Total Cost of Preservation (TCP)
Cost and Price Modeling for Sustainable Services

Rev. 2.2.1 – 2014-03-11

1 Introduction

Information technology and resources are thoroughly integrated with, and indispensable to, today's web-based commerce, science, education, and entertainment. The digital assets underpinning these activities, however, are inherently fragile with respect to ever increasing disruptive technological change. Without effective, affordable, and proactive curation management, today's digital assets will not remain viable and useful in the future. In order to sustain long-term preservation efforts and results, the *total cost of preservation* (TCP), that is, the full economic costs of preservation activities over archival timespans, must be well understood. The University of California Curation Center (UC3) [31] at the California Digital Library (CDL) [8], is engaged in the application of a TCP model in order to move many of its core service offerings to a cost recovery operational basis. A robust cost model is a necessary pre-condition for the development of a sustainable pricing scheme.

TCP analysis can be applied usefully in the development of two specific price models for preservation service pricing:

- Pay-as-you-go
- Paid-up

The pay-as-you-go model is appropriate for situations where a reliable and predictable annual income stream is available to the client purchasing preservation services. When this is not the case, for example, with organizations facing irregular/boom-or-bust budgetary cycles or for fixed term, grant-funded research projects, the paid-up model may be more attractive; indeed, in many circumstances it may be the only realistic option.

2 Cost and price modeling

Long-term digital preservation is a complex activity, involving sophisticated technical infrastructure as well as significant human competencies, analysis, and decision making.

Given the difficulties in accounting for the myriad aspects of preservation activity, and of projecting the cost of those activities into the indefinite future, any analysis of TCP must rely on a number of fundamental abstractions and assumptions about those activities and the environment in which they take place. (See Appendix A for a summary of prior work on preservation cost analysis.)

2.1 Preservation landscape

The TCP analytical framework abstracts preservation activities and their environment into the following 10 high-level categories, each of which is a cost component in TCP modeling:

1. Content ***Owners*** (or ***Creators*** or ***Curators***), with primary *intellectual* and *administrative* stewardship responsibility for digital content ***Collections***, use …
2. Submission ***Streams*** to send content to a …
3. Preservation ***System***, with a primary *technical* stewardship responsibility, composed of ***Ingest***, ***Data management***, and ***Access*** service functions, all running on …
4. ***Servers*** and making use of …
5. ***Storage***, to enable discovery and exploitation by …
6. Content ***Consumers*** (or ***Users***), with ongoing …
7. Preservation ***Analysis and planning*** and …
8. ***Interventions*** as necessary to ensure ongoing viability and accessibility, all subject to …
9. Direct operational ***Administration*** and …
10. High-level ***Management*** oversight.

This abstraction (see Figure 1) is based on the ISO 14721 Open Archival Information System (OAIS) reference mode [18], which is widely used within the digital library and preservation communities. (More information about OAIS is found in Appendix F.) However, the model and some of its terminology have been modified slightly in order to broaden its applicability and facilitate understanding by a non-specialist audience. (For example, the OAIS *producer* entity has been replaced with *owner*.) With appropriate definitional adjustment, the TCP model should be applicable to *any* online information service.

Each of these 10 categories can be further refined into subsidiary costs: capital and operational expenses (CAPEX/OPEX); fixed and marginal expenses; one-time, term, and annual expenses; etc. The full set of hierarchical categories can be examined in the accompanying spreadsheet that implements the TCP pay-as-you-go and paid-up price models [32]. The spreadsheet is fully parameterized to allow other institutions and service providers to incorporate appropriate assumptions and costs based on local policy decisions.

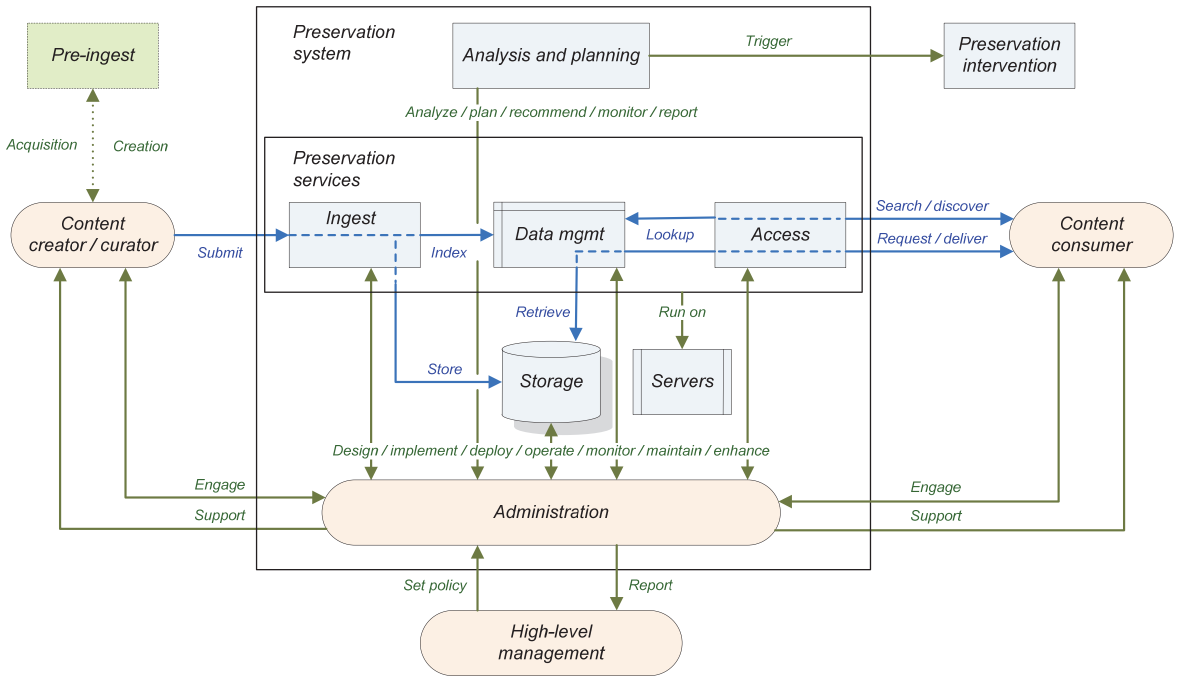

Figure 1 – TCP entities and interactions

A content *Owner* or *Curator* is an individual or corporate agent with primary *intellectual* stewardship responsibility for that body of content; the *technical* stewardship responsibility rests with the preservation *System*.

