

# EL CÁLCULO DE COSTES DE LA PRESERVACIÓN DIGITAL: UN ANÁLISIS DE MODELOS

*José Ramón Cruz Mundet\**

Universidad Carlos III de Madrid

*Carmen Díez Carrera\*\**

Universidad Carlos III de Madrid

**Resumen:** Conocer, medir y parametrizar cuánto cuesta la preservación digital permanente ha sido objeto de atención desde los primeros estudios a mediados de los años 90 del siglo XX.

El objeto es resultado de una investigación más amplia sobre los costes de la preservación digital, un tema al que se viene prestando un interés menor frente al más atractivo de los desarrollos tecnológicos y la materialización de grandes repositorios digitales. Sin embargo, se trata de una actividad que comporta unos costes de estructura compleja y de difícil concreción.

Tratar de desentrañar esta cuestión viene siendo objeto de diferentes proyectos en los últimos veinte años y, aunque se han producido algunos avances, nos encontramos bastante lejos de alcanzar un modelo de general aceptación.

Partiendo de esta constatación, nos proponemos analizar los nueve modelos empíricos desarrollados hasta el momento, y hacerlo desde una doble perspectiva: su caracterización y su comparativa, de cara a obtener como conclusión una imagen lo más precisa posible del estado en el que se halla la cuestión en el momento presente.

**Palabras clave:** Activos digitales; costes de la preservación digital; economía digital; mercado digital; OAIS; preservación digital.

**Title:** COSTS CALCULATION OF DIGITAL PRESERVATION: A MODELS ANALYSIS.

**Abstract:** Understand, measure and parameterize how much costs the long term digital preservation, has received attention from the first studies in the mid-90s of XX century.

How to unravel this question has been the subject of several projects in the last twenty years and, although there has been some progress, we are quite far from reaching a generally accepted model.

With this in mind, we analyze the nine empirical models developed so far, and we have try to do so from a dual perspective: characterization and comparison, in order to obtain as finding a picture as clear as possible of the state of the art in the question at the present moment.

**Keywords:** Digital assets; costs of digital preservation; digital economics; digital market; digital preservation; OAIS.

## 1. INTRODUCCIÓN

El cálculo de los costes es uno de los puntos débiles más importantes de la gestión de la preservación digital. Si bien se trata de una actividad reciente en el tiempo, aún se encuentra lejos de haber alcanzado un consenso y un compromiso general sobre su necesidad, más allá de la comunidad implicada en su práctica.

Tanto en el sector público, como en el privado, se coincide en que hace falta conservar datos y documentos necesarios para el funcionamiento diario, para documentar y probar actos, para cubrir las necesidades propias y de terceros. Son requisitos de las organizaciones digitales.

Existe la creencia bastante extendida de que los costes son de naturaleza tecnológica, y que se centran por tanto en el hardware, cada vez más barato, y en el software, que se supone incorporado en los sistemas de gestión. Un planteamiento erróneo e inviable para la preservación de información auténtica, fiable, íntegra y accesible a largo plazo.

Todavía estamos bastante lejos de poder contestar a la sencilla pregunta de ¿y esto, cuánto vale? Por eso mismo es difícil conseguir fondos para que las organizaciones implicadas y las instituciones de memoria conserven el patrimonio digital. Por esa misma falta de certeza resulta complicado convencer a quienes toman las decisiones, sobre todo en las organizaciones sin obligaciones sociales sobre ello, acostumbradas a conocer y controlar los costes de cada una de sus actividades.

\* joseramon.cruz@uc3m.es

\*\* carmen.diez@uc3m.es

Este tema ha sido objeto de atención desde muy temprano, ya a mediados de los 90 del pasado siglo, en los primeros estudios sobre conservación digital, se puso sobre la mesa, y veinte años después continúa sin resolverse. Este es en definitiva el campo de estudio del presente trabajo, que se propone analizar los diferentes modelos para el cálculo de los costes de la preservación digital y compararlos, para obtener una imagen lo más precisa posible del estado de la cuestión.

Se ha procurado una selección exhaustiva de todos los esfuerzos que cumplieran con los siguientes condicionantes:

1. Que fueran fruto del trabajo de expertos en las disciplinas implicadas: la economía, la tecnología y la gestión de la información, al menos.
2. Que se tratara de proyectos de investigación, bien financiados en convocatorias competitivas, bien por parte de entidades interesadas en los resultados.
3. Que los resultados hayan sido publicados y demostrado su aplicabilidad, con independencia de su continuidad.

Sobre esta base, hemos procedido a realizar una completa revisión de la literatura y a recoger los que cumplieran con el condicionado expuesto, de cuyo resultado hemos identificado nueve proyectos. Al tratarse de un asunto de enjundia, nos hemos encontrado con dos tipos de resultados: estudios teóricos y propuestas, que no han trascendido de su publicación; y proyectos, que en todos los casos han resultado cumplir con las condiciones para su selección.

La mayoría de los modelos ha surgido como resultado de proyectos de investigación, sus páginas web son la fuente de información principal, junto con las publicaciones derivadas de su desarrollo y difusión. En concreto nos hemos basado en los siguientes, ordenados según la secuencia en que han ido surgiendo:

1. Digital Preservation Testbed, de los Archivos Nacionales de Holanda, posteriormente unido al proyecto PLANETS.
2. CET (Costs Estimation Toolkit) de la NASA.
3. LIFE (Lifecycle Information for E-Literature), de la British Library y el University College London.
4. El modelo del proyecto ESPIDA.
5. El modelo KRDS (Keeping Research Data Safe) aplicado a los datos científicos.
6. Cost Model of Digital Archiving, del instituto Data Archiving and Networked Services (DANS) de la Real Academia Holandesa de Artes y Ciencias (KNAW).
7. CMDP (Cost Model for Digital Preservation) de los Danish National Archives (DNA).
8. TCP (Total Costs of Preservation Analysis) del Preservation Center de la Universidad de California.
9. 4C Project (Colaboration to Clarify the Costs of Curation) financiado por la UE.

