Getting to the Bottom Line
20 Cost Questions for Digital Preservation

Matt Schultz, MetaArchive Cooperative
Aaron Trehub, Auburn University, ADPNet

ASERL Webinar Series
Wednesday, March 11th, 2015
Background

- Dodging the Memory Hole Saving Digital News
- ASERL Webinar Series, November 17, 2014
  - Katherine Skinner (Educopia Institute)
  - Matt Schultz (MetaArchive Cooperative)
- Comparative Analysis of Distributed Digital Preservation Systems

https://vimeo.com/112488449
Cooperative Preservation

**Principles**
- Transparency
- Trust
- Accountability
- Collaboration
- Flexibility
- Shared Investment
- Shared Knowledge
- Shared Solutions

**Practices**
- Preserving our missions through mutual support and collaboration
- Solving common problems together and producing field-wide benefits
- Gaining back control of our digital futures through ownership, control, and freedom
MetaArchive Cooperative

MetaArchive is a true cooperative:

- All hardware and software assets are owned by members
- Membership fees and storage fees go to a central pool of support for members’ co-op activities
  - Members recently increased network storage and lowered their storage fees over 40%
- Common, neutral center (Educopia Institute)
- Distribution of work
- Community of engagement
- Building knowledge
- Accomplishing preservation
- Concentrated effort toward unified goal
Founded on the premise that memory organizations should maintain their historical role as cultural stewards
- Preservation of digital assets as corollary to preserving physical ones
- Libraries/archives need in-house expertise and knowledge
- Value of curators and librarians and archivists

Chose technical and organizational infrastructure that capitalizes on memory organization proven methodologies
- LOCKSS (Lots of Copies Keep Stuff Safe)
- Distributed digital preservation
- Embedded technologies
- Partnerships to keep costing affordable
Ownership vs. Outsourcing

- Libraries are challenged to retain ownership & control of their core assets and resources
  - Digitization units
  - Scholarly journals & databases
  - Monographs (ebooks)
  - Web services
  - IT infrastructure
  - Digital collections
The commercial cloud is revolutionizing IT
  ▪ Reducing hardware and staff costs
  ▪ Economies of scale (maybe)

The commercial cloud is raising challenging questions
  ▪ Data security
  ▪ Data privacy
  ▪ Data retrieval
Turning to the Cloud for Help?

- Records in the Cloud (http://www.recordsinthecloud.org/)
  - Building upon the work of InterPARES
  - Ensuring authenticity and reliability of electronic records
  - We may be paying a high price for loss of control
Turning to the Cloud for Help?

- David Rosenthal - DSHR's Blog (http://blog.dshr.org/)
  - Economic Model of Long-Term Storage
  - Cost of cloud storage will not magically decrease over time
    - If your collection grows more than 10% year storage cost as a proportion of the total budget will increase
    - Cloud providers are benefitting from economies of scale - not the consumer
  - Time-tested laws (Moore & Kryder) may be changing - Bezo's Law?
Library-Oriented Solutions
How Do I Make Sense of the Costs?

- **Economics of Digital Preservation**
  - Blue Ribbon Task Force on Sustainable Digital Preservation & Access
    - Ingest & Staffing are costly
    - Economies of Scale
    - Automation
    - Focus on Return on Investment (ROI)
  - 4C Project – Collaboration to Clarify the Costs of Curation
    - Towards a cost calculator
  - Total Cost of Preservation (TCP)
    - Pay as you go; or
    - Pay once, preserve forever
  - Economic Models for Long-Term Storage (Rosenthal)
    - Use of the Cloud & Long-Lived Media

- **Great resources** – all require deep reading and training
Organizations need a simple, ready-to-use, but comprehensive set of questions to assist with their data gathering and comparative analyses.

Features and functionality are important, but those are often the easy pieces of information to learn about.

Getting to the bottom line, gathering the short- and long-term costs, is often more difficult.
Getting to the Bottom Line:
20 Cost Questions for Digital Preservation

The following questions are being made freely available by the MetaArchive Cooperative’s Outreach Committee to assist comparisons with their competitive offerings or various digital preservation solutions. Ask these of any digital preservation solution provider. Features and functionality are important, but those are often the easy pieces of information to learn about. Getting to the bottom line, understanding the short- and long-term costs, is often more critical. Use of all or any of the questions is disclaimed.

1. What are the solution provider’s licensing, subscription or membership fees?
   ✓ Have these fees increased or decreased over the past three years, and why?
   ✓ How often is the fee structure reviewed? And how are they set?
   ✓ How are current/subscribers/members (or member) cost increased during any such reviews?

2. Are additional memberships required to participate in the solution?
   ✓ If so, what are the fees associated with those memberships?

