Personal Digital Archiving

Jeff Ubois

http://personalarchiving.com
Agenda

~ Intro

~ PDA 2011, and some examples

~ Economics, and the boundary between commercial/non-commercial sectors

~ What’s next
personal archives

personal archives and a new archival calling
personal archives exchange 2010
marilyn monroe the personal archives
richard cox personal archives

About 158,000,000 results (0.19 seconds)

THE PERSONAL ARCHIVE: ON RETRIEVING VALUABLE CULTURAL RESOURCES
visits - 12/16/10
The devastating destruction of personal archives in America and globally, costing billions in cultural and scientific losses. Plans for financing and ... www.grazian-archive.com/projects/archvpt.html - Cached - Similar

Understanding Personal Archives: Exchange 2010 SP1 Help
2 Dec 2010 ... Personal archives help you regain control of your organization's messaging data by eliminating the need for personal store (.pst) files and ... technet.microsoft.com/en-us/library/dd979795.aspx - Cached - Similar

Email Archiving and Retention - Microsoft Exchange 2010
Personal Archive is a specialized mailbox associated with a user's primary ... www.microsoft.com/exchange/.../email-archiving-and-retention.aspx - Cached

Reading Archives: Personal Archives
21 Jan 2009 ... Here is information from my publisher of a new book about personal archives. I thought since I spend so much commenting on other books, ... readingarchives.blogspot.com/2009/...personal-archives.html - Cached - Similar

Personal Archives Accessible in Digital Media (Paradigm)
1 Oct 2008 ... A JISC-funded project of the Universities of Oxford and Manchester to produce an online workbook on Digital Private Papers of use to ... www.paradigm.ac.uk/ - Cached - Similar

Personal Archives
Papers and other archival material submitted by one individual or organization. The list below lists a selection of personal archives at the LA&M. ...
Grazian Archive

Assuming that creative work has an estimable value -- but aren't we paying tens of thousands of professors and as many artists and leaders according to this assumption? -- the sum of all such creative values lost, not saved, in a decade must amount to billions of dollars.  - Alfred de Grazia
The materials are uncensored and complete, even where they are not considered by the author or his readers to be praiseworthy; the ideal archive should reveal its subject and author to be true to himself and others, he must be fully visible and recorded authentically and critically.
Why it matters

~ Today’s Shakespeares, Einsteins, Rembrandts are producing their works digitally.

~ Eyewitness news

~ What would you pay for a video of the last 7 generations of your family?
Why now?

- Surplus of fakery creates demand for authenticity
- Privilege of preservation being democratized by new technology
- Cheap capture (baby pix)
- The “Archival Instinct”
But ...

〜 it’s still very early days
Web Coverage

- Personal Archiving http://personalarchiving.com
- Rudy Rucker http://www.rudyrucker.com/blog/2011/02/24/speakage/
- Litbrarian http://litbrarian.wordpress.com/2011/02/
- Waki Librarian http://thewakilibrarian.wordpress.com/
- Birkin James Diana http://thewakilibrarian.wordpress.com/
Problems

- What to save
  - publications, private data, 3rd party...
  - email, photos, home movies, SN data, health records ... any media type
- How to access it later
- How to pay for it (institutional design)
Highlights

- Cameras: how they skew what we see
- Amateurs: what you can do yourself
- User practices: what people do
- Interface designs & preservation systems
- Naivety: unexamined dependencies (stone CDs, startups as custodians)
Economics

~ Predictability: What do expenses & income look like, long term?

~ Boundaries: what is the relation between the commercial & non-commercial sectors?

~ Institutions / Individuals: top down vs bottom up approaches to preservation
Figure 5: The Emerging Gap
Information Creation > Storage Available

Petabytes

- Digital Information Created
- Digital Storage Available

The Digital Universe Gap

Source: IDC Digital Universe Study, sponsored by EMC, May 2010
Costs

~ Predictable costs essential for organizations that promise long term access

~ Ingest, storage, other costs vary

~ Complex cost models illustrate more uncertainty (LIFE, CET, KRDS, DANS)
Reported Scanning

- Text pages: €0.1 - €0.8 / page
- Images: €0.72 - 6 / page
- Audio: €6.42 - €78.84 / hour
- Videotape: €5.46 - €120 / hour
- Film: €8 - €1040 / hour

Source: Numeric
Price Per Box

Generally we find 2000 – 3000 pages in a standard box. A ball park for your project would be somewhere from $.10 per page to $.23 per page “All In” depending on what services we do for you and if we scan color pages in color. Your “All In” total would be from $200.00 to $690.00.
Storage costs

COST OF A PETABYTE

- RAW DRIVES: $81,000
- BACKBLAZE: $117,000
- DELL MD1000: $826,000
- SUN X4550: $1,000,000
- NETAPP FAS-6000: $1,714,000
- AMAZON S3: $2,806,000
- EMC NS-960: $2,860,000

* Amazon S3 Storage over three years (minus electricity, co-location and administration).
Sequence Read Archive (SRA) and Trace Archive repositories have been discontinued

