

# Deleting to Preserve: appraisal in the digital age

*Simon Wilson, Senior Archivist*

Why do we appraise?

## Why do we appraise?

I asked a few colleagues;

- to remove duplicates and to weed out the rubbish
- make the items we retain easier for users to discover
- save space and costs of storing material un-necessarily
- because we can't keep everything

The principles of appraisal are applied in different ways and for different reasons

## Why do we appraise?

Evidential, informational or historical values as criteria in making the decision whether to retain an item

Process not undertaken lightly because items not retained will be destroyed and the content in all likelihood will be lost

Trained to work with material in a range of formats  
- we do adopt different approaches to storing & managing different non-digital formats eg maps, photographs etc

Appraisal decision based on content and context not format

# What does appraisal look like?



# How we appraise

Series of questions:

1. What is it?
2. Can I (or a colleague) read it?
3. Do I want to keep it?
4. Then consider description and storage/preservation issues
5. Next document...



Appraisal in the digital age?

# Appraisal in the digital age

Why should it be any different with born-digital archives?

Apply the same evidential, informational or historical values as criteria in making the decision whether to retain an item

Trained to work with material in a range of formats

But there are significant differences – not least the issues surrounding media, file formats and scale



# What does appraisal look like in the digital age?

2010-15 DirPrnInfo and catalogue notes.xls [Compatibility Mode] - Microsoft Excel

Filename	Date created	Date last accessed	File extension	File Size	Checksum	Filepath
Bookplate.pub	18/02/2009 20:27	08/01/2009 17:55	pub	26,112	B6B81A1D799335CDAA4989C9F4CB78FE	D:\born digital\Stephen Gallagher\Bookplate.pub
CompSlip.pub	24/06/2008 23:42	08/01/2009 17:56	pub	20,992	868A497B18FAA988597B48300240D48	D:\born digital\Stephen Gallagher\CompSlip.pub
Letterhead template.doc	24/06/2008 23:42	06/09/2007 19:14	doc	21,504	40265233F0064BF21F0386D99A4774BF	D:\born digital\Stephen Gallagher\Letterhead template.doc
A Year in Oktober.doc	24/06/2008 23:42	16/10/1998 17:32	doc	55,296	AD0E4497263F85CE3C15694C35C61FF0	D:\born digital\Stephen Gallagher\Articles\A Year in Oktober.doc
Abner Stein Books CV April 2006.doc	24/06/2008 23:42	17/04/2006 15:14	doc	32,768	B95AC91D8B84EB21145E9C53BA6032AA	D:\born digital\Stephen Gallagher\Articles\Abner Stein Books CV April 2006.doc
Agency CV November 2005.doc	24/06/2008 23:42	22/11/2005 01:28	doc	40,960	CB63602C9473330A4C39B8585411015A	D:\born digital\Stephen Gallagher\Articles\Agency CV November 2005.doc
American notes 2007.doc	24/06/2008 23:42	13/07/2007 12:15	doc	72,704	F5C662A91F11EA9FC65DF8AA31575EEC	D:\born digital\Stephen Gallagher\Articles\American notes 2007.doc
ARCHIVET.doc	24/06/2008 23:42	16/09/2001 21:17	doc	34,304	R49R4R4R6FF4756C65E9F232B123DC3F	D:\born digital\Stephen Gallagher\Articles\ARCHIVET.doc

	A	B	C	D	E
1	Folder Name	Filepath	Number of Sub Folders	Number of Files	Folder Size (kb)
11	Runaways Project	D:\born digital\Stephen Gallagher\JBTV DEVELOPMENT 2009\Runaways Project\	0	18	5,093,923
12	Train Wreck Jane	D:\born digital\Stephen Gallagher\JBTV DEVELOPMENT 2009\Train Wreck Jane\	0	1	32,768
13	My Radio	D:\born digital\Stephen Gallagher\My Radio\	10	0	0
14	Gallagher - A Resistance To Pressure 256 kbps	D:\born digital\Stephen Gallagher\My Radio\Gallagher - A Resistance To Pressure 256 kbps\	0	1	165,969,792
15	Gallagher - An Alternative To Suicide 96 kbps	D:\born digital\Stephen Gallagher\My Radio\Gallagher - An Alternative To Suicide 96 kbps\	0	1	61,613,349

# How we appraise (in the digital age)

1. What is it?
2. Can I (or a colleague) read the media?  
Can we read the contents?
3. Do I want to keep it?
4. Then consider description and storage/preservation issues
5. Look at the next file...