3. What are the solution provider’s storage fees?
   ✓ How have the fees increased or decreased over the past three years, and why?
   ✓ How often is the fee structure reviewed? And how are they set?
   ✓ How are current/subscribers/members (or member) cost increased during any such reviews?

4. How does the solution provide for increases in storage capacity?
   ✓ How do these increases affect fees, if at all?

5. Do any fees paid (licensing/subscription/membership/subscription) include geographically distributed copies of our content?
   ✓ If not, what’s the additional cost for this option?
   ✓ If so, how many copies, and in what geographic locations?

6. Roughly, what percentage of fees paid (all included) are used to pay for the solution provider’s support/staff?

7. Does the solution provider charge a one-time setup, implementation, or initial ingest fee, license and beyond any ongoing fees?

8. Does the solution provider provide a one-time setup, implementation, or initial ingest fee, license and beyond any ongoing fees?
   ✓ If so, does the solution provider charge any more? Is there a reference?
   ✓ If not, do any fees being paid cover things like verifying their proprietary, ensuring their backup, and removing any fees that may get charged in the process?

9. Is there a charge for retrieving content from the solution?
   ✓ If yes, how much is it?
   ✓ How is the charge calculated and what does it cover?

http://metaarchive.org/cost-questions
Basic Costs

- What are the solution provider’s licensing, subscription or membership fees?
  - Have these fees increased or decreased over the past three years, and why?
  - How often is the fee structure reviewed? And how are fees set?
  - How are customers/subscribers/members consulted during any such reviews?
- Are additional memberships required to participate in the solution?
  - If so, what are the fees associated with those memberships?
- Is there a minimum licensing/subscription/membership term?
- On what schedule are customers/subscribers/members billed?
Storage Costs

- What are the solution provider’s storage fees?
  - Have these fees increased or decreased over the past three years, and why?
  - How often is the fee structure reviewed? And how are fees set?
  - How are customers/subscribers/members consulted during any such reviews?
- Are there limits on the number or size of collections that can be deposited?
  - Is there a maximum limit on the amount of storage that can be utilized?
- How does the solution provide for increases in storage capacity?
  - And how do these increases affect fees, if at all?
Ingest Costs

✓ Does the solution provider charge a one-time setup, implementation, or initial ingest fee, above and beyond any ongoing fees?
✓ On average, how long does it take to begin using the solution once a contract or service license agreement (SLA) has been signed
  ✓ What steps are involved?
✓ Does the solution provider prepare content for ingest?
  ✓ If so, do any fees being paid cover tasks like verifying inventories, performing fixity checks, and/or repairing any files that may get damaged in the process of sending content to the solution provider?
  ✓ If not, does the solution provider cover/defer any costs to train or outsource?
Is there a charge for retrieving content from the solution?
- If yes, how much is it?
- How is this charge calculated and what does it cover?
- Under what terms? Are there any restrictions or limitations?

Is there a charge for deleting content from the solution?
- If yes, how much is it?
- How is this charge calculated and what does it cover?
- Under what terms? Are there any restrictions or limitations?

SELF-QUESTION: Am I, as a customer/subscriber/member, also paying for local backups above and beyond the preservation service costs?
- Recognizing that both preservation and backup are important and have their own unique and important places in an overall strategy, how can I best balance both costs?
What are a few specific examples of tasks, services, or resources that my fees for this solution are supporting (e.g., staff salaries, infrastructure upgrades, research and development, etc.)?

If the solution is a non-profit (or run by one), are copies of the annual operating budget and financial statement shared with the members?

If not, are there plans to?

In terms of sustainability, does the solution provider have a strategic plan, succession plan, or disaster recovery plan?

If so, how up-to-date are such plans?

Has the solution provider engaged in any audits or risk assessments?

Are any of the plans or audit/assessment results publicly available?
Using the Questions

- Pretty handout or printer-friendly version
- Make use of all or as many of the questions as desired
- Logically arranged, but feel free to ask questions in whatever order seems appropriate
- Just suggestions, re-phrase as necessary
- Maybe you will think of some questions that are not listed...let us know
MetaArchive Outreach Committee

- OUTREACH-COMM-L@metaarchive.org
- MetaArchive Outreach Committee
  - Aaron Trehub, Co-Chair (Auburn University)
  - Cinda May, Co-Chair (Indiana State University)
  - Holly Mercer (University of Tennessee, Knoxville)
  - Sammie Morris (Purdue University)
  - David Ratledge (University of Tennessee, Knoxville)
  - Deanna Ulvestad (Greene County Public Library)
Getting to the Bottom Line

- MetaArchive has worked with inquiring libraries, archives, and other memory organizations, and its own members for over a decade to respond transparently to many of the questions they have about the cost of doing digital preservation.