A submission *Stream* is the technical mechanism, and its accompanying workflow, used by *Owners* to transfer digital content to the preservation *System*.

A preservation *Intervention* is a non-routine, non-periodic, or unpredictable activity, such as a format migration, necessary to ensure ongoing access to content managed within the preservation *System*.

2.2 Assumptions

Properly accounting for the full economic costs of long-term preservation is a complicated undertaking. In order to make the analysis tractable, the TCP analysis makes the following simplifying assumptions:

1. Only the costs pertaining to preservation service *providers* are considered. In particular, the costs associated with the local activities of content *Owners* (e.g., content creation or acquisition, reformatting, packaging, submission, etc.) are considered out of scope. On the other hand, the cost of supporting *Owners* in making use of preservation *System* functions *is* in scope.
2. The *primary* customers of the preservation service provider are the content *Owners*, who are responsible for paying for utilized services and allocated and consumed resources. In particular, there is no assumption of payment from content *Consumers*. This arrangement reflects a common pattern of operation by cultural and scientific memory organizations and programs, for whom providing public access to managed content is a primary institutional goal.
3. Individual cost components of the model can be categorized unambiguously as either *fixed*, which are incurred regardless of level of use, or *marginal*, which scale proportionally with use.
4. Individual cost components can be categorized unambiguously as *one-time*, *term*, or *annual*. Term costs are those incurred for an activity or capacity for a fixed number of years. Term costs can be *annualized* on an inflation-adjusted basis over their effective lifespan. (See Appendix E.)
5. The values defined for individual cost components represent *nominal* costs, that is, the costs defined for generic instances of activities or system capacities. This is reasonable under a further assumption of a policy of uniformity of preservation *effort* rather than *outcome*. For example, on the basis of its form, structure, accompanying metadata, etc., some digital content may be more inherently amenable to preservation care and will naturally receive a higher level of preservation service and outcome.
6. Most routine preservation actions performed on content, e.g., characterization, fixity, normalization, etc., are substantially automated or automatable. Thus, the main costs associated with an action are in the acquisition and deployment of the implementing software, which is *independent* of the number of objects against which the action is performed.
7. The preservation service provider can carry forward budgetary *surpluses* across fiscal year boundaries. Furthermore, these surplus funds can be *invested* at market rates, with the income further contributing to the surplus.

8. The model is ultimately *revenue neutral*, that is, there should be no surplus (or deficit) funds remaining at the end of the period under consideration.
9. Values can be determined for various *annual adjustment factors*, such as inflation, investment rate of return, cost of living adjustments (COLA), merit pay raises, changes to the per-unit cost of preservation, etc., which, although held *constant* over the full period of TCP consideration, are nevertheless reliably *predictive* of long-term economic, technological, and organizational trends.

The last assumption, although the mainstay of the standard economic forecasting technique of *discounted cash flow* (DCF) analysis [10], can be problematic over extended time periods.

2.2 Total cost of preservation

The total cost of preservation (TCP) to a preservation service provider across all of its content *Owners* is represented numerically as:

$$TCP = n \cdot O + m \cdot T + A + \ell \cdot R + k \cdot S + j \cdot C + G + i \cdot V + D + M \quad (1)$$

where:

TCP	Total cost of preservation (incurred by the service provider across all *Owners*).
n	Number of content *Owners* or *Creators*.
O	Unit cost of an *Owner* or *Creator*.
m	Number of unique submission *Streams*.
T	Unit cost of a *Stream*.
A	Fixed cost of the preservation *System*.
ℓ	Number of *Servers*.
R	Unit cost of a *Server*.
k	Number of units of *Storage*.
S	Unit cost of *Storage*.
j	Number of content *Consumers*.
C	Unit cost of a *Consumer*.
G	Fixed cost of *Analysis and Planning*.
i	Number of preservation *Interventions*.
V	Unit cost of an *Intervention*.
D	Fixed cost of *Administration*.

M	Fixed cost of *Management*.

The preservation *System*, *Administration*, and *Management* are considered *fixed* costs, as they are (essentially) incurred regardless of level of adoption. Thus they are represented in Eq. (1) by single terms, A, D, and M, respectively. All other components are *marginal* costs, which (essentially) scale with use, and thus are each represented by two terms: a unit *cost* and the *number* of units consumed or allocated.

2.3 Pay-as-you-go price model

The pay-as-you-go price is based on an annual billing cycle, $t = 0,1,2,\ldots$ The per-unit *Owner*, *Streams*, and actual *Storage* cost components are specific to a given *Owner*. All other cost components – the preservation *System*, *Servers*, content *Consumers*, ongoing *Analysis and planning*, *Interventions*, *Administration*, and *Management* – are considered “common good” costs; that is, they are generally applicable and beneficial to all content *Owners* equally, and as such, are properly apportioned across all *Owners*. In Eq. (2) the terms for these components are collected together and divided by n, the number of *Owners*.

$$X_t = \frac{A + \ell \cdot R + j \cdot C + i \cdot V + G + D + M}{n} + O + m_O \cdot T + k_O \cdot S \qquad (2)$$

where:

X_t	Pay-as-you-go price of a given *Owner* in year t.
m_O	Number of submission *Streams* attributable to the *Owner*.
k_O	Number of units of *Storage* attributable to the *Owner*.

2.3.1 Cost categories

It may useful to distinguish subsets of costs on the basis of the putative source of funding for the underlying activity or capacity. This requires splitting some of the terms in Eqs. (1) and (2) into constituent parts reflecting meaningful administrative distinctions. For example, the following breakdown of costs may reflect the administrative reality for many academically-based preservation service providers.

