Born-digital archives @ Hull: early steps and lessons learnt (so far)

Simon Wilson, Digital Archivist (AIMS Project)
What is the AIMS Project?

Steps taken @ Hull

Steps still to take

Questions
AIMS Project

An inter-Institutional Model for Stewardship

Each partner employing a Digital Archivist

Two year project (ends Sept 2011)

To process born-digital collections

To use Hydra, a Fedora repository-based solution
AIMS Approach

Starting point - traditional archival theory/principles

AIMS Model – identify commonality across the 4 partners - not looking to create a single solution

To disseminate the methods & lessons learnt

Virtual collaboration – weekly Skype calls and GoogleDocs - multiple editors work on same text but also face to face mtgs

Toad in the Hull, design for a toad that was never built  (Larkin with Toads archive)
look in your store

Survey your holdings for existing (digitised or) born-digital material - search your catalogue for CD, disc etc

Identify distinct media and content issues, possible level of cataloguing etc

Knowledge of quantity, media & range of media formats is a step forward
Relationship more **critical** than ever before, new **questions** we need to ask – hardware, software, passwords etc

Stephen Gallagher:
2005 - 42 boxes paper archives inc some of his TV work on video/DVD

2010 - born-digital material – 14,320 digital files (13.6GB) via an external hard drive
quarantine space

Setting-up a quarantine PC - test files for viruses

Purchase new PC with lots of expansion space, look to create some backward compatibility eg USB floppy / zip drive

Establish policies and procedures for receiving and managing born-digital archives
Karen’s Dir Printer

Useful **free** software tool allows you to retrieve key information about each file/folder including date created, checksum etc.

Archives and/or depositor – create a file **manifest** at the time of transfer.

Also use **retrospectively** for material already deposited.

Rubymatica

Tool (developed by AIMS project programmer and currently in beta testing)

Takes the files that are to be ingested into Fedora and creates a metadata wrapper

Use DROID, FITS and other established tools
steps still to take

Getting the material into Fedora will be a huge achievement....

... but we still have the following on our to do list:

1. Email (watching brief)
2. Preservation actions
3. Arrangement & Description
4. Discovery & Access

Staromestska Metro Station, Prague [http://www.flickr.com/photos/dcdead/4678139577/]
We will be looking to start using the Plato tool to test migration from one file format to another.

Experiments are run on test corpus of files.

Establish policies and procedures for receiving and managing born-digital archives.

Plato — see [http://www.ifs.tuwien.ac.at/dp/plato/intro.html](http://www.ifs.tuwien.ac.at/dp/plato/intro.html)
Hypatia

How do we arrange and describe born-digital archives?

Ideally we would like to view the file & integrate it with finding aids etc

Hypatia - new tool defined by archivists for archivists

Proof of concept is currently being built by developers – showcase at SAA in Chicago (tbc)
Axiell

Working with colleagues at LSE, Parliamentary Archives, Wellcome Library etc

Looking at how CALM can exchange information / data (via API) with digital repository’s workflow.
user experience

Paper archives offer a sense of discovery

Notebooks - plot outlines, notes of meetings with producers, snippets of dialogue etc for different work all *intermingled*

Born-digital material the same information is *dispersed* between multiple digital files

Acc 2008 box 15, Chimera file 2
archives workflow...

Quarantine
Check for mould & infestation etc

Archives Work room
Initial review / appraisal

Donor survey
talk to depositor

Re-box the material into archive folders / boxes etc

Cataloguing Room
first serious look at material, appraisal, arrangement & description; identify any legal/access restrictions etc

REPOSITORY
---
Produce surrogate copies for user access (m/film)

original item produced in search room

Catalogue

WEB PLATFORM – catalogue available online
AIMS, CALM & Fedora workflow...

**Digital Objects**

- Quarantine
  - Virus check, create basic manifest etc
- Network storage
  - Initial review/appraisal
- Donor survey
  - Secure web-form on a SQL Lite database

**Rubymatica**
- Virus check, checksum, create manifest (with copy of original manifest and Donor Survey for ref) creates SIP

**FEDORA**
- (Technical metadata)
- (Access & Security)
- Ingest (BAG-IT)
- Transformation layer – migrate file on to access format etc

**Hypatia (Arrangement & Description Tool)**
- Inc browse/view/delete files; set intellectual arrangement and align digital objects with this; set user permissions; create descriptive record etc

**CALM**
- (Descriptive metadata)
- Accession record

**WEB PLATFORM**
- (merged Solr indexes & Blacklight)
- EAD Finding aid inc links to digital objects
- Solr index

**KEY**
- Digital Asset
- Metadata
scale of the task

We don’t have a complete solution in-place yet....

... but we already have 40,000 born-digital files from just 4 collections

Expect to have 1m+ born-digital files in 5 years and that backlog will be measured in TB
conclusion

The nature (and format) of archives has changed dramatically in recent years.

It is not for archivists to try and predict the future.

We can’t “keep everything” and hope Google create an algorithm to enable access.

Archivists: already have the skills – just need the confidence to appreciate they are applicable.
on monday

There are some things you can do on Monday ...

1. Survey your collections

2. Install Karen’s Directory Printer or DROID etc

3. Have a play – hands-on learning does count

You already have the skills – apply to this new domain

any questions?

Contact details

Simon Wilson
Digital Archivist
Hull History Centre

Tel 01482 317506
s.wilson@hull.ac.uk

AIMS Project blog –
http://born-digital-archives.blogspot.com

Portrait of Claude-Henri Watelet blogging, after Jean-Baptiste Greuze
http://www.flickr.com/photos/notionscapital/2497196140/in/photostream/