Due to budget constraints, NCBI will be discontinuing its Sequence Read Archive (SRA) and Trace Archive repositories for high-throughput sequence data. Closure of the databases will occur in phases. SRA and Trace will stop accepting some types of submissions in the coming weeks, and all submissions within the next 12 months. Over the next several months, NCBI will be working with staff from NIH Institutes that fund large-scale sequencing efforts to develop an approach for future access to and storage of the existing data. NCBI will continue to support and develop information resources for biological data derived from next-generation sequencing such as genotypes, common variations, rare variations, sequence assemblies and gene expression data. We therefore encourage the research community to continue submissions of these data to the applicable databases, including:

1. RNA-Seq and epigenomic data to GEO
2. Variants, genotypes, phased haplotypes, and polymorphisms to dbVar, dbGaP and dbSNP
3. Genomic assemblies to GenBank/WGS
4. Transcript assemblies to GenBank/TSA
5. 16S ribosomal RNA and other targeted locus survey assemblies to GenBank

NCBI expects new applications will continue to emerge for next generation technology. We are excited to work with the community to develop strategies for archiving other summary experimental measures that are informative, efficient, and valuable to the biomedical research community.

For further information about submissions, contact NCBI's Help Desk.

To get started with SRA, enter one or more search terms.

The Sequence Read Archive (SRA) stores sequencing data from the next generation of sequencing platforms including Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® System, Helicos Heliscope®, and others.

SRA is a service of the U.S. National Library of Medicine
Complex Cost

- Life Cycle Information for E-Literature
- CET (NASA)
- Keeping Research Data Safe
- CMDP (Danish Library, National Archive)
- Curtis & Cartwright
Pay Once Models

~ Gap between project based funding, need for perpetuity (eg NSF, NIH, you)
~ Real estimates:
  ~ Princeton: $5K/TB (~100x media cost)
  ~ PrestoPRIME: 40x raw media cost
~ Basis for buy a brick: endow a TB
Roles for Commerce

- Cataloging & indexing
- Clearance
- Long term promises...?
Memnon offers archives a range of services to digitise, restore, preserve and provide access to classic and current sound and video recordings in any format. Our service is suitable for a wide range of archive owners - from cultural institutions, libraries and universities to record labels, archive websites and private collectors. We provide a professional, high-quality and cost-effective service, that combines innovative techniques, in-depth experience of audiovisual processing and databases, professional research in archiving standards, and optimised production process.

Memnon is currently involved in the digitisation of over 500,000 hours of audiovisual archives (350,000 hours of sound and 150,000 hours of video) for prestigious cultural institutions and audiovisual organisations throughout Europe.

Memnon digitises all analogue formats (sound: magnetic tapes, audio cassettes, vinyl records) (video: Betacam SP, U-Matic, 1 inch, 2 inch, VHS, D3, D9, etc.) and digital supports (sound supports: DAT, CD) (video: DVD, Digital Beta, DVC Pro, DV Cam, P2, etc.).

Digitization of other archive media (papers, pictures ...) is arranged through specialist partners.

Memnon has developed a series of software tools and solutions to assist archives in achieving significant savings in the time required for indexation and documentation and to facilitate its post-digitisation operation.

We offer secure long-term storage and indexing of audiovisual archives, making them available worldwide.
Welcome to mylastemail.com where we offer you special Online Memorial and Online Obituary services.

The two longest words in the English language are “If only.” We may regret not telling someone how much we love them, or regret the harsh words we once had with a friend. We regret not letting friends and family into our lives a little more and sharing our thoughts and feelings with them. “If only I had told them”, “If only they could have known.”

And once we die that opportunity is lost – but it need not be.

Here at mylastemail.com we offer you the opportunity to make that “If only...” into a reality.
NEW! A beta version of an updated Wayback Machine interface and index (with more recent results) is now in testing! You can try <http://mylastemail.com/> at the test site for improved results. Note that availability and functionality of the beta will fluctuate during the test period.

Note some duplicates are not shown. See all.
* denotes when site was updated.
Material typically becomes available here 6 months or more after collection, with some exceptions See FAQ.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0 pages</td>
<td>0 pages</td>
<td>0 pages</td>
<td>0 pages</td>
<td>0 pages</td>
<td>0 pages</td>
<td>0 pages</td>
<td>2 pages</td>
<td>6 pages</td>
<td>15 pages</td>
<td>30 pages</td>
<td>21 pages</td>
<td>1 pages</td>
<td>0 pages</td>
</tr>
</tbody>
</table>