1. Common good.
    a. *Development*. Annualized costs of major technical development of the preservation *System* and its subsidiary *Services*, generally funded through external grants.
    b. *Operation*. Annual costs of operating the *System*, supporting content *Consumers*, and preservation *Administration* and *Management,* generally funded by central University funds.
    c. *Infrastructure*. Annual costs of operating the *Servers* underlying the *System*, generally funded by central University funds.
    d. *Intervention*. One-time costs associated with a preservation *Intervention* necessary to ensure ongoing viability of and accessibility to managed content. Due to the irregular necessity for preservation *Interventions*, their cost may be best covered through special one-time *subventions*.

    Note that in the case of UC3, cost categories 1(b) and (c) have essentially been pre-paid by the 10 UC campuses as part of the funding for central University service units such as UC3.

2. *Owner*-specific costs.
    a. *Customer Relations Management (CRM)*. One-time costs of accepting and registering for use a new *Owner* and the submission *Streams* used by that *Owner*, generally funded by direct billing to the *Owner*.
    b. *Dedicated resources*. Annual or annualized costs of resources solely dedicated for use by an *Owner*, for example, *Storage*, generally funded by direct billing to that *Owner.*

The modified pay-as-you-go formulas for these six cost categories are derived in Appendix B.

2.4 Paid-up price model

The paid-up price model is based on a one-time, up-front payment, E, sufficiently large to pay for all subsequent preservation activities over a term of T years, assuming an annual investment rate of return, r, applicable to surplus funds and a compounding discount (or markup) factor, d, applicable to the cost of the activities.

$$E_T = X_0 \cdot \frac{(1+r)^T - (1-d)^T}{(1+r)^T \cdot (r+d)} \qquad (3)$$

where:

E_T	Paid-up price for T years.
X_0	Pay-as-you-go price established for the *base* year $t = 0$.
r	Investment rate of return, as a decimal (2% = 0.02).
d	Discount (or markup) factor, as a decimal.
T	Term, in years.

See Appendix C for the derivation of Eq. (3).

The determination of the annual discount/markup factor, d, is highly dependent on the nature of the underlying cost component. It may be useful to assign a different factor to each of the cost categories defined in Section § 2.3.1, based on the following assumptions:

1. The cost of staff-dependent components, for example, *Owner* and *Consumer* CRM (customer relationship management) and *Analysis and planning*, tends to *increase* over time on the basis of staff cost-of-living (COLA) and merit salary increases, although this increase may be mitigated by increases in staff *productivity* resulting from improved tools and techniques.
2. The cost of infrastructure-dependent components, for example, *Servers* and *Storage*, tends to *decrease* over time, at least is has historically through steady increases in semiconductor and storage media density, as expressed through the well-known Moore's law [17] and Kryder's law [35].

A modification to Eq. (8) incorporating distinct discount factors is derived in Appendix C.

The TCP paid-up cost model is based on the economic forecasting technique of *discounted cash flow* (DCF) [10]. Its predictive reliability is highly dependent on the values specified for r and d. In general, the longer the term T, the more difficult it is to estimate values for r and d that will accurately reflect changing economic and technological trends over T years. In light of this, UC3 uses $T = 10$ as the maximum paid-up term. See [1] and [28] for alternative stochastic modeling approaches.

It is desirable for an *Owner* to be able to switch easily between the pay-as-you-go and paid-up price basis. The relevant factor in this transition is the "coefficient of persistence," the one time premium or multiplier of the then current pay-as-you-go price to achieve paid-up persistence for T years.

$$\varphi_T = \frac{E_T}{X_t} = \frac{(1+r)^T - (1-d)^T}{(1+r)^T \cdot (r+d)} \tag{4}$$

φ_T	Coefficient of persistence for T years.

Assuming that sufficient funds are available for the one-time payment, the paid-up price basis is fiscally more attractive whenever $\varphi_T < T$, as that indicates that the paid-up price is less than T times the pay-as-you-go price.

3 Responding to predictive failure

In the event that revenues are *not* sufficient to cover expenses, a preservation service provider can attempt to implement a number of prospective ameliorating strategies.

1. *Raise* additional revenues through:
    a. *Permanently* increased institutional funding levels.
    b. *One-time* or *time-limited* grant funding or special institutional subventions.
    c. Direct or indirect monetization of content *consumption*.

2. *Lower* costs through:
    a. Reduction in the aggregate *complexity* of the content under management through a reduction in any or all of the:
        i. *Number*,
        ii. *Size*, or
        iii. *Heterogeneity*

        of the managed content. Note that any decision to *deaccession* content may temporarily *increase* operational costs, especially if manual selection is necessary.

    b. Lower *service levels* applied to managed content through:
        i. Downgrading from *on-line* to *near-line* or *off-line* content access.
        ii. Downgrading from *bright* to *dark* content access.
        iii. Downgrading from *functional* to *bit-level* preservation.

    c. Lower *technology* costs through:
        i. Increasing *automation* of routine procedures. Note that this may be facilitated through strategy 2(a)(iii), reduction in content *heterogeneity*.

ii. Extending the *service life* of infrastructural components such as *Servers* and *Storage*. Note that this may lead to increased technology failures with concomitantly increased costs.

iii. Replacing infrastructural components with less expensive replacements, for example, moving from disk-based on-line access to tape-based near-line access. Note that this may necessitate a concomitant adoption of strategy 2(b), reduction in supported service level.

iv. Exploitation of infrastructural *virtualization* under the assumption that it is less costly than physical *Services* or *Storage*.

Note that any infrastructural *migration* may temporarily *increase* costs.

d. Lower *staff* costs through:

i. Staffing *reductions* through natural attrition or targeted layoffs. Note that increasing individual staff workloads may be counterproductive with respect to overall productivity and cost reduction.

ii. Increasing *productivity* levels through increased professional development and training and the use of better tools and greater automation.

