

# Getting to the Bottom Line

## 20 Cost Questions for Digital Preservation

Katherine Skinner (Educopia Institute)

Matt Schultz (Grand Valley State University)

Aaron Trehub (Auburn University, ADPNet)

# Aligning National Approaches to Digital Preservation



National Library of Estonia  
May 2011

Consorti de Biblioteques  
Universitaries de  
Catalunya (CBUC) and the  
Biblioteca de Catalunya



November 2013

# Exploring Costs of DP

**4C**

● ● ● ●

Collaboration to Clarify  
the Costs of Curation

UC3 Cost modeling  
UC3 Curator Center

- Total cost of preservation
  - ▶ Nominal units of activity or capacity
  - ▶ Fixed vs. incremental costs

$$TCP = n \cdot O + m \cdot T + A + \ell \cdot R + k \cdot S + j \cdot C + G + i \cdot V + D + M$$

Interventions Administration Management  
Owners Streams System Servers Storage Consumers Analysis

- Pay-as-you-go pricing
  - ▶ Common good vs. direct costs

$$X = \frac{A + \ell \cdot R + j \cdot C + i \cdot V + G + D + M}{n} + O + m_o \cdot T + k_o \cdot S$$


# Under-Resourced Institutions



<http://digitalpowrr.niu.edu/>

The screenshot shows a software interface with a top navigation bar containing colored tabs: "Digital", "Preservation", "Access", "Migration", "Dissemination", and "Other". Below the navigation bar, there are several sections with various options and settings, including "Logical Objects", "Physical Objects", "CD/DVD Images", "Software Packages", "Web to File", "Other Storage", and "Other Storage".

DCC Lifecycle Stages\*

	Access Use and Reuse	Creation/ Review/ Update	Cost Lifecycle Transition	Disposal	Migration	Preservation Action	Preservation Planning	Score
Acidic	1	4	3		11	10		
Binary Data			4					
Container						5		
Database		1	3		3	13		
Disk Image		5	3		1	1		
Document	3	1	4		32	10		
EBook					5	2		
Email			4			4	1	
Geospatial						1		
Image	2		2		20	10		
Project Management Data	1					1		
Research Data	1	8	12		4		12	15



# Getting to the Bottom Line



MetaArchive Cooperative is a digital preservation network created and hosted by and for cultural memory organizations.

## Getting to the Bottom Line: 20 Cost Questions for Digital Preservation

The following questions are being made freely available by the MetaArchive Cooperative Outreach Committee to assist institutions with their comparative analyses of 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 difficult. Users of this resource are encouraged to make use of all or as many of the questions as desired.

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 fees set?
  - How are customers/subscribers/members consulted 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?
  - 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?
4. How does the solution provide for increases in storage capacity?
  - And how do these increases affect fees, if at all?
5. Do any fees paid (licensing/subscription/membership/storage) 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 service/support staff?
7. Does the solution provider charge a one-time setup, implementation, or initial ingest fee, above and beyond any ongoing fees?
8. Does the solution provider prepare my content for ingest or am I responsible?
  - If not, does the solution provider cover/defer any costs to train or outsource?
  - If so, do any fees being paid cover things like verifying inventories, performing fixity checks, and repairing any files that may get damaged in the process?
9. 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?

1 of 2



MetaArchive Cooperative is a digital preservation network created and hosted by and for cultural memory organizations.

## Getting to the Bottom Line: 20 Cost Questions for Digital Preservation

10. Is there a charge for deleting content from the solution?
  - How is this accomplished?
11. Am I responsible for obtaining any additional hardware or software at my own expense in order to work with the solution?
  - If Yes, what are the average hardware/software costs?
  - Does the solution provider publish specifications or requirements?
  - What is the recommended replacement cycle?
  - Are there any additional costs for security (e.g., encrypted storage, etc.)?
12. Is there a minimum licensing/subscription/membership term?
13. 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?
14. To what degree are the steps related to ingest, description, preservation, etc. automated?
15. 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?
16. Does the solution provide basic documentation or instructions on getting started?
17. 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?
18. On what schedule are customers/subscribers/members billed?
19. 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?
20. SELF-QUESTION: Are you, 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 places in an overall strategy, how can you best balance both costs?