Los hemos elegido porque son, hasta el momento, los únicos intentos de establecer modelos de validez general, unas veces, y orientados, otras, a satisfacer necesidades locales y concretas.

## 2. ORÍGENES Y PRIMEROS ENSAYOS

A mediados de los años 90 del siglo XX se produjeron los primeros ensayos, centrándose en unos pocos factores de costes: almacenamiento, metadatos e infraestructura técnica, o en cuestiones amplias como el aumento de los costes de la comunicación académica o en la economía de las bibliotecas digitales (Eakin y otros, 2008).

Los primeros estudios centrados en el tema fueron dos y coincidieron en el año de publicación:

- El de Beagrie y Greenstein financiado por las JISC, A Strategic Framework for Creating and Preserving Digital Resources: A JISC/NPO Study Within the Electronic Libraries (elib) Programme on the Preservation of Electronic Materials (1998)
- y el de Hendley, con la misma financiación, Comparison of Methods and Costs of Digital Preservation (1998).

El primero contiene referencias a factores de costes, mientras que el segundo se refiere a los costes ligados a las distintas técnicas de preservación.

Estos primeros ensayos no proporcionaron cálculos de costes explícitos, sino que comenzaron a evaluar las fases funcionales del ciclo de vida de la preservación (comenzando con la creación de los datos) y a determinar los factores de costes particulares asociados a cada etapa en el ciclo de vida. Hendley por ejemplo señalaba siete áreas a evaluar: la creación de datos, la selección y evaluación, la gestión, la descripción, el uso, la preservación y la gestión de derechos.

Poco después, Russell y Weinberger (2000) establecieron la diferencia entre costes de preservación y costes de acceso, en tanto procesos tecnológicos y de gestión diferentes con distintos costes asociados; aunque vinculados por el

hecho de que el acceso al material, incluso en condiciones de alta restricción, confiere valor a la información almacenada.

Ashley (2000) subrayó los costes y la ayuda que su concreción representaba para los archiveros, de cara a elaborar sus propios cálculos; algo parecido hizo Sanet (2002), mientras que Kenney y Rieger (2000) trataron de los costes del reformateado, Crespo y García Molina (2001) sobre el diseño del repositorio de archivo, Granger sobre los costes de emulación frente a los de migración. Sin embargo, ninguno de los autores presenta estimaciones de costes concretas y detalladas, la mayoría se centra en comparar métodos técnicos como emulación frente a migración, pero no los costes específicos atribuibles (Dollar, 1999; Holdsworth, 2001; Lorie, 2001; Rothenberg, 1999; Rothenberg y Bikson, 1999; Wheatley, 2001).

Al mismo tiempo estos primeros estudios reflejaron la creciente preocupación sobre la complejidad de analizar el coste en sus componentes. También comenzaron a mostrar conciencia sobre las relaciones entre las tecnologías y la amplia gama de actividades necesarias para alcanzar la preservación permanente (Eakin y otros, 2008). Los esfuerzos pusieron el foco en: (a) el ciclo de vida, (b) los costes de partida y (c) los costes en el tiempo. Así, (a) algunos proyectos desvelaron que es posible distribuirlos de forma desigual a lo largo del ciclo de vida de los objetos digitales. Beagrie y su equipo (b) demostraron que los de ingreso comportan una pesada carga, pero que si se hace bien los costes de preservación pueden ser despreciables a largo plazo (c); en caso contrario, los gastos de reparación y recuperación pueden resultar excesivos.

Shelby Sanet (2002 y 2003) estudió y comparó el modelo de Beagrie (1998) y el estudio de InterPARES en un contexto de preservación permanente de documentos auténticos, y concluyó señalando la falta de datos, de modelos y de cálculos de costes. Los esquemas existentes para jerarquizar los factores de costes eran poco realistas o poco aplicables.

Sanet propuso un modelo funcional basado en 3 categorías:

1. Costes de preservar documentos electrónicos: que incluye costes de capital, costes operativos directos e indirectos.
2. Costes por el uso de documentos electrónicos preservados.
3. Grupos de usuarios: que ofrece información sobre el acceso y los usuarios.

A las dos primeras categorías agregó el modelo desarrollado por InterPARES1 (Preservation Task Force), para dar lugar al suyo propio, una metodología orientada a la actividad, que permitía comparar categorías similares de costes en modelos y marcos, y en el contexto de cualquier estrategia particular de preservación.

La sostenibilidad de la preservación digital depende, entre otras razones, de conocer los costes. “Si las instituciones tienen una idea realista de los costes, pueden planificar en consecuencia. Un modelo de costes hace posible la planificación inteligente” (Sanet, 2003).

### 3. MODELOS DE COSTES

Al tiempo que Sanet publicaba sus hallazgos, arrancaron los primeros proyectos orientados -en cierta medida al menos- al desarrollo de modelos de costes. Debemos advertir que en este epígrafe resumiremos al máximo posible la descripción de los modelos, ya que de otro modo sería un trabajo imposible de presentar bajo este formato, y nuestro objetivo es la comparación; además de que se pueden consultar en sus publicaciones fundamentales reseñadas.

#### 3.1 Digital Preservation Testbed (Archivos Nacionales de Holanda)

Concluido en 2005 y posteriormente unido al proyecto Planets<sup>1</sup> (2007-2010), fue un proyecto de investigación acerca de las opciones de acceso seguro y continuado a documentos auténticos, sobre la base de evaluar la efectividad, las limitaciones, potencial aplicación y costes de tres enfoques de preservación digital permanente: migración, XML y migración según el modelo UVC (Universal Virtual Computer).

Entre otros resultados, el proyecto produjo unos indicadores de costes en forma de hoja de cálculo, para que se fueran actualizando a medida que se incorporaran nuevas experiencias y se identificaran nuevos elementos de juicio.

Diferenciaba los siguientes indicadores:

1. Los costes de un sistema de archivo digital (repositorio o depósito digital) y una funcionalidad para la preservación permanente de los documentos digitales (sistema de preservación), donde se detectan diversos componentes: espacio físico, hardware y software.