- Aug 05, 2003
- Nov 21, 2003
- Mar 23, 2004
- May 20, 2004
- Jun 10, 2004
- Jun 12, 2004
- Jun 18, 2004
- Jun 25, 2004
- Mar 09, 2005
- Apr 04, 2005
- Jun 26, 2005
- Oct 05, 2005
- Nov 04, 2005
- Nov 26, 2005
- Dec 01, 2005
- Dec 02, 2005
- Dec 20, 2005
- Dec 21, 2005
- Dec 21, 2005
- Dec 23, 2005
- Dec 23, 2005
- Dec 24, 2005
- Dec 25, 2005
- Jan 04, 2006
- Jan 12, 2006
- Jan 18, 2006
- Jan 17, 2007
- Feb 08, 2006
- Feb 09, 2006
- Mar 07, 2006
- Mar 07, 2006
- Mar 24, 2006
- Apr 02, 2006
- Apr 02, 2007
- Mar 04, 2007
- Apr 09, 2006
- Apr 19, 2006
- May 23, 2006
- Jun 11, 2006
- Jun 13, 2006
- Jun 16, 2006
- Jul 03, 2006
- Jul 11, 2006
- Aug 13, 2006
- Aug 14, 2006
- Aug 20, 2006
- Aug 21, 2006
- Oct 14, 2006
- Oct 27, 2006
- Nov 07, 2006
- Nov 14, 2006
- Nov 15, 2006
- Jan 02, 2007
- Jan 07, 2007
- Jan 12, 2007
- Jan 12, 2007
- Jan 22, 2007
- Jan 22, 2007
- Jan 27, 2007
- Feb 04, 2007
- Feb 06, 2007
- Feb 12, 2007
- Mar 04, 2007
- Mar 22, 2007
- May 06, 2007
- Jun 04, 2007
- Jun 08, 2007
- Jul 02, 2007
- Jul 03, 2007
- Jul 03, 2007
- Aug 11, 2007
- Sep 11, 2007
- Oct 13, 2007
- Feb 21, 2008
The idea has once again arisen to sell University of Iowa’s Jackson Pollock masterpiece “Mural” (1943), but this time it’s not to pay for flood damage to the University’s museum. Representative Scott Raecker (R-Urbandale) of the Iowa House of Representatives has introduced a bill to force the sale of the artwork to pay for student scholarships.

In a conflict that pits the abstract value of art against the literal value of money, the Pollock is now a pawn in what has been called an “attack on the University of Iowa” by Democratic Iowa Senator Robert Dvorsky. Raecker argues that the sale of the painting, valued at $140 million, would finance a “$5 million a year endowment,” and “could let 750 to 1,000 students get a full-ride scholarship every year.” But the museum argues that the Pollock provides Iowa with its most significant piece of artwork, an invaluable cultural resource and educational tool.
Bridging Individuals

- Opportunity for LAMs to
  - support self archiving, harness non-market motivations
  - collaborate to develop software
  - collect small endowments (crowdsourced funding)
  - reach new constituencies

Monday, May 9, 2011
Questions & Requests

~ Feedback re: design competition
~ Cases
~ Luxembourg/Future of the Past
~ PDA 2012
Thank you

jeff@ubois.com
www.personalarchiving.com
Appendix

...
BRTF’s Funding

- Internal Budgeting. Examples: movie studios
- Public funding. Examples: national archive
- Fees for Preservation. Examples: Portico, Chronopolis
- Fees for Access. Examples: ICPSR, LexisNexis

Source: BRTF / Sustainable Economics for a Digital Planet
PrestoCentre

- Aggregating reported data
- Collecting tenders & responses
- Partnering with others collecting similar information (FOCAL, others)
- Moving towards fixed units, fixed costs...don’t know quantities!
Robots.txt Query Exclusion.

We're sorry, access to http://www.finalthoughts.com has been blocked by the site owner via robots.txt.

You may want to:

- Read more about robots.txt
- See the site's robots.txt file
- Try the page on the live web: http://www.finalthoughts.com
- Search for all pages on the site finalthoughts.com/
- Try a different page address, at top

See the FAQs for more info and help, or contact us.
Society’s ever-increasing demands for vast amounts of data to be kept for the future are not matched by suitably lavish funds. Thus, absent a technological miracle, bit preservation is a problem with which we are doomed to struggle indefinitely. - David S.H. Rosenthal
PDA 2011

- Second year, ~140 people
- Diverse group from government, industry, academia, public
- Broad subject coverage from library management to imaging to healthcare to religion to privacy to sci-fi to SNs...
- Coverage via #PDA2011
Partnerships

- Partnerships aren’t free: huge commercial uncertainties can mean harsh terms
- Mixed record
- Serious downside risks
THE NEW RENAISSANCE

REPORT OF THE ‘COMITÉ DES SAGES’

REFLECTION GROUP ON BRINGING EUROPE’S CULTURAL HERITAGE ONLINE
Report warns of 'digital Dark Age' if digitisation is left to private sector

Research commissioned by EU urges member states to take greater responsibility

Benedicte Page

guardian.co.uk, Wednesday 12 January 2011 16.25 GMT

Article history
Comité des Sages:

“With no exaggeration, we can state that what is at stake is a common good of humanity and not just of Europe.”

“The public sector has the primary responsibility to fund digitisation, and Member States will need to considerably step up their investments in digitisation. The current financial crisis cannot be ignored, but equally cannot be a reason for not acting.”
On Privatization

“The power to remember, as well as the power to forget, are thus gradually being concentrated in clusters of commercial enterprises with very particular interests, beliefs, ideologies, and preferences.” - Guy Pessach