2 of 2

<http://metaarchive.org/cost-questions>

# Getting to the Bottom Line



MetaArchive Cooperative is a digital preservation network created and hosted by and for cultural memory organizations.

## Getting to the Bottom Line: 20 Cost Questions for Digital Preservation

The following questions are being made freely available by the MetaArchive Cooperative institutions with their comparative analyses of various digital preservation providers. Features and functionality are important, but those are often the bottom line, understanding the short- and long-term costs, is often more important. Use all or as many of the questions as desired.

1. What are the solution provider's licensing, subscription or membership fees?
  - > Have these fees increased or decreased over the past year?
  - > How often is the fee structure reviewed? And how are they reviewed?
  - > How are customers/subscribers/members consulted during the review process?
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 start-up costs?
  - > Have these fees increased or decreased over the past year?
  - > How often is the fee structure reviewed? And how are they reviewed?
  - > How are customers/subscribers/members consulted during the review process?
4. How does the solution provide for future growth?
  - > And how do these increase over time?
5. Do any fees paid (licensing/subscription) vary?
  - > If not, what's the additional cost for additional data?
  - > If so, how many copies, and how often?
6. Roughly, what percentage of fees are for hardware?
  - > How often is the hardware replaced?
7. Does the solution provider charge fees?
  - > If so, what are the fees associated with those memberships?
8. Does the solution provider prepare for disaster recovery?
  - > If not, does the solution provider have a disaster recovery plan?
  - > If so, do any fees being paid cover that may get damaged in the event of a disaster?
9. Is there a charge for retrieving content?
  - > If yes, how much is it?
  - > How is this charge calculated and what does it cover?
10. Is there a charge for archiving content?
  - > If yes, how much is it?
  - > How is this charge calculated and what does it cover?
11. Is there a charge for deleting content?
  - > If yes, how much is it?
  - > How is this charge calculated and what does it cover?
12. Is there a minimum licensing/subscription/membership term?
13. Are there limits on the number or size of collections that can be deposited?
  - > If so, what are the limits?
  - > Are there any exceptions to these limits?

1 of 2



## DIGITAL PRESERVATION NEWS

Lovely cost questions from the MetaArchive for thinking around service providers in DP:



**Neil Grindley**  
@neilgrindley

[Follow](#)

Here's something useful. [metaarchive.org/cost-questions](http://metaarchive.org/cost-questions)  
Pragmatic preservation advice from the folks @metaarchive  
Relevant for us @4c\_project

12:13 PM - 11 Mar 2015

2

2 of 2

<http://metaarchive.org/cost-questions>

# Why cost/pricing transparency matters

- 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

# An exercise

Applying some of the 20 questions to  
DP networks, products, and services\*

(\*Disclaimer: based on what I was able  
to find on the Web)



# ADPNet

- Annual membership/subscription 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
- Data in fee: \$0
- Data out fee: \$0
- Equipment cost: ? [varies; typically \$3,000-\$5,000 every 3-4 years]
- Staff time/sweat equity: ? [varies—e.g. with committee service]
- Multiple copies: Yes

# Amazon Glacier

- Annual membership/subscription fee(s): \$0
- Annual storage fee(s): \$0.007/GB/month = \$7/TB/month = \$84/TB/year
- Upload and retrieval request fee: \$0.05 per 1,000 requests
- Data in fee: \$0
- Data out fee: varies; starts at \$0.02/GB to another AWS region
- Equipment cost: n/a
- Staff time/sweat equity: ?
- Multiple copies: Yes

# APTrust

- Annual membership/subscription fee(s): \$20,000
- Annual storage fee(s): 10TB allocation; members may purchase additional capacity in 5TB increments at \$4,250/5TB/year (expected to drop to at least \$3,250/5TB/year “at scale”)
- Upload and retrieval request fee: ?
- Data in fee: ?
- Data out fee: ?
- Equipment cost: ?
- Staff time/sweat equity: ?
- Multiple copies: Yes

# Archivematica

- Annual membership/subscription fee(s): \$0
- Annual storage fee(s): see ArchivesDirect, DuraCloud, and Private LOCKSS Networks
- Upload and retrieval request fee: n/a
- Data in fee: n/a
- Data out fee: n/a
- Equipment cost: ?
- Staff time/sweat equity: ?
- Multiple copies: depends on DP/storage service