Additionally, a service provider could minimize the likelihood of funding shortfalls by building in to its price schedule a small *contingency surcharge*, σ, for purposes of building up a reserve, or “rainy day” fund, encumbered for future use in responding to unexpected conditions. Eqs. (2) and (3) can be rewritten to incorporate the surcharge term as a percentage of the baseline prices:

$$X'_t = (1 + \sigma) \cdot X_t \tag{5}$$

$$E'_T = X'_0 \cdot \frac{(1 + r)^T - (1 - d)^T}{(1 + r)^T \cdot (r + d)} \tag{6}$$

X'_t	Pay-as-you-go price incorporating contingency surcharge.
E'_T	Paid-up price incorporating contingency surcharge.

4 Conclusion

An understanding of the full economic costs of offering preservation and curation services is necessary to the long-term sustainability of those services and service providers. The Total Cost of Preservation (TCP) model presented in this paper provides a useful analytical framework for assessing and accounting for the full range of costs attendant to preservation and curation services, and has been used to derive two price models: pay-as-you-go, which is most appropriate for customers with reliable and predictable sources of continuing funding; and paid-up, which is most appropriate for customers with term-limited funding.

The accompanying spreadsheet developed by UC3 to implement the TCP model is fully parameterized and can be customized easily by other organizations to reflect local conditions, assumption, policies, and policies [32].

Appendix A Prior work

A number of international efforts have studied the question of long-term preservation costs; most notably, the Nationaal Archief of the Netherlands in 2005 [24]; the LIFE[2] project (Life Cycle Information for E-Literature) work on a Generic Preservation Model (GPM) in 2008 [6][16][21]; the KRDS project (Keeping Research Data Safe) in 2010 [7][9]; the Princeton DataSpace initiative in 2010 [12][27]; the Danish National Archives and Royal Library CMDP (Cost Model for Digital Preservation) in 2011 [11][19]; the ENSURE project in 2011 [13]; the APARSEN project [3]; and the 4C project in 2013 [1]. (See [36] for a summary of these and other relevant activities; see [22][26] for bibliographies of the cost modeling literature.)

All of these models analyze preservation costs throughout the full lifecycle of preserved assets. Both the Nationaal Archief and LIFE project work employ very fine-grained analysis of cost components and are based on representative *actual* costs rather than the *nominal* costs employed by TCP. The LIFE model includes cost components for content creation and acquisition, which are considered out of scope in the TCP analysis. Furthermore, LIFE scales preservation action costs by the number of expected objects. TCP assumes that there is no per-object marginal cost; instead, marginal costs are associated only with the various *Content Types* of which objects are members. The KRDS model is specific to the research data lifecycle, but its findings appear to be applicable to other contexts. Unlike the Nationaal Archief and LIFE models, KRDS assumes a discounting function that annually decreases the aggregate cost of preservation service. The DataSpace model is also based on a discounting function, but its analysis covers only the costs associated with preservation storage, which are defined on a pay once, store forever (POSF) basis. The OAIS activity-based CMDP work concentrates on post facto measuring of preservation costs, rather than on forecasting, although it is possible using the framework. While DataSpace is explicitly concerned with “forever” pricing, neither it nor any of the other models assume the benefit of an annual investment return in offsetting a portion of ongoing costs. Most, if not all, of the individual cost components of these models can be mapped to OAIS environmental and functional entities or TRAC criteria [25], facilitating common points of reference and comparison. The APARSEN project bases its work on the ISO 16363 standard for trusted digital repositories [19], a formalization of TRAC. The ENSURE project has identified rigorous cost/benefit analysis as a significant missing factor from most cost modeling efforts [36]. The 4C project has also completed a comprehensive evaluation of a number of cost models [30] to identify common characteristics and gaps in preparation of an economic sustainability reference model [21].

A number of institutions, both commercial and non-profit/academic, now offer long-term preservation services. (Or at least persistent storage; the important distinction between the two is not always apparent from the description of these service offerings.) Carbonite is representative of commercial preservation service offerings. Its cost is $599/year for up to 500 GB, with further increments of 100 GB available for $89/year. Amazon S3 pricing is complex, starting at $0.14/GB/month for the first TB with discounts for additional amounts, and additional transactional charges for access [5]. The Amazon Glacier service provides near-line storage at a significantly discounted price of $0.011/GB/month, with transaction frees being applied when access requests exceed a 5% monthly threshold [3]. The Chronopolis repository charges $1,500/TB/year [29]. The Princeton DataSpace repository charges a one-time fee of $0.006/MB [26]. The DuraCloud service has a variety of price points ranging from $1,500/TB/year to $6,900/TB/year for basic through enterprise service [12]. Note that these are the baseline prices for the first TB, with a lesser rate for subsequent use and further price decreases for content over 10 TB in size. The HathiTrust storage-based pricing for content-contributing partners is $3.29/GB/year plus a one-time fee of 25% the of the first year cost [15]. The LifeTime Library at the University of North Carolina offers students permanent storage “and associated services” for 250 GB, apparently with no associated fees [33]. The USC Digital Repository offers a paid up license for 20 years of preservation service for $1,000/TB [34]. In general, however, little information is available explaining the basis of their business models.

Appendix B Pay-as-you-go cost categories

It may useful to distinguish subsets of costs on the basis of the putative source of funding for the underlying activity or capacity. This requires splitting some of the terms in Eqs. (1) and (2) into constituent parts reflecting meaningful administrative distinctions. For example, the following breakdown of costs may reflect the administrative reality for many preservation service providers.

