Types

The file type report shows how many URLs were captured for each specified file type as well as total byte size of the archived documents.

File Type	Uris	Bytes
text/html	107101	1.4 GB
application/pdf	47	39.74 MB
application/msword	38	8.88 MB
image/jpeg	23	401 KB
image/gif	17	177 KB
text/css	6	0.1 KB

Download

Help on Reports

Archive-It provides four post crawl downloadable reports to assist partners in analyzing and understanding the data that has been archived. The reports can be downloaded as CSV files so they can be easily opened in Excel.

- [Learn more about reports](#)
- [Learn more about the host report](#)
- [Learn more about robots.txt](#)

Start Inbox - Microsoft Outlook Microsoft PowerPoint - [...] Archive-It: Crawl Rep... MLA Yorkshire 17:12

Archive-It - <http://www.archive-it.org/>

UK Web Archiving Consortium <http://www.webarchive.org.uk>

MLA Yorkshire Website

MLA Yorkshire - Windows Internet Explorer provided by IE7 WYJS

File Edit View Favorites Tools Help

Google G Go Bookmarks 5 blocked Check Look for Map AutoFill Send to Settings

http://wayback.archive-it.org/1237/20081202031535/http://www.mlayorkshire.org.uk/webdav/harmonise?Section/@id=4511

MLA Yorkshire

You are viewing an archived Web site, archived on 3:15:35 Dec 02, 2008, that is part of a collection of archived websites created using [Archive-It](#). The information on this web page may be out of date. External links, forms, and search boxes may not function within this collection. [[hide](#)]

MLA MUSEUMS, LIBRARIES AND ARCHIVES YORKSHIRE

Search

Home
[News & Events](#)
[The Region](#)
[Archives](#)
[Libraries](#)
[Museums](#)
[Resources & Publications](#)
[Training](#)
[Careers](#)
[Networks & Groups](#)
[About MLA Yorkshire](#)
[Contact Us](#)

MLA Yorkshire is the lead strategic agency for museums, libraries and archives in the region. We work to improve people's lives through access to these collections and resources – building knowledge, supporting learning, inspiring creativity, and celebrating identity.

> Latest News

- MLA Appoints New Area Director North Accreditation - Final Update
- groupthing.org: Engaging Young People with Reading
- Their Past Your Future – Final Round Year 3 Grants
- Strategic Commissioning - The Future

Subscribe to our **E-Info Service**

> Resources and Publications

- Free Family Learning in Museums Resource Pack
- Yorkshire Museums Support Consultation - Final Report
- Building a Diverse Workforce: Libraries and Archives in Yorkshire
- Yorkshire GSO Pilot Report and Case Studies

> Conferences & Seminars

- Do Games Tell A Story? 2 December 2008
- EUCLID: CulturEuro Seminar, 2 December 2008
- Tactile Graphics 2008: 3-5 December 2008
- MA: Designing for Disability, 5 December 2008
- Just Playing? The Role of Play in Children's Lives: 14 January 2009

> Training Events

- SINTO Courses - Promoting Reading to Young People 04-12-2008
Sheffield Hallam University
- Object Handling Essentials 11-12-2008
Yorkshire Museum, York
- Disability Awareness Workshops 2009 14-01-2009

Programmes

- [2012 Olympics](#)
- [Inspiring Learning for All](#)
- [Museums & Archives Alive!](#)
- [Museums Development](#)
- [Skills for Life](#)
- [Widening Participation](#)

Choose a project...
Go

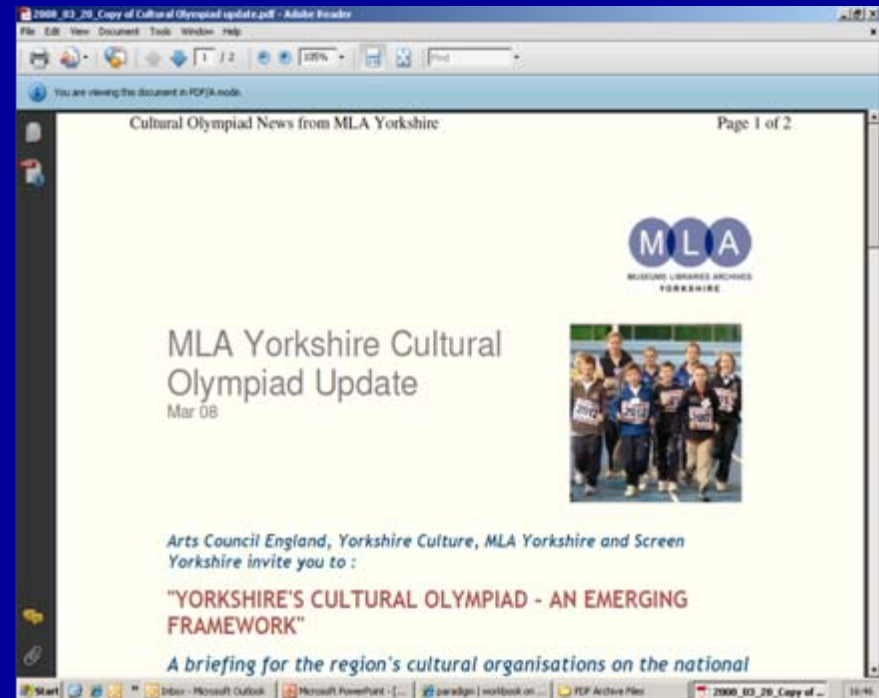
Inspiring Yorkshire

Download the latest edition of our [newsletter](#)

Done

Start Inboxes - Microsoft Outlook Microsoft PowerPoint - [...] Archive-It: Access - Way... MLA Yorkshire MLA Yorkshire - Wind... 17:13

MLA Yorkshire E-Bulletins - A Practical Example



“I never have these types of problems with digital”

- **Don't panic at the format and overcomplicate – look for simple solutions that are ‘good enough’**
 - Make multiple copies
 - Research software tools and preservation services (eg PRONOM <http://www.nationalarchives.gov.uk/pronom>)
 - Archival theory is still relevant. Develop new procedures alongside traditional workflows.
- **Don't start with long obsolete material – it'll be much harder**
- **Deal with reality, not with ‘what ifs’.**
- **Do not underestimate your expertise with regards to cultural factors/interpersonal skills vital to digital preservation** (eg privacy concerns, depositor liaison)

The Preservation Chain

“Digital information lasts forever, or five years – whichever comes first” *Jeff Rothenberg*

You don't need to aim for 'forever'.
Just the next generation of recordkeepers.



Alexandra Eveleigh
AHRC Doctoral Award Holder
Email alexandra.eveleigh.09@ucl.ac.uk

or
West Yorkshire Archive Service
PO Box 5
Nepshaw Lane South
Morley
Leeds
LS27 0QP
Email archives@wyjs.org.uk

