

Editora

Luisa Angélica Paraguai Donati

Conflicto de intereses

Los autores declaran que no existe conflicto de intereses.

Recibido

15 feb. 2024

Versión Final

11 sept. 2024

Aprobado

10 oct. 2024

A la búsqueda de un modelo: 25 años calculando los costes de la preservación digital

In Search of a Model: 25 Years Calculating the Costs of Digital Preservation

Carmen Díez-Carrera¹ , José Ramón Cruz-Mundet¹ 

¹ Universidad Carlos III de Madrid. Facultad de Humanidades, Comunicación y Documentación. Madrid, España. Correspondencia para/Correspondence to: C. Díez-Carrera