1. Common good.
   a. *Development*. Annualized costs of major technical development of preservation *System* and its subsidiary *Services*, generally funded through external grants.
   b. *Operation*. Annual costs of operating the *System*, supporting content *Consumers*, and preservation *Administration* and *Management,* generally funded by central University funds.
   c. *Infrastructure*. Annual costs of operating the *Servers* underlying the *System*, generally funded by central University funds.
   d. *Intervention*. One-time costs associated with a preservation *Intervention* necessary to ensure ongoing viability of and accessibility to managed content, possibly funded by special one-time subventions or surcharges.
2. *Owner*-specific costs.
   a. *Customer Relations Management (CRM)*. One-time costs of accepting and registering for use a new *Owner* and the submission *Streams* used by the *Owner*, generally funded by direct billing to the *Owner*.
   b. *Dedicated resources*. Annual or annualized costs of resources solely dedicated for use by an *Owner*, for example, *Storage*, generally funded by direct billing to that *Owner.*

$$Xd = \frac{A_D}{n} \tag{B.1}$$

$$Xo = \frac{A_O + j \cdot C + G + D + M}{n} \tag{B.2}$$

$$Xr = \frac{\ell \cdot R}{n} \tag{B.3}$$

$$Xv = \frac{i \cdot V}{n} \quad \text{(B.4)}$$

$$Xc = O + m_O \cdot T \quad \text{(B.5)}$$

$$Xs = k_O \cdot S \quad \text{(B.6)}$$

where:

Xd	Pay-as-you-go common good development cost.
A_D	Pay-as-you-go common good *Archive* development cost.
Xo	Pay-as-you-go common good operational cost.
A_O	Pay-as-you-go common good *Archive* operational cost
Xr	Pay-as-you-go common good infrastructure cost.
Xv	Pay-as-you-go common good *Intervention* cost.
Xc	Pay-as-you-go *Producer* CRM cost.
Xs	Pay-as-you-go Producer dedicated resource cost.

Appendix C Paid-up price model derivation

The paid-up price model is based on a one-time, up-front payment, E_T, sufficiently large to pay for all subsequent preservation activities over a term of T years, assuming an annual investment rate of return, r, or surplus funds and a compounding discount (or markup) factor, d, applicable to the cost of the activities. This can be illustrated by the cash flow diagram (CFD) in Figure 4.

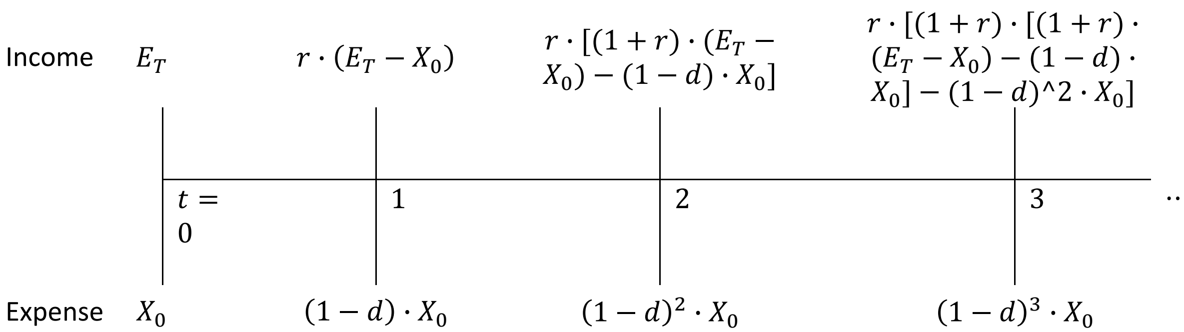

Figure C.1 – TCP paid-up price cash flow diagram

At beginning of the base year ($t = 0$) there is the one-time payment E_T and expense X_0. At year end there is the additional income generated by investing the surplus funds, $E_T - X_0$, with a return rate of r, and the following years expenses, which is the initial year cost adjusted by the compounding discount factor, d.

Since the paid-up price model is intended to be ultimately revenue neutral, at the end of the term the net financial position, equal to the previous year's net plus the current year's income minus expenses, should be 0. Thus, at the end of year T:

$$0 = (1+r) \cdot [(1+r) \cdot [(1+r) \cdot \cdots \cdot [(1+r) \cdot (E_T - X_0) - (1-d) \cdot X_0] - (1-d)^2 \cdot X_0] - \cdots - (1-d)^2 \cdot X_0] \qquad \text{(C.1)}$$

Solving Eq. (C.1) for E_T gives the one-time price.

$$0 = (1+r)^T \cdot (E_T - X_0) - (1+r)^{T-1} \cdot (1-d) \cdot X_0 - (1+r)^{T-2} \cdot (1-d)^2 \cdot X_0 - \cdots - (1+r) \cdot (1-d)^{T-1} \cdot X_0 \qquad \text{(C.2)}$$

$$(1+r)^T \cdot (E_T - X_0) = (1+r)^{T-1} \cdot (1-d) \cdot X_0 + (1+r)^{T-2} \cdot (1-d)^2 \cdot X_0 + \cdots + (1+r) \cdot (1-d)^{T-1} \cdot X_0 \qquad \text{(C.3)}$$

$$(1+r)^T \cdot E_T = (1+r)^T \cdot X_0 + (1+r)^{T-1} \cdot (1-d) \cdot X_0 + (1+r)^{T-2} \cdot (1-d)^2 \cdot X_0 + \cdots + (1+r) \cdot (1-d)^{T-1} \cdot X_0 \qquad \text{(C.4)}$$

$$(1+r)^T \cdot E_T = \sum_{t=0}^{T-1} (1+r)^{T-t} \cdot (1-d)^t \cdot X_0 \tag{C.5}$$

$$E_T = \sum_{t=0}^{T-1} X_0 \cdot \frac{(1+r)^{T-t} \cdot (1-d)^t}{(1+r)^T} \tag{C.6}$$

$$E_T = \sum_{t=0}^{T-1} X_0 \cdot \frac{(1-d)^t}{(1+r)^t} \tag{C.7}$$

An equation of the form of Eq. (C.7) is known as a *power series* and has the general solution derived in Appendix D.