2. Costes de personal contemplados en dos momentos: el inicial (uno o dos años) que comporta la fase de diseño y puesta en marcha; y el de desarrollo, que depende del crecimiento, del grado de automatización y de la sostenibilidad económica.
3. Costes de desarrollo (o adquisición) de software y métodos para la preservación de documentos digitales, como determinar los requisitos de autenticidad de cada lote de documentos, establecer las estrategias de preservación, evaluarlas y documentarlas.
4. Costes del funcionamiento real de las acciones de preservación, como la migración, la conversión en XML, o el uso del UVC para mantener la accesibilidad.
5. Otros factores que influyen en el coste total como: los niveles de uso de los clientes, el tiempo transcurrido entre acciones de preservación, el seguimiento y la evaluación de la tecnología, las necesidades de almacenamiento adicional, los enlaces a los sistemas de gestión de documentos electrónicos, el volumen de documentos, la variedad de formatos y la velocidad de acceso, las exigencias de autenticidad y fiabilidad y la tasa de reposición de los sistemas.

(Slats, Verdegem, 2005).

El modelo se materializaba en una hoja de cálculo con tres partes:

1. Base de coste (los de personal, sistema de archivo digital y sistema de preservación)
2. Cálculo del tiempo (el que lleva ejecutar las actividades de archivado, desarrollar los métodos de preservación y ejecutar las actividades de preservación)
3. Cálculo de coste (por lote de documentos, por año y sobre un periodo de N años)

Se trataba, en todo caso, de un modelo inicial y estimativo puesto en marcha para animar a los futuros proyectos a ponerlo a prueba y a mejorarlo. Siendo un modelo sencillo y fácil de usar, se encuentra desactualizado y, probablemente, fuera de uso.

### 3.2 Cost Estimation Toolkit (CET)

Es una herramienta de la NASA para estimar el cálculo general de costes, que también se utiliza para el de los datos científicos, pero no está diseñado ni soporta el de la preservación digital. La literatura sobre la materia lo cita e incluye entre los modelos al efecto, sin embargo no contempla la planificación de la preservación, ni la migración (Fontain y otros, 2007). Ha sido diseñada para estimar el ciclo de vida completo de los sistemas de datos científicos procedentes de las operaciones terrestres y del espacio, lo que comprende: su introducción, uso y conservación; aunque esta no se plantee a largo plazo, uno de los inconvenientes más señalados por los expertos (Xue y otros, 2011).

### 3.3 El proyecto LIFE (Information For E-Literature)

Resultado de la colaboración entre el University College London (UCL) y la British Library, para elaborar una metodología de modelado del ciclo de vida digital y para calcular los costes de su preservación, ha sido desarrollado en tres fases entre 2005 y 2010: LIFE1, LIFE2 y LIFE3 (Holea y otros, 2010).

Ha sido una gran contribución a la comprensión de los costes a largo plazo de la preservación digital, como primer paso para ayudar a las instituciones a planificar el futuro de sus colecciones digitales. Partió de una doble base:

- Una revisión de la literatura desde 1978 sobre ciclos de vida de costes, especialmente enfocado a las bibliotecas (LIFE, 2005). El proyecto desarrolló una metodología basada en el ciclo de vida para calcular los costes de preservar información a 5, 10 y más años (McLeod, Wheatley, Ayrís, 2006).
- El trabajo de Oltmans y Kol (2005) para construir un modelo más detallado con la creación del Generic Preservation Model (GPM). El GPM permitía estimar los costes de preservación del contenido de una forma básica, que se mejoraría en la siguiente fase (LIFE2).

La falta de trabajos anteriores, también motivó el desarrollo del Modelo Genérico de Preservación, para un mayor desarrollo de la etapa de preservación. Ello permitiría a las instituciones empezar a identificar y reducir los picos de costes, así como la frecuencia de las acciones de preservación. En dicho modelo se identificaron y modelaron los factores de costes y, como parte del mismo, se elaboró una hoja de cálculo como herramienta de medición.

El modelo LIFE se basa en el ciclo de vida como concepto clave (Davis, 2008) y se compone de una serie de etapas:

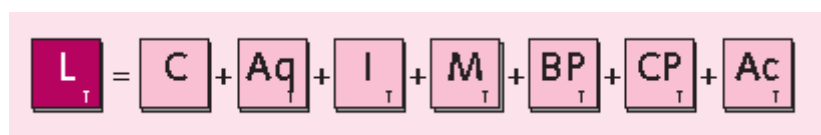


Figura 1. Etapas del Modelo LIFE3. (Fuente: The Life3, 2010).

L: costes del ciclo de vida/ C: creación o compra/ Aq: adquisición/ I: Ingreso/ M: creación de metadatos/ BP: preservación de flujo de bits/ CP: preservación del contenido/ Ac: acceso - T: tiempo.

Proporciona una visión de los procesos aplicados a los objetos digitales durante su ciclo de vida, por parte de una organización que actúa como responsable de su custodia, basándose en el modelo de referencia OAIS. Los procesos se organizan libremente en un orden cronológico desde su creación hasta su eventual acceso. Aunque en realidad se pueden agrupar y ejecutar en diferente orden. Lo que el modelo pretende es recoger los procesos observados en la mayoría de los ciclos de vida digitales. Si bien algunos pueden no ser aplicables a todos los ciclos de vida, la intención es proporcionar marcadores comprensibles para la mayoría de los procesos típicos del ciclo de vida.

El resultado final se materializa en una herramienta predictiva de costes que mejora significativamente la capacidad de las organizaciones para planificar y gestionar la preservación de contenido digital (Wheatley y Hole, 2009). Dicha herramienta, presentada como una hoja de cálculo Excel<sup>2</sup>, cubre una amplia gama de supuestos de conservación, incluyendo sonido, web y archivo de revistas electrónicas, además de impresos. Además, en asociación con el Humanities Advances Technology and Information Institute de la Universidad de Glasgow, produjo una versión web de la herramienta.<sup>3</sup>

La herramienta exige que el usuario introduzca información en sólo cinco campos. Esta información se utiliza para rellenar previamente el modelo con datos promedio de los estudios de caso relevantes en los que está disponible, y ofrece como resultado una estimación de costes.