Its complicated...

other factors/approaches to  
consider

## The scale of things

If we have the hardware, software to process the material...

The biggest challenge is the sheer quantity of material

- collections measured in terms of x thousand files or x GB
- 1 PDF file is 500MB, 7249 pages (= 1.5 linear metres)

Recent hybrid collection – 3 boxes of paper & 1 hard drive (contained 18,453 files – 78GB) with more digital to come!

Look to be consistent in our approach – but can we treat the paper and digital material in the same collection equally?

## Cost factors

We have treated different collections differently based on content and expected usage – usually archival description

Cost equations:

**Description**     Archivist time vs anticipated use

**Appraisal**     Time to undertake appraisal vs space saving

**Storage**     Cost of storing rubbish vs time to appraise

**Access**     Cost to provide access (but nobody uses it)

# What do our users want? (in the digital age)

Keep everything

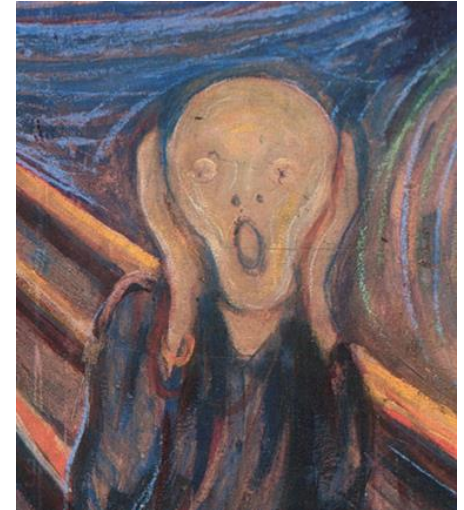
For “everything” to be accessible

A service that is as easy to use as Google

Online access to digital material rather than a physical visit

To copy material and re-use it for their own purposes

Access to the original files to conduct analysis



# Depositors understanding about digital archives?

Depositors aren't as familiar with digital archives as they are for paper archives

- new conversation with depositors we've worked with
- questions about their digital life – might seem intrusive
- need to explain why we need to act quickly to preserve media and their contents

“Wikileaks affect” - genuine concern that digital archives will be ‘exposed’- we need to demonstrate our custodial care and principles still apply in this new environment

# Don't appraise anything?

- 1) Storage is cheap
- 2) Our users don't want us to throw anything away
- 3) We don't have the time to go through each and every file





## Impact of non-appraisal?

Material of no research/historical value that needs to be:

- stored, described, migrated, transformed, managed etc

**Forever** (and that's a mighty long time)

or until you re-evaluate it, but as we couldn't be bothered first time - will we ever go back?

# Automating the appraisal process?

Sounds like a good use of technology...

How could it work?

- delete solely based on file format, file type, by filename?
- none of these approaches would be accepted for paper files

How can we define the parameters when they change (or be different) for each archival collection?

# Managed risk approach

Likelihood and impact of sensitive material being accessed

Sensitive information

- not just health, identity, credit card details
- might be a telephone number, an email address...

Worst case scenarios:

- breach of data protection legislation, trust, institutional reputation, copyright...



# Collections assessment approach

Personal experience of working on collections containing digital material – is that (as with traditional collections) some collections benefit from appraisal and others don't

Develop a strategy to identify issues surrounding sensitive content, level of research interest and use

Conduct an assessment for each collection

# New ways of working

## Closer working with **depositors**

- identify sensitive material and possible format issues
- possible role in pre-processing / sorting
- risk of “good stuff” not being deposited – retained by depositor or not considered to be of interest

Use visualisation **tools** offering researcher (and archivist) new perspectives on the collections/information in our care

Use checksum tools to identify duplicate content

## Conclusion

We have always adopted different approaches to how we process collections in our care, digital material is no different

Appraisal is just one element of the challenge of working with born-digital archives.

We need to employ new skills, work more closely with depositors and use tools to solve the problems that the sheer quantity of digital material brings and develop our services to ensure they are still relevant and appropriate to our users

# Contact

Simon Wilson

Senior Archivist

Hull History Centre

Tel 01482 317506

Email [s.wilson@hull.ac.uk](mailto:s.wilson@hull.ac.uk)

[www.hullhistorycentre.org.uk](http://www.hullhistorycentre.org.uk)