$$E_T = X_0 \cdot \frac{1 - \left(\frac{1-d}{1+r}\right)^T}{1 - \frac{1-d}{1+r}} \tag{C.8}$$

$$E_T = X_0 \cdot \frac{(1+r)^T - (1-d)^T}{(1+r)^T \cdot (r+d)} \tag{C.9}$$

As discussed in Section § 2.3.1, distinct discount factors should be determined for the six cost categories: development, operations, infrastructure, intervention, CRM, and storage. Thus, a more accurate formulation for the paid-up price is:

$$\begin{aligned} E_T = {} & Xd_0 \cdot \frac{(1+r)^T - (1-d_D)^T}{(1+r)^T \cdot (r+d_D)} + Xo_0 \cdot \frac{(1+r)^T - (1-d_O)^T}{(1+r)^T \cdot (r+d_O)} + \\ & Xr_0 \cdot \frac{(1+r)^T - (1-d_R)^T}{(1+r)^T \cdot (r+d_R)} + Xv_0 \cdot \frac{(1+r)^T - (1-d_V)^T}{(1+r)^T \cdot (r+d_V)} + \\ & Xc_0 \cdot \frac{(1+r)^T - (1-d_C)^T}{(1+r)^T \cdot (r+d_C)} + Xs_0 \cdot \frac{(1+r)^T - (1-d_S)^T}{(1+r)^T \cdot (r+d_S)} \end{aligned} \tag{C.10}$$

where:

Xd_0	Pay-as-you-go development cost established for the *base* year $t = 0$.
d_D	Annual development discount (or markup) factor, as a decimal.
Xo_0	Pay-as-you-go operational cost for the base year.
d_O	Annual operational discount factor.
Xr_0	Pay-as-you-go infrastructure cost for the base year.

d_R	Annual infrastructure discount factor.
Xv_0	Pay-as-you-go intervention cost for the base year.
d_V	Annual intervention discount factor.
Xc_0	Pay-as-you-go CRM cost for the base year.
d_C	Annual CRM discount factor.
Xs_0	Pay-as-you-go storage cost for the base year.
d_S	Annual storage discount factor.

The revenue neutrality of the paid-up pricing is illustrated in the example in Table C.1, with a ten year term ($T = 10$), an initial pay-as-you-go cost of $390 ($X_0 = \$390$), an annual investment return of 2% ($r = 0.02$), an annual aggregate discount in providing preservation service of 5% ($d = 0.05$), and a paid-up price of $2,892 ($X_{10} = \$2{,}892$). As shown, at the end of the 10th year ($t = 9$), the pricing yields a surplus of $0. (All figures rounded to the nearest dollar).

Year **(t)**	***Income***	***Expense***	***Surplus***
0	$ 2,892	$ 390	$ 2,502
1	$ 50	$ 371	$ 2,182
2	$ 44	$ 352	$ 1,873
3	$ 37	$ 334	$ 1,576
4	$ 32	$ 318	$ 1,290
5	$ 26	$ 302	$ 1,014
6	$ 20	$ 287	$ 748
7	$ 15	$ 272	$ 490
8	$ 10	$ 259	$ 241
9	$ 5	$ 246	$ 0

Table C.1 – Paid-up price model revenue neutrality example

Appendix D Power series

An equation of the form:

$$p(x, N) = a + a \cdot x + a \cdot x^2 + \cdots + a \cdot x^{N-1} \quad \text{(D.1)}$$

which can be represented in compact form as:

$$p(x, N) = \sum_{i-0}^{N-1} a \cdot x^i \quad \text{(D.2)}$$

is known as a *power series*. Its general solution is derived as follows: Eq. (D.1) can be scaled by a factor of x:

$$x \cdot p(x, N) = a \cdot x + a \cdot x^2 + \cdots + a \cdot x^{N-1} + a \cdot x^N \quad \text{(D.3)}$$

Subtracting Eq. (D.3) from (D.1) yields:

$$p(x. N) - x \cdot p(x, N) = a + (a \cdot x - a \cdot x) + (a \cdot x^2 - a \cdot x^2) + \cdots + (a \cdot x^{N-1} - a \cdot x^{N-1}) - a \cdot x^N \quad \text{(D.4)}$$

Most of the terms cancel, leaving:

$$p(x. N) - x \cdot p(x, N) = a - a \cdot x^N \quad \text{(D.5)}$$

which can be rearranged to form the general power series solution:

$$(1 - x) \cdot p(x. N) = a - a \cdot x^N \quad \text{(D.6)}$$

$$p(x. N) = \frac{a - a \cdot x^N}{1 - x} \quad \text{(D.7)}$$

$$p(x. N) = a \cdot \frac{1 - x^N}{1 - x} \quad \text{for } 0 \leq x \leq 1 \quad \text{(D.8)}$$

An *infinite* power series, one for which $N = \infty$ can be solved by observing that $x^N \rightarrow 0$ as $N \rightarrow \infty$ so long as x satisfies the convergence criterion $0 \leq x \leq 1$. Thus:

$$p(x, \infty) = \lim_{N \rightarrow \infty} p(x, N) = a \cdot \frac{1}{1 - x} \quad \text{for } 0 \leq x \leq 1 \quad \text{(D.9)}$$

Various other equations in the TCP model are derived by substituting appropriate terms for a and x into Eqs. (D.8) and (D.9).

Appendix E Annualizing term costs

Many preservation costs are incurred for an activity or capacity with an effective lifespan of more than one year, for example, preservation analysis or servers or storage. In the TCP model these costs are *annualized* on an inflation-adjusted basis over their effective lifespan. For example, suppose a preservation service provider procures a *Server* for $5,000 with the intention of replacing it after five years. The *nominal* annualized cost is $5,000/5 years = $1,000/year. But this does not take into account the inflationary effect on the annualized cost, since future dollars are worth less than current dollars. The *net present value*, NPV, of series of T nominally-equal annual cash flow amounts, c, subject to an inflation rate, f, is:

$$NPV = c + \frac{c}{1+f} + \frac{c}{(1+f)^2} + \cdots + \frac{c}{(1+f)^{T-1}} \tag{E.1}$$

Eq. (E.1) is a power series that can be solved by substituting c for a, $1+f$ for x, and T for N in Eq. (D.8) of Appendix D.

$$NPV = c \cdot \frac{1 - \frac{1}{(1+f)^T}}{1 - \frac{1}{1+f}} \tag{E.2}$$

which can be simplified to:

$$NPV = c \cdot \frac{\frac{(1+f)^T - 1}{(1+f)^T}}{\frac{f}{1+f}} \tag{E.3}$$

$$NPV = c \cdot \frac{(1+f)^T - 1}{(1+f)^T} \cdot \frac{1+f}{f} \tag{E.4}$$

$$NPV = c \cdot \frac{(1+f)^T - 1}{f \cdot (1+f)^{T-1}} \tag{E.5}$$

In the TCP case the NPV amount is known for the base year $t = 0$, so Eq. (D.5) can be solved for the unknown adjusted annualized cost c.

$$c = NPV \cdot \frac{f \cdot (1+f)^{T-1}}{(1+f)^T - 1} \tag{E.6}$$

In the example above, the adjusted annualized value is $1,039.99, assuming 2% inflation over 5 years. That is, the total nominal amount $5,199.85 collected as five annual payments of $1,039.99 has the net present value in the base year $t = 0$ of $5,000.