Los cinco tipos de información son: años (inicio/finalización), categoría (tipo de activos), fuente (de costes), número de objetos por año y tamaño de la organización.

La herramienta es compatible con los siguientes tipos de activos: sitios web, revistas electrónicas, artículos impresos (digitalizados), grabaciones sonoras (digitales o analógicas) y documentos de investigación (MS Office, archivos PDF y pequeñas bases de datos).

El coste total del ciclo de vida estimado se calcula para cada etapa y elemento.

### **3.4 El modelo desarrollado por el Proyecto ESPIDA<sup>4</sup>**

Es un conjunto de recomendaciones estratégicas que puede ayudar a elaborar proyectos que no precisen ofrecer beneficios económicos de manera inmediata a una organización, sino beneficios intangibles. Ofrece ventajas tanto a los que proponen los proyectos, como a los que toman las decisiones, en el sentido de entender las propuestas, y asegurar que el proceso decisional es transparente y cuenta con toda la información relevante. Y permite a ambas partes una comunicación eficiente y en beneficio de la organización.

El trabajo se concentró en los modelos de costes, no en los proyectos, partiendo de la base de que los activos deben ser preservados. Su filosofía responde a varias preguntas: cuánto cuesta, por qué necesitamos hacerlo, y por qué debemos gastar dinero en ello, en lugar de gastarlo en la actividad central de la organización. Y mientras la precisión en los costes es vital para un buen proyecto, estas preguntas exigen respuestas muy diferentes (Curral, MacKinney, 2007).

El pensamiento estratégico se guía, además de por los temas financieros y de costes, por la visión y la percepción de las organizaciones que asumen riesgos cuando invierten en desarrollar nuevas ideas. El proyecto ESPIDA pretende asegurar que, cuando es necesario, la organización reconoce el valor de sus activos de información y tiene la perspectiva para constatar que su pervivencia es materia de decisiones, antes que de tecnología. Y que dichos activos están alineados con el plan estratégico de la organización. El reto consiste en expresarlo de una forma comprensible para los directivos. Además de que las decisiones de financiación se proyectan a corto plazo (de uno a tres años), mientras que las perspectivas de la preservación son a largo plazo, lo que hace necesaria la confluencia de ambas visiones (Curral, MacKinney, 2007).



universidades y otras instituciones de educación superior del Reino Unido, así como recomendaciones al respecto. El equipo de investigación fue dirigido por Neil Beagrie, Julia Chruszcz y Brian Lavoie, liderado por las universidades de Cambridge, Southampton y Kings College. Emplearon como punto de partida dos modelos de costes de actividad: LIFE y CET (*Cost Estimation Tool*) de la NASA, sobre la base del modelo OAIS.

Finalmente el estudio vio la importancia de alinear su modelo con otros existentes en las instituciones de educación superior y para ello también estudió el modelo TRAC<sup>6</sup> (Transparent Approach to Costing) para evaluar al completo los costes económicos de la universidades del país, ya que se trata de la metodología empleada en dichas instituciones (Beagrie y otros, 2008). El estudio concluía con un decálogo de recomendaciones acerca del valor de la conservación de los datos de la investigación y de las estrategias para lograrlo (Beagrie y otros, 2008).

KRDS es un modelo genérico, en el que se reutilizan términos y definiciones de OAIS, y que cada cual debe adaptar:

1. Fase de prearchivo: en la que se crean los datos e implica las actividades de difusión, iniciación y creación.
2. Fase de archivo: las actividades requeridas para la preservación digital permanente. Actividades: adquisición, disposición, ingreso, instalación, planificación de la preservación, inversión en innovación (I+D), gestión de datos, y acceso.
3. Servicios de soporte: servicios y funciones necesarios para controlar de forma continuada el funcionamiento de otras entidades funcionales. Actividades: administración, servicios comunes, gestión de inmuebles y servicios.

El análisis de los costes de la preservación de los conjuntos de datos de investigación no es suficiente para evaluar la viabilidad económica, debe ir acompañado de un cuadro de beneficios. Gran parte de la literatura sobre la materia se centra solo en el lado de los costes, comparativamente se presta poca atención a la articulación de los beneficios; en cambio, a menudo se dan por sabidos o se expresan en términos demasiado genéricos para ser de utilidad práctica para la toma de decisiones. Desafortunadamente, la medición de los beneficios es con frecuencia muy difícil, sobre todo cuando no se prestan fácilmente a su expresión en términos cuantitativos (KRDS, 53).

Como resultado se obtuvo la constatación de que los costes de las actividades de archivo (instalación y preservación) son una porción muy pequeña (en torno al 15%), frente al ingreso (55%) y el acceso (31%). La automatización en el ingreso mejorará los costes y la rentabilidad en el futuro. Y los costes fijos, especialmente el personal, tienen un alto peso (Beagrie y otros, 2010).

### 3.6 Cost Model for Digital Archiving

Basado en el modelo ABC (Activity-Based Costing Model) es un modelo para el cálculo de costes para la preservación y difusión de datos de investigación que ha sido desarrollado en el entorno del instituto Data Archiving and Networked Services<sup>7</sup> (DANS) de la Real Academia Holandesa de Artes y Ciencias (KNAW), un repositorio de confianza.

Parte de la constatación de que hay más información que recursos para la preservación digital permanente, y de que el cálculo de costes sigue siendo una actividad reciente y escasa.

El modelo DANS ABC se organiza en base de cinco componentes básicos:

1. Los grupos de recursos
2. Los factores de coste de los recursos
3. Las actividades (agrupadas)
4. Los factores de coste de las actividades
5. Los objetos de costes  
(Palaiologk y otros, 2012; Ooi, Soh, 2003).

