Pay Once, Preservation Forever
A "Paid-Up" Cost Model for Long-Term Preservation
http://wiki.ucop.edu/display/Curation/Cost+Modeling

Stephen Abrams, Patricia Cruse, and John Kunze

UC Curation Center
California Digital Library, University of California
http://www.cdlib.org/uc3

Sustainable services in an era of budgetary constraint

Digital preservation and curation are rapidly maturing disciplines, drawing on a rich set of community best practices, tools, and service providers to ensure the long-term viability of the digital assets that thoroughly pervade all aspects of contemporary culture, commerce, science, education, and entertainment. In many ways, however, the most significant risk to that viability is financial, rather than technical. Unlike the conservation of analog materials, the effective preservation of digital resources necessitates ongoing and proactive intervention, and any interruption in these activities may result in irretrievable data loss. In an era of significant budgetary constraints, however, many institutions are finding it difficult to dedicate ongoing funding streams in support of long-term preservation efforts. To address this concern, the UC Curation Center (UC3) at the California Digital Library (CDL) has developed a cost model for long-term preservation that can be applied on either a "pay-as-you-go" or a "paid-up" (or "pay once, preserve forever") basis. The latter is particularly appropriate for data produced as research outputs of grant-funded projects, since, in the absence of a paid-up option, the status of project outputs often becomes problematic when project funding ceases.

Cost modeling for preservation

For tractability, the UC3 analysis abstracts the complexity of the preservation environment and activities into the 10 high-level cost components shown in the table below. While these are sufficient to model the full economic cost of preservation, they can be customized easily to cover only specific subsets of those costs as determined by local policy

Preservation activities are embodied in an archival:
1. ***System***, composed of various
2. ***Services*** supporting necessary and desirable functions, running on
3. ***Servers***, designed, deployed, maintained, enhanced, and used by
4. ***Staff***, in support of content
5. ***Producers***, who use
6. ***Workflows*** to submit instances of
7. ***Content types***, which occupy
8. ***Storage***, and are subject to ongoing
9. ***Monitoring*** and periodic
10. ***Interventions***.

The operation of the *Archival* system (A), which subsumes the costs of *Services* and *Servers,* is considered a fixed cost; all other components are marginal costs each represented by a nominal unit cost and number of units consumed. Support for the *Archive*, *Workflows* (W), *Content Types* (C), *Monitoring* (M), and *Interventions* (I) are considered "common goods." Since they are beneficial to all *Producers* (P) their costs are properly apportioned across all *Producers. Storage* costs (S), on the other hand, are specific to a given *Producer*. The cost of supporting a new *Producer* is assessed only in the initial year. *Staff* costs are subsumed under the other model components. The annual pay-as-you-go price Q therefore can be represented as:

$$Q = \frac{A + m \cdot W + \ell \cdot C + j \cdot M + i \cdot V}{n} + k \cdot S + \begin{cases} P & if \ t = 0 \\ 0 & if \ t > 0 \end{cases}$$

Using this as a basis it is possible to define three cost functions for estimating total costs over time.

$H(T)$ The cumulative pay-as-you-go cost Q over T years.

$G(T, d)$ The discounted pay-as-you-cost, which assumes an annual decrease d in the cost Q of providing preservation service.

$F(T,d,r)$ The paid-up price, which assumes an up-front payment that earns an investment return r on the annual unspent surplus.

The relationship between these cost functions is shown in the graph below. Note that the functions G and F both approach limits as $T \rightarrow \infty$.
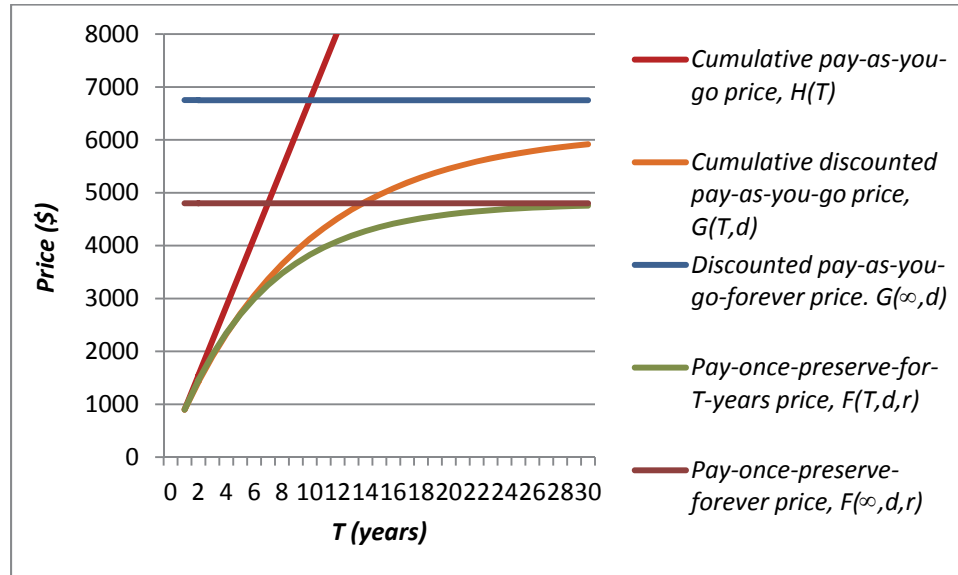
More information about UC3 cost modeling efforts, including a parameterized spreadsheet implementing the models, can be found at http://wiki.ucop.edu/display/Curation/Cost+Modeling.

The UC Curation Center (UC3) is a creative partnership bringing together the expertise and resources of the California Digital Library (CDL), the ten UC campuses, and the international curation community. Together, the UC3 partnership provides innovative curation solutions to its campus constituencies and external content partners.