- *Getting to the Bottom Line: 20 Cost Questions for Digital Preservation* is one step forward to encourage more transparency field-wide.
What Else Can We Do?

- Cannot ultimately demand transparency or enforce it
- How can we create the conditions for that to happen?
- How can we produce and host community-based resources to sustain it?
In our experience (North America), cost is one of the two most important constraints on instituting a DP/DC program (the other is staff expertise, or the lack of it).

Cost is also one of the most difficult elements to analyze systematically.

Cost bears directly on long-term sustainability, arguably the most important DP/DC attribute.

Then there’s this...
“C’mon guys...
You expect me to work through a bunch of equations?! Don’t give me theories, give me facts. Just tell me what it’s gonna @#$%! cost.”

(With thanks to Neil Grindley of JISC)
Let’s Get Started

- Build a list
  - Library-Oriented Solutions
  - Commercial Solutions
- Hit the Websites
  - Features
  - Costs/Pricing
- Request Quotes/Price Sheets
  - CAUTION: Often custom, non-standard
  - Least transparent
- Annual membership fee(s)
- Annual storage fee(s)
- Upload and retrieval request fee
- Data transfer IN fee
- Data transfer OUT fee
- Staff time/sweat equity
- Distributed network?
**ADPNet**
(http://adpn.org/docs/pdf/ADPNet_Governance_Policy.pdf)

- Annual membership fee(s): $300, $800, $2,500, $5,000
- Annual storage fee(s): $50/100MB, $25/500MB, $250/100GB, $350/500GB above base allotments for the four membership levels
- Upload and retrieval request fee: $0.00
- Data transfer IN fee: $0.00
- Data transfer OUT fee: $0.00
- Cost of new equipment: $3,000-$5,000 every 4 years
- Staff time/sweat equity: ?
- Distributed network: Yes
Amazon Glacier
(http://aws.amazon.com/glacier/pricing/)

- Annual membership fee(s): $0.00
- Annual storage fee(s): $0.01/GB = $10/TB
- Upload and retrieval request fee: $0.05 per 1,000 requests
- Data transfer IN fee: $0.00
- Data transfer OUT fee: varies; $0.02/GB to another AWS region
- Cost of new equipment: n/a
- Staff time/sweat equity: ?
- Distributed network: Yes (?)
- Annual membership fee(s): $20,000
- Annual storage fee(s): 10TB allocation; members may purchase additional capacity in 5TB increments
- Upload and retrieval request fee: ?
- Data transfer IN fee: ?
- Data transfer OUT fee: ?
- Cost of new equipment: ?
- Staff time/sweat equity: ?
- Distributed network: ?
Annual membership fee(s): $4,500 (assessment only)-$11,900
Annual storage fee(s): 1TB allocated; additional storage at $1,000/TB/year
Upload and retrieval request fee: ?
Data transfer IN fee: ?
Data transfer OUT fee: ?
Cost of new equipment: ?
Staff time/sweat equity: ?
Distributed network: Yes
CDL Merritt
(UC only: http://www.cdlib.org/services/uc3/merritt/)

- Annual membership fee(s): $0.00
- Annual storage fee(s): ?
- Upload and retrieval request fee: ?
- Data transfer IN fee: ?
- Data transfer OUT fee: ?
- Cost of new equipment: ?
- Staff time/sweat equity: ?
- Distributed network? Yes
Chronopolis
(http://chronopolis.sdsc.edu/pricing/index.html)

- Annual membership fee(s): ?
- Annual storage fee(s): ?
- Upload and retrieval request fee: ?
- Data transfer IN fee: ?
- Data transfer OUT fee: ?
- Cost of new equipment: ?
- Staff time/sweat equity: ?
- Distributed network: Yes
Annual membership fee(s): [$20,000?]
Annual storage fee(s): ?
Upload and retrieval request fee: ?
Data transfer IN fee: ?
Data transfer OUT fee: ?
Cost of new equipment: ?
Staff time/sweat equity: ?
Distributed network: Yes
Annual membership fee(s): ?
Annual storage fee(s): $1,875-$6,750 for 1st TB, $700-$1,400 for each additional TB
Upload and retrieval request fee: ?
Data transfer IN fee: ?
Data transfer OUT fee: ?
Cost of new equipment: ?
Staff time/sweat equity: ?
Distributed network: Yes
Ex Libris Rosetta
(http://www.exlibrisgroup.com/category/RosettaOverview)

- Annual membership fee(s): ?
- Annual storage fee(s): ?
- Upload and retrieval request fee: ?
- Data transfer IN fee: ?
- Data transfer OUT fee: ?
- Cost of new equipment: ?
- Staff time/sweat equity: ?
- Distributed network: Yes
HathiTrust
(http://www.hathitrust.org/help_new_cost_model)