El modelo tiene también sus limitaciones en relación con su transferibilidad y extensión. Mientras que el enfoque que se presenta en el trabajo para la construcción del modelo, se puede aplicar fácilmente a otras organizaciones, el modelo en sí puede requerir una personalización significativa, antes de que pueda ser utilizado en otros lugares (Palailogk y otros, 2012).

### 3.7 Cost Model for Digital Preservation (CMDP) del Archivo Nacional de Dinamarca<sup>8</sup>

Parte de la idea de que es necesario superar las limitaciones de los dos modelos anteriores: LIFE y KRDS. El proyecto desarrolla un modelo de costes basado en la actividad, que contabiliza todos los costes económicos, directos e indirectos, y estructurado en torno al desglose funcional proporcionado por el modelo de referencia OAIS (Ingreso, Instalación, Gestión de datos, Administración, Planificación de la preservación, Acceso, y Servicios comunes), así

como los roles de Productor, Consumidor y Gestión, como marcadores de posición para factores de coste externos, que influyen en el coste de la preservación. Cada una de estas entidades comprende una serie de funciones descritas en OAIS. Según su experiencia, en relación con la migración digital, OAIS proporciona un nivel de detalle que equivale o excede el de cualquier otro modelo funcional utilizado para el cálculo de costes.

Incluye una herramienta que calcula los costes presentes y futuros de las colecciones digitales de las instituciones del patrimonio cultural, basada en diversas aportaciones de usuarios, tales como la cantidad y el tipo de datos.

El modelo se caracteriza porque ha sido diseñado para proporcionar un enfoque coherente, para estimar los costes económicos totales de la preservación de materiales digitales en instituciones normalmente eficientes y que cumplen con OAIS. Asimismo, el enfoque basado en la actividad permite el seguimiento de los costes en el tiempo.

El modelo se ha hecho para la entidad funcional Planificación de la preservación, aplicado a la migración y basado en datos de costes limitados a la experiencia en los archivos. Si bien el objetivo es modelar todo el ciclo de vida de la preservación digital, la primera versión del modelo sólo se ocupa del coste de la planificación y de las migraciones digitales; donde la complejidad de los formatos y de los objetos digitales tiene una gran influencia en estos costes

Los autores del proyecto concluyen reflexionando sobre la inadecuación de los modelos de costes y la importancia de desarrollar métodos más precisos, aunque destacan el alto grado del suyo; para lo que hace falta poner a prueba constantemente los modelos empíricos. Sin embargo, constatan, faltan datos de pruebas, así como estudios teóricos, por ejemplo, sobre frecuencia de migración y expectativa de vida de los formatos, factores que mantienen un alto grado de incertidumbre y contaminan de imprecisión el modelo (Kejser, Nielsen y Thirifays, 2011).

Aun y todo, reconocen los autores, las predicciones de costes de preservación son inciertas, más allá de unos pocos años y esta incertidumbre aumenta con el tiempo debido a la dificultad de predecir los costes futuros de la tecnología. Si bien la instalación de archivo representa el área menos compleja y más madura del ciclo de vida de la preservación digital, además de la que ofrece mayor cantidad de datos de costes empíricos (Kejser, Nielsen y Thirifays, 2012).

Falta por desarrollar un modelo común y datos reales para comparar, así como evaluación sobre la calidad de los activos y lo que ello significa para los interesados.

### 3.8 Total Cost of Preservation (TCP) Analysis

El *Curation Center* de la Universidad de California elaboró el Total Cost of Preservation (TCP) Analysis<sup>9</sup>, que puede ser aplicado en el desarrollo de dos modelos específicos: el modelo de pago sobre la marcha, basado en una facturación anual (Pay-as-you-go) y el modelo de pago por desembolso, abonado al comienzo del periodo en cuestión (Paid-up). El primero es adecuado para los servicios de preservación de pago, donde existe un flujo de ingresos anuales fiable y predecible; mientras que el segundo está pensado para organizaciones con financiación puntual y proyectos de investigación con plazo determinado.

El modelo parte de la base de algunos supuestos, entre los cuales destacan:

- Sólo se consideran los costes pertenecientes a los proveedores del servicio de preservación, los asociados con el Archivo en terminología OAIS, mientras que los del productor quedan fuera.
- Los componentes de costes pueden categorizarse como:
  - fijos, en los que se incurre con independencia del nivel de uso, y como marginales, establecidos en proporción a su uso;
  - como de una sola vez y como recurrentes.
- Las acciones ejecutadas en la preservación de contenido están esencialmente automatizadas.
- Se conoce el tamaño y el alcance del contenido y el coste de preservarlo se ha establecido al comienzo.

El análisis TCP abarca los costes económicos asociados a la conservación permanente de los activos digitales, aunque los modelos de precios resultantes se pueden personalizar fácilmente para su uso en otros casos.

### 3.9 4C Collaboration to Clarify Costs of Curation<sup>10</sup>

Es un proyecto financiado por el 7º Programa Marco de la UE que se ha desarrollado entre 2013 y 2015. Entre sus objetivos está analizar los modelos existentes, elaborar nuevas propuestas y favorecer la conciencia y el intercambio de información y de experiencias para el desarrollo de un modelo y de una herramienta robusta para el cálculo de costes.



Su visión es que “en cinco años (2020) será más fácil diseñar o disponer de servicios de conservación digital más rentables y eficientes, porque los costes, los beneficios y los ejemplos para hacerlo serán más y mejor comprendidos por todos los interesados relevantes en el ciclo de vida. El modelado de costes formará parte de la planificación y actividades de gestión de todos los repositorios digitales” (4C, 2015).