Appendix F Open Archival Information System (OAIS)

The ISO 14721 Open Archival Information System (OAIS) reference model provides a useful reference for the development of a comprehensive archives "consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community" [18].

The simplified OAIS entities and interactions used by TCP are depicted in Figures F.1 and F.2.

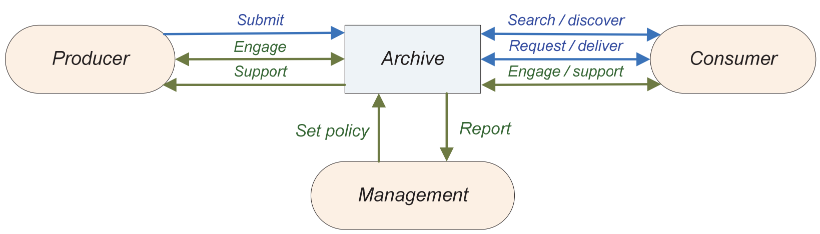

Figure F.1 – High-level OAIS entities and interactions

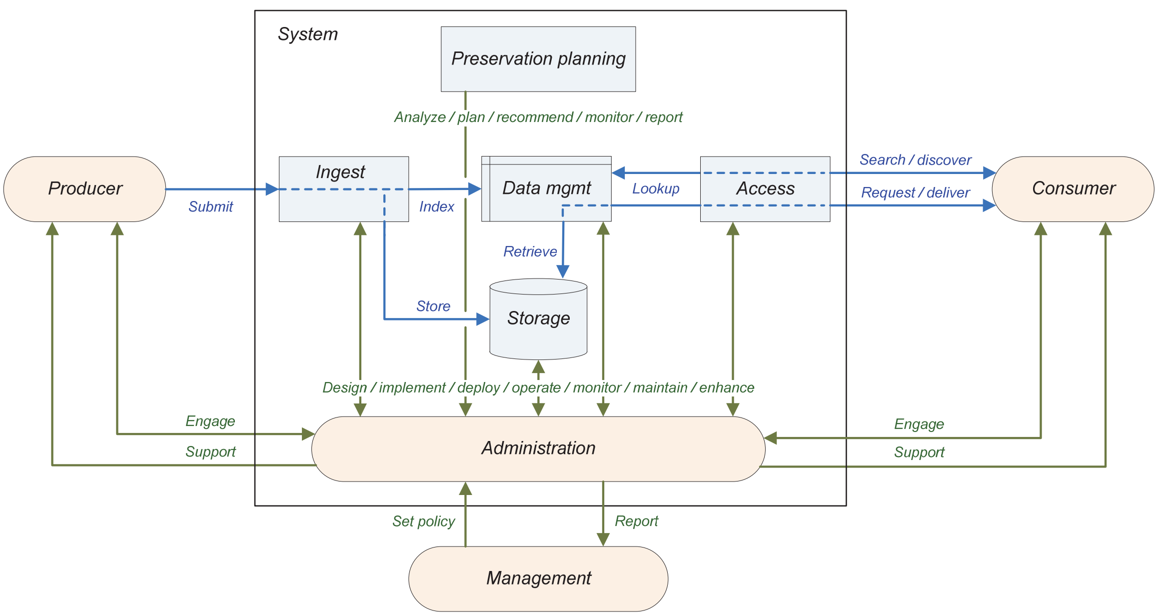

Figure F.2 – Intermediate OAIS entities and interactions

For purposes of accounting for and allocating the full economic cost of long-term preservation, the TCP model (see Figure 1) distinguishes between a narrowed *System* entity, composed of the *Ingest, Data management*, and *Access* services, and the *Servers* or infrastructural platform on which the *System* operates and the *Storage* that the preserved content occupies.

References

[1] 4C Project (2013), *4C: Collaboration to Clarify the Cost of Curation* <http://www.4cproject.eu/>.

[2] Addis, Matthew (2012), "Cost modeling for forever storage," *PASIG 2012*, Dublin, October 16-19.

[3] Alliance for Permanent Access (2012), *Report on Cost Parameters for Digital Repositories* <http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2013/03/APARSEN-REP-D32_1-01-1_0.pdf>.

[4] Amazon, *Amazon Glacier Pricing*, accessed 2012-11-02 <http://aws.amazon.com/glacier/pricing>.

[5] Amazon, *Amazon S3 Pricing*, accessed 2012-11-02 <http://aws.amazon.com/s3/pricing>.

[6] Ayris, Paul, Richard Davies, Rory McLeod, Rui Miao, Helen Shenton, and Paul Wheatley (2008), *LIFE2 Final Project Report*, <http://eprints.ucl.ac.uk/11758>.

[7] Beagrie, Neil, Brian Lavoie, and Matthew Woollard (2010), *Keeping Research Data Safe 2* <http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf>.

[8] California Digital Library (2012), *California Digital Library* <http://www.cdlib.org/>.

[9] Charles Beagrie (2011), *Keeping Research Data Safe (KRDS)* <http://www.beagrie.com/krds.php>.

[10] Damodaran, Aswath (2002), *Investment Valuation: Tools and Techniques for Determining the Value of Any Asset* (2n ed.; New York, Wiley).