# ArchivesDirect

- Annual membership/subscription fee(s): \$4,500 (assessment only)-\$9,999
- Annual storage fee(s): 1TB allocated; additional storage at \$1,000/TB/year
- Upload and retrieval request fee: ?
- Data in fee: ?
- Data out fee: ?
- Equipment cost: n/a
- Staff time/sweat equity: ?
- Multiple copies: Yes



# Arkivum/100

- Annual membership/subscription fee(s): ?, but 1TB/year commitment required
- Annual storage fee(s): ?
- Upload and retrieval request fee: No
- Data in fee: No
- Data out fee: No
- Equipment cost: ?
- Staff time/sweat equity: ?
- Multiple copies: Yes

# Backblaze

- Annual membership/subscription fee(s): \$0
- Annual storage fee(s): \$0.005/GB/month = \$5/TB/month = \$60/TB/year
- Upload and retrieval request fee: \$0.004 per 10,000 downloads; \$0.004 per 1,000 other transactions
- Data in fee: \$0
- Data out fee: \$.05/GB
- Equipment cost: n/a
- Staff time/sweat equity: ?
- Multiple copies: Yes

# DPN

- Annual membership/subscription fee(s): Formerly \$20,000, currently forming membership committee to explore tiered pricing (FAQ document, June 2015)
- Annual storage fee(s): ?
- Upload and retrieval request fee: ?
- Data in fee: ?
- Data out fee: ?
- Equipment cost: ?
- Staff time/sweat equity: ?
- Multiple copies: Yes

# DuraCloud

- Annual membership/subscription fee(s): \$1,875-\$6,750, with 1TB storage
- Annual storage fee(s): \$700-\$1,400 for each additional TB, depending on subscription plan
- Upload and retrieval request fee: ?
- Data in fee: ?
- Data out fee: ?
- Equipment cost: n/a
- Staff time/sweat equity: ?
- Multiple copies: depends on subscription plan

# Ex Libris Rosetta

- Annual membership/subscription fee(s): ?
- Annual storage fee(s): ?
- Upload and retrieval request fee: ?
- Data in fee: ?
- Data out fee: ?
- Equipment cost: ?
- Staff time/sweat equity: ?
- Multiple copies: ?



# Google Nearline Storage

- Annual membership/subscription fee(s): \$0
- Annual storage fee(s): \$0.01/GB/month = \$10/TB/month = \$120/TB/year
- Upload and retrieval request fee: varies, starting at \$0.01/GB/month within multi-regional location
- Data in fee: \$0
- Data out fee: varies, starting at \$0.12/GB/month
- Equipment cost: n/a
- Staff time/sweat equity: ?
- Multiple copies: ?

# MetaArchive

- Annual membership/subscription 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
- Data in fee: \$0
- Data out fee: \$0
- Equipment cost: ~ \$5,500 in first year and every three years thereafter
- Staff time/sweat equity: ~.02 FTE
- Multiple copies: Yes

# OCLC Digital Archive

- Annual membership/subscription fee(s): ?
- Annual storage fee(s): ?
- Upload and retrieval request fee: ?
- Data in fee: ?
- Data out fee: ?
- Equipment cost: ?
- Staff time/sweat equity: ?
- Multiple copies: Yes

# Preservica (Cloud Edition)

- Annual membership/subscription 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 in fee: ?
- Data out fee: ?
- Equipment cost: ?
- Staff time/sweat equity: ?
- Multiple copies: Yes

# Takeaways

1. The digital preservation and storage services landscape is changing very rapidly.
2. Cost/pricing information can be hard to find (on the Web), collate, and compare.
3. Cloud storage is cheap. Cloud activity may not be.
4. Nobody is doing a very good job of devising an affordable and sustainable DP solution for small/mid-sized, poorly resourced LAMs (and this includes the community-based solutions).



# Thank You

Katherine Skinner

([katherine@educopia.org](mailto:katherine@educopia.org))

Matt Schultz

([schultzm@gvsu.edu](mailto:schultzm@gvsu.edu))

Aaron Trehub

([trehuaj@auburn.edu](mailto:trehuaj@auburn.edu))