El aspecto en el que más incide es en el planteamiento de un borrador de nuevo modelo: Economic Sustainability Reference Model (ESRM), tributario del elaborado por Rusbridge y Lavoie en 2011, que se basa a su vez en el informe del Blue Ribbon Task Force for Sustainable Digital Preservation and Access. Modelo que por el momento continúa en un nivel teórico, a falta de su concreción, y en él Rusbridge y Lavoie utilizan deliberadamente el término conservación digital, en lugar de preservación digital, porque comprende “mantener, preservar y agregar valor a la información digital a lo largo de su ciclo de vida” (Rusbridge, Lavoie, 2011). Un término que el proyecto 4C mantendrá en todo momento como preferente.

La definición constituye un buen ejemplo de lo ya apuntado, el modelo es “una herramienta estratégica para planificar y discutir, dirigida al personal ejecutivo y de administración, más que al operativo. Su propósito es proporcionar los cimientos para progresar en el desarrollo de estrategias sostenibles de éxito para la conservación digital. Ello lo hace organizando la materia, proporcionando un punto de referencia común de conceptos y términos, e introduciendo una capa de abstracción que esconde las complejidades e idiosincrasias de implementaciones y de contextos individuales, mientras que, al mismo tiempo, incorpora detalle suficiente para sostener discusiones sustantivas de temas compartidos” (4C, 2013).

El modelo de costes y beneficios que plantea es un modelo conceptual con los siguientes componentes:

1. Objetivos y estrategias: describe las metas de una organización (consumidor) en términos de conservación de los activos digitales de los que es responsable y cómo alcanzar dichas metas. El consumidor define las necesidades del servicio de conservación basándose en los objetivos y estrategias, y evalúa los costes y beneficios del servicio.
2. Contexto organizacional: que define los objetivos y estrategias. Los consumidores toman decisiones a la luz de la naturaleza de las organizaciones y de los activos de información que poseen, en tanto interesados.
3. Riesgos: influyen en los objetivos y estrategias que deben articularse y gestionarse para minimizar las amenazas y maximizar las oportunidades.
4. Costes y beneficios: asociados a la confluencia con los objetivos de conservación de la organización, materializados en el servicio de conservación.

El modelo se ha concebido como una base común, que permita a las herramientas de evaluación de costes emplear elementos comparables, pues cuanto más representativos sean de situaciones concretas, tanto más adecuados serán para el cálculo, pero menos comparables (4C, 2014).

Es importante para una organización que proporciona servicios de conservación, comprender bien la distribución de los costes y cuáles son los más importantes, los que necesitarán una atención especial en su gestión.

Como resultado del proyecto 4C se ha creado el espacio Curation Costs Exchange<sup>11</sup>, una plataforma para compartir, comparar y comprender datos sobre costes de la preservación digital, con el objetivo de ayudar a que las decisiones estén mejor informadas y las inversiones cuenten con datos fiables. Además de información extractada del proyecto, el aspecto más novedoso y práctico es la herramienta para la parametrización y comparación de factores de costes individuales, una herramienta a la que se puede acceder tras darse de alta y permite introducir datos y obtener resultados comparativos.

#### 4. ANÁLISIS Y CONCLUSIONES

A partir del análisis de los modelos anteriores desde una óptica comparativa, cabe extraer algunas conclusiones, que en cierto modo se derivan de las carencias detectadas.

En primer lugar, partimos de una constatación sobre la que existe acuerdo unánime: nos encontramos en la infancia de la preservación digital, caracterizada además por desarrollarse en un entorno mixto o dual, en el que lo analógico sigue teniendo mucho peso. Una situación que probablemente se prolongue en el tiempo, si nos encaminamos hacia un futuro en el que convivan lo analógico y lo digital, horizonte que no parece descartable.

A pesar del poco tiempo transcurrido y del camino aún por recorrer, llama la atención el enorme esfuerzo realizado por tantos investigadores. Una parte de la comunidad implicada en la preservación digital, está en la senda de alcanzar

uno o más modelos, que sirvan para parametrizar los costes de esa actividad y, todo parece indicar, que se logrará en el plazo de unos años.

Entrando en detalle, los modelos estudiados revelan la conciencia generalizada acerca de que los costes, no se pueden aislar del medio en el que se producen; se insertan en un marco más amplio que incluye los intereses de terceros, la estructura organizacional y el entorno cultural. Los intereses se han ampliado para incluir los costes de preservar un objeto en todo su ciclo de vida, como han ejemplificado muy bien los modelos del proyecto LIFE y el de Beagrie, Chruszcz, y Lavoie (Eakin y otros, 2008).

El modelado de costes tiene un enfoque de ciclo de vida, aspecto en el que coinciden desde hace tiempo, cuando ya se enunciara en un temprano estudio sobre métodos de preservación y modelos de costes (Hendley, 1998). El motivo es la naturaleza recurrente de los costes de preservación, y el hecho de que son difíciles de separar de otros costes del ciclo de vida, como la creación y el acceso (Granger, Russell y Weinberger, 2000). Y que los costes de preservación dependen en gran medida del rango de servicios que ofrece una institución.

Las diferencias surgen cuando se discute cómo se debería estructurar o desglosar el ciclo de vida de los costes de la preservación digital, y cómo detallarlo. Estamos a falta de un modelo funcional bien definido, así como de un acuerdo sobre principios contables generalmente aceptados. “Otro tema sin resolver es el desarrollo de fórmulas para hacer modelos de costes operativos” (Kejser, Nielsen y Thirifays, 2011).

Como constatación de las diferencia está la existencia misma de varios modelos para distintos tipos de información, en los que se han identificado adecuadamente los factores de costes, pero no se han desarrollado modelos que sean aplicables a la realidad. Los resultados están mayoritariamente basados en situaciones individuales y no se pueden extrapolar ni son susceptibles de comparación, salvo excepciones, y aún estas se refieren a categorías de instituciones (sobre todo las de educación superior) o a tipologías de materiales (audiovisuales, textuales...). Cada uno de los modelos se ha establecido generalmente para dar respuesta a una necesidad concreta, por lo que sus objetivos son también diferentes (Zeller, 2010). Y esto mismo hace que sean difíciles de manejar, a pesar de la abundante documentación con que se acompañan, y resulten poco fiables por la carencia de prácticas de auditoría y de certificación (4C, 2014).