[11] Danish National Archives/Danish Royal Library (2011), *Cost Model for Digital Preservation (CMDP)* <http://www.costmodelfordigitalpreservation.dk/>, accessed 2011-09-02. See also, <http://www.costmodelfordigitalpreservation.dk/contact/cmdp-1---preservation-planning-and-digital-migrations/documentation/Kostenmodel_in_Excel_versie_1.2_%28final%29.xls> and <http://www.costmodelfordigitalpreservation.dk/contact/cmdp-2---ingest-and-archival-storage/documentation/Tool_Version_2.xls>.

[12] DuraSpace (2012), *DuraCloud Subscription Plans*, accessed 2012-11-02 <http://www.duracloud.org/pricing>.

[13] ENSURE project (2011), *ENSURE – Enabling Knowledge Sustainability, Usability, and Recovery for Economic Value* <http://ensure-fp7-plone.fe.up.pt/site>.

[14] Goldstein, Serge J., and Mark Ratliff (2010), *DataSpace: A Funding and Operational Model for Long-Term Preservation and Sharing of Research Data*, <http://arks.princeton.edu/ark:/88435/dsp01w6634361k>.

[15] HathiTrust (2012), *Cost* <http://www.hathitrust.org/cost>.

[16] Hole, Brian, Li Lin, Patrick McCann, and Paul Wheatley (2010), "LIFE^3: A predictive costing tool for digital collections," *7th International Conference on Preservation of Digital Objects (iPRES2010)*, Vienna, September 19-24, 2010 <http://www.life.ac.uk/3/docs/Ipres2010_life3_submitted.pdf>.

[17] Intel Corporation (2005), *A Conversation with Gordon Moore: Moore's Law* <ftp://download.intel.com/museum/Moores_Law/Video-Transcripts/Excepts_A_Conversation_with_Gordon_Moore.pdf>.

[18] ISO 14721 (2003), *Space data and information transfer systems – Open archival information system – Reference model*. See also, CCSDS 650.0-B-1 (2002), Blue Book <http://public.ccsds.org/publications/archive/650x0b1.pdf>.

[19] ISO 16363 (2012), *Space data and information transfer systems – Audit and certification of trustworthy digital repositories <http://public.ccsds.org/publications/archive/652x0m1.pdf>*.

[20] Kejser, Ulla Bøgvad, Anders Bo Nielsen, and Alex Thirifays (2011), "Cost model for digital preservation: Cost of digital migration," *International Journal of Digital Curation* 6(1): 255-267 <http://www.ijdc.net/index.php/ijdc/article/viewFile/177/246>.

[21] Lavoie, Brian, and Neil Grindley (2013), *A Draft Economic Sustainability Reference Model* <http://www.4cproject.eu/component/docman/doc_download/32-ms9-draft-esrm>.

[22] Library of Congress (2012), "A digital asset sustainability and preservation cost bibliography," *The Signal: Digital Preservation* <http://blogs.loc.gov/digitalpreservation/2012/06/a-digital-asset-sustainability-and-preservation-cost-bibliography/>.

[23] LIFE project (2011), *LIFE: Life Cycle Information for E-Literature* <http://www.life.ac.uk/>.

[24] Nationaal Archief (2005), *Costs of Digital Preservation*, <http://www.nationaalarchief.nl/sites/default/files/docs/kennisbank/codpv1.pdf>, accessed 2011-09-02. See also, <http://www.nationaalarchief.nl/sites/default/files/docs/kennisbank/kostenmodel_in_excel_versie_1.2_final.xls>.

[25] OCLC/Center for Research Libraries (2007), *Trusted Repository Audit & Certification: Criteria and Checklist* <www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf>.

[26] Open Planets Foundation (2013), *Digital Preservation and Data Curation Costing and Cost Modeling* <http://wiki.opf-labs.org/display/CDP/Home>.

[27] Princeton University (2009), *DataSpace Policies and Guidelines*, July 15 <http://dataspace.princeton.edu/jspui/about/DataSpacePnG.pdf>.

[28] Rosenthal, David S. H., Daniel C. Rosenthal, Ethan L. Miller, Ian F. Adams, Mark W. Storer, and Erez Zadok (2012), "The economics of long-term digital storage," *Memory of the*

World in the Digital Age: Digitization and Preservation, Vancouver, September 26-28, 2012 <http://www.lockss.org/locksswp/wp-content/uploads/2012/09/unesco2012.pdf>.

[29] San Diego Supercomputer Center, *Chronopolis Pricing*, accessed 2012-11-02 <http://chronopolis.sdsc.edu/pricing>.

[30] Stokes, Paul, ed. (2014), *Evaluation of Cost Models and Needs & Gap Analysis* <http://www.4cproject.eu/component/docman/doc_download/34-d3-1-evaluation-of-cost-models-and-needs-gaps-analysis?Itemid=>.

[31] University of California Curation Center (2012), *University of California Curation Center* <http://www.cdlib.org/uc3>.

[32] University of California Curation Center (2012), *Cost Modeling* <https://confluence.ucop.edu/display/Curation/Cost+Modeling>.

[33] University of North Carolina (2011), *Life Time Library thought to be the first nationwide*, September 15, 2011 <http://sils.unc.edu/news/2011/SILS-lifetime-library>.

[34] University of Southern California (2011), *USC Digital Repository: Pricing* <http://repository.usc.edu/pricing/>.

[35] Walter, Chip (2005), “Kryder’s law,” *Scientific American* 293 (July 25): 32-33 <http://www.scientificamerican.com/article.cfm?id=kryders-law >.

[36] Xue, Pengfei, Mohamed Badawy, Essam Shehab, and Paul Baguley (2011), "Cost modeling for long-term digital preservation: Challenges and issues," *9th International Conference on Manufacturing Research*,

[37] Zeller, Jean-Daniel (2010), “Cost of digital archiving: Is there a universal model?” *8th European Conference on Digital Archiving*, Geneva, April 28-30, 2010 <http://regarddejanus.files.wordpress.com/2010/05/costsdigitalarchiving-_jdz_eca2010.pdf>.