- Annual membership/storage fee(s): different formulas for public-domain (PD) and in-copyright (IC) volumes.
  - Public domain = \((PD \times X \times C) / N\), where \(PD\) = number of public domain volumes in HT, \(X\) = flexible multiplier (currently 1.5), \(C\) = basic infrastructure cost in HT (currently $.19 per volume), and \(N\) = total number of partners in HT
  - In copyright = \((IC \times X \times C) / H\), where \(IC\) = number of in-copyright volumes that overlap with volumes in HT, \(X\) = flexible multiplier (currently 1.5), \(C\) = basic infrastructure cost in HT (currently $.19 per volume), and \(H\) = number of partners that hold a particular in-copyright volume
- Example: Partner with 2 million PD and 2 million IC volumes would pay \(\$56,693.55\) per year (2012-2013)
HathiTrust (cont’d)

- Upload and retrieval request fee: ?
- Data transfer IN fee: ?
- Data transfer OUT fee: ?
- Cost of new equipment: ?
- Staff time/sweat equity: ?
- Distributed network: Yes
MetaArchive
(http://www.metaarchive.org/costs)

- Annual membership fee(s): $3,000, $5,500, and “collaborative member” rate (by negotiation)
- Annual storage fee(s): $.59/GB/Year = $585/TB/Year
- Upload and retrieval request fee: $0.00
- Data transfer IN fee: $0.00
- Data transfer OUT fee: $0.00
- Cost of new equipment: ~ $5,500 in first year and every three years thereafter
- Staff time/sweat equity: ~.02 FTE
- Distributed network: Yes
OCLC Digital Archive


- Annual membership fee(s): ?
- Annual storage fee(s): ?
- Upload and retrieval request fee: ?
- Data transfer IN fee: ?
- Data transfer OUT fee: ?
- Cost of new equipment: ?
- Staff time/sweat equity: ?
- Distributed network: ?
### Preservica


- **Annual membership fee(s):** ?
- **Annual storage fee(s):** $3,950 (up to 100GB)-$11,950 (1TB-10TB), with discounts available for >10TB
- **Upload and retrieval request fee:** ?
- **Data transfer IN fee:** ?
- **Data transfer OUT fee:** ?
- **Cost of new equipment:** ?
- **Staff time/sweat equity:** ?
- **Distributed network:** Yes
2. Web-based DP/DC Cost Calculator. If TIAA-CREF can do a Web-based retirement calculator, we can do this.
3. A place to put them.
1. DP/DC Cost-Price Spectrum

From Amazon Glacier to...what, exactly? (e.g. on a $/€ per TB basis?)
2. DP/DC Cost Calculator

- **CDL-TCP Cost and Price Model 2.2.1** (2014: USA)
  - [https://wiki.ucop.edu/display/Curation/Cost+Modelling](https://wiki.ucop.edu/display/Curation/Cost+Modelling)
  - Format: Excel spreadsheet
  - Status: available; in development

- **4C Curation Costs Exchange** (CCEx) (2014: EU)
  - [http://www.curationexchange.org/](http://www.curationexchange.org/)
  - Format: Web-based
  - Status: available; in development
Suggestions

- Decide on single Web-based DP/DC cost calculator; put resources into it (avoid duplication).
- Agree on basic set of cost variables (no more than 20?), including external costs (e.g. membership in DP/DC networks and/or commercial DP/DC products).
- Require transparency about costs and prices from all players.
- Enlist development (programming) support from the DP/DC community.
- Reduce dependence on soft money. Do it ourselves.
- Make the code publicly available (e.g. on GitHub)
3. The Place

“Establish a digital preservation resource centre (DPRC) – Provide decision-makers at LAMs with a single place for current information on various digital preservation solutions to enhance uptake and to foster a broader understanding of options.”

Options

- **COPTR**?
  - [http://coptr.digipres.org/Main_Page](http://coptr.digipres.org/Main_Page)
- **Digital Preservation Coalition**?
- **4C Project**?
  - [http://4cproject.eu/](http://4cproject.eu/)
- **Open Preservation Foundation**?
- **POWRR**?
  - [http://digitalpowrr.niu.edu/](http://digitalpowrr.niu.edu/)
- **APA/APARSEN**?
  - [http://www.alliancepermanentaccess.org/](http://www.alliancepermanentaccess.org/)
- **ANADP Web site**?
  - [http://www.educopia.org/events/ANADP](http://www.educopia.org/events/ANADP)
- **Someplace else**?
“Even if we don’t have a perfect model, we need to move forward now with a ‘good enough’ model.”

Thank You

Matt Schultz
MetaArchive Cooperative
matt.schultz@metaarchive.org

Aaron Trehub
Auburn University, ADPNet
trehuaj@auburn.edu