Hay un aspecto que trasciende el cálculo de costes y afecta a la preservación en su totalidad, y es que los objetivos y los incentivos no están a menudo bien alineados entre las comunidades de interesados: el poseedor de los derechos, el archivo o custodio tercero, y los beneficiarios. Un aspecto sobre el que pronto llamaron la atención Lavoie (2003) y Bradley (2005).

Queda por construir un modelo general que, más allá de una simple suma de los costes de los diversos elementos, pueda integrar los factores que influyen entre los diferentes elementos de costes. Y es que casi todos los modelos excluyen la propia parte archivística de la conservación, en la descripción de los costes (Zeller, 2010, 7). Se trata, a nuestro parecer, de uno de los puntos débiles a un tiempo más evidente y fácil de superar. Conviene mantener unidos la preservación con los procesos antecedentes: creación y gestión, pues existen factores como los metadatos que se imputan a estas fases (costes) y otros que dejan de computarse por quedar en un limbo, como sucede con la selección (beneficios).

Esto último nos permite recordar que todos los objetos no son activos digitales, sino tan solo aquellos que tienen un valor presente y futuro y han sido seleccionados. La valoración y la selección son procesos incuestionables para establecer esa diferencia y aunque hay experiencia en la materia, los procesos aplicados a los materiales digitales deben ser semiautomatizados y escalables para ser eficientes. Hace falta identificar el valor de los activos digitales y elegir, tomar decisiones (4C, 2015).

Hace falta que se publiquen datos, que las organizaciones los den a conocer y los compartan. Se trate de textos electrónicos formateados, datos científicos, documentos... es importante constituir una base de conocimiento para la comparativa y la modelización. Ahora mismo, el espacio Curation Costs Exchange, al que ya nos hemos referido, es una vía válida, aunque otros intentos anteriores han concluido prematuramente sin resultados. Todo depende de que se consolide con el tiempo una comunidad de interesados bien definida y de que se alineen sus expectativas e intereses.

Con todo, están pendientes, aspectos determinantes en la concienciación sobre el valor y la necesidad misma de preservar, cuya solución no parece fácil. ¿Cómo es posible casar un compromiso a largo plazo, como el de la conservación, con un horizonte de actividad a corto o medio plazo, como es el de gran parte de las organizaciones? En este sentido, como ya ocurre en el entorno analógico, las Administraciones Públicas poseen ventaja frente a las

empresas, organizaciones privadas, familias e individuos. La falta de atractivo económico es una rémora, pues más allá de las exigencias legales y de las necesidades de memoria que cada uno asuma, la preservación digital carece del incentivo que el patrimonio cultural analógico ofrece desde el punto de vista inversor, sea para la posesión particular (coleccionismo), sea para el disfrute general.

Queda camino por recorrer y resistencias por superar. Hace falta mejorar el mercado de las soluciones, con sistemas más eficientes; desarrollar servicios e infraestructuras escalables, a la medida de las diferentes necesidades. Colaborar compartiendo recursos, infraestructuras y esfuerzos (4C, 2015).

## NOTAS

- <sup>1</sup> <<http://www.planets-project.eu>> [Consulta: 8 de mayo de 2015].
- <sup>2</sup> <<http://www.life.ac.uk/tool/>> [Consulta: 8 de mayo de 2015].
- <sup>3</sup> <<http://lifedev.hati.arts.gla.ac.uk/>> [Consulta: 8 de mayo de 2015].
- <sup>4</sup> <<http://www.gla.ac.uk/services/library/espida/>> [Consulta: 8 de mayo de 2015].
- <sup>5</sup> <<http://www.beagrie.com/krrds/>> [Consulta: 8 de mayo de 2015].
- <sup>6</sup> <<http://www.jcpcg.ac.uk/guidance/>> [Consulta: 8 de mayo de 2015].
- <sup>7</sup> <<http://www.nwo.nl/en/about-nwo/organisation/nwo-divisions/dans>> [Consulta: 8 de mayo de 2015].
- <sup>8</sup> <<http://www.costmodelfordigitalpreservation.dk/>> [Consulta: 8 de mayo de 2015].
- <sup>9</sup> <<https://wiki.ucop.edu/display/Curation/Cost+Modeling>> [Consulta: 8 de mayo de 2015].
- <sup>10</sup> <<http://4cproject.eu/>> [Consulta: 8 de mayo de 2015].
- <sup>11</sup> <<http://www.curationexchange.org/>> [Consulta: 8 de mayo de 2015].

## 5. BIBLIOGRAFÍA

- 4C PROJECT. Draft Economic Sustainability Reference Model, [s. l.]. 2013.
- 4C PROJECT. Evaluation of Cost Models and Needs & Gaps Analysis. [s. l.]. 2014.
- 4C PROJECT. Roadmap. Investing in Curation. A Shared Path to Sustainability. [s. l.]. 2015.
- ASHLEY, K. Digital Archive Costs: Facts and Fallacies. En: Proceedings of the DLM Forum on Electronic Records (DLM '99), 1999, p. 121-128.
- BEAGRIE, N. y GREENSTEIN, D. *A Strategic Policy Framework for Creating and Preserving Digital Collections: A Report to the Digital Archiving Working Group*. Londres: British Library Research and Innovation Centre, 1998.
- BEAGRIE, N., CHRUSZCZ, J. y LAVOIE, B. *Keeping Research Data Safe: A Cost Model and Guidance for UK Universities*. Londres: Joint Information Systems Committee, 2008.
- BEAGRIE, N.; LAVOIE, B. y WOOLLARD, M. *Keeping Research Data Safe 2*. Final Report - abril 2010. Londres: Joint Information Systems Committee, 2010.
- BLUE RIBBON TASK FORCE. Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information. Final Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access. [s.l.]. 2010.
- BRADLEY, K. *Digital sustainability and Digital Repositories*. Sidney: National Library of Australia, 2015.
- CRESPO, A. y GARCÍA-MOLINA, H. Cost-Driven Design for Archival Repositories. En: Joint Conference on Digital Libraries 2001 (JCDL'01); 24-28 junio; Roanoke, Virginia, USA, 2001.
- CURRALL, J. y MCKINNEY, P. Investing in Value: A Perspective on Digital Preservation. En: *D-Lib Magazine*, 2006, vol. 12, nº 4.
- DAVIES, R. (ed.); AYRIS, P. y otros (LIFE Project Team). *The LIFE2 Final Project Report*. Londres: Joint Information Systems Committee, 2008.
- DOLLAR, C.M. *Authentic Electronic Records: Strategies for Long-term Access*. Chicago, IL: Cohasset Associates, 1999.
- EAKIN, L.; FRIEDLANDER, A. y SCHONFELD R. with contributions by CHOUDRHURY, S. A Selective Literature Review on Digital Preservation Sustainability. Report. [s. l.]. 2008.
- FONTAINE, K.; HUNOLT, G; BOOTH, A. y BANKS, M. Observations on Cost Modeling and Performance Measurement of Long-Term Archives. En: PV2007 Conference Proceedings, 2008, p. 7.
- GRANGER, S; RUSSELL, K. y WEINBERGER, E. Cost elements of digital preservation. En: *D-Lib Magazine*, 2000, vol. 6 nº 10.
- HENDLEY, T. *Comparison of Methods & Costs of Digital Preservation*. Londres: British Library Research and Innovation Report 106, 1998.
- HOLDSWORTH, D. Emulation, Preservation, and Abstraction. En: *RLG DigiNews*, 2001, vol. 5 nº 4.
- HOLEA, B. y otros. The Life3 Predictive Costing Tool for Digital Collections. En: *New Review of Information Networking*, 2009, vol. 15, nº 2, p. 81-93.

- KEJSER, U.B.; NIELSEN, A.B. y THIRIFAYS, A. Cost Model for Digital Preservation: Cost of Digital Migration. En: *The International Journal of Digital Curation*, 2011, vol. 1, nº 6, p. 255-267.
- KEJSER, U.B.; NIELSEN, A.B. y THIRIFAYS, A. Modelling the Costs of Preserving Digital Assets. En: *The Memory of the World in the Digital Age: Digitization and Preservation*, 26-28 septiembre. Vancouver: UNESCO, 2012, p. 529-539.
- KENNEY, A.R. y RIEGER, O.Y. *Moving Theory into Practice: Digital Imaging for Libraries and Archives*. Mountain View, CA: Research Libraries Group, 2000.
- LAVOIE, B. *The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making*. Londres: OCLC Office of Research, 2003.
- LIFE3 PROJECT, THE. Bringing Digital Preservation to LIFE Lifecycle Information for E-literature An Introduction to the third phase of the LIFE Project A JISC and RIN funded joint venture Project. Londres, 2010.
- LORIE, R.A. Long Term Preservation of Digital Information. En: *First ACM/IEEE-CS Joint Conference on Digital Libraries*, Roanoke, VA, 2001.
- MCLEOD, R.; WHEATLEY, P. y AYRIS, P. *Lifecycle information for e-literature: full report from the LIFE project*. Londres: LIFE Project, 2006.
- OLTMANS, E. y KOL, N. A Comparison Between Migration: An Overview of Life Cycle Management at the National Library of the Netherlands and Emulation in Terms of Costs. En: *RLG Diginews*, 2005, vol. 9, nº 2.
- OOI, G. y SOH, C. Developing an activity-based costing approach for system development and implementation. En: *The Data Base for Advanced Information Systems Summer*, 2003, vol. 34, nº 3, p. 54-70.
- PALAILOGK, A.S. y ECONOMIDES, A.A.; TJALSMA H.D. y SESINK, L.B. An Activity-Based Costing Model for Long-Term Preservation and Dissemination of Digital Research Data: the Case of DANS. En: *International Journal of Digital Libraries*, 2012, vol. 12, p. 195-214.
- ROTHENBERG, J. *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. Washington: Council on Library and Information Resources, 1999.
- ROTHENBERG, J. y BIKSON, T.K. Carrying Authentic, Understandable and Usable Digital Records Through Time [s. l.]. RAND Europe, 1999.
- RUSBRIDGE, C. y LAVOIE, B. Draft Economic Sustainability Reference Model. [s. l.]. 2011.
- RUSELL, K. y WEINBERGER, E. *Cost Elements of Digital Preservation*. Leeds: Cedars Project, 2000.
- SANETT, S. Toward Developing a Framework of Cost Elements for Preserving Authentic Electronic Records into Perpetuity. En: *College and Research Libraries*, 2002, vol. 63, nº 5, p. 388-404.
- SANETT, S. The Cost to Preserve Authentic Electronic Records in Perpetuity: Comparing Costs across Cost Models and Cost Frameworks. En: *RLG Diginews*, 2003, vol. 7, nº 4.
- SLATS, J. y VERDEGEM, R. Cost Model for Digital Preservation. National Archief of the Netherlands. En: *DLM Forum*. Budapest: DLM Forum, 2005, p. 8.
- WHEATLEY, P. Migration - a CAMiLEON Discussion Paper. En: *Ariadne*, 2001, vol. 29.
- WHEATLEY, P. y HOLE, B. LIFE3. Predicting Long Term Digital Preservation Costs. En: *iPRES 2009: the Sixth International Conference on Preservation of Digital Objects*. San Francisco: California Digital Library, UC Office of the President, 2009, p. 206-210.
- XUE, P.; BADAWY, M.; SHEHAB, E. y BAGULEY, P. Cost Modelling for Long-term Digital Preservation: Challenges and Issues. En: *Proceedings of the 9th International Conference on Manufacturing Research ICM*, 2011, p. 6.
- Zeller, J.D. Cost of Digital Archiving: Is there a Universal Model? En: *8th European Conference on Digital Archiving*, Geneva, 28 - 30 abril, 2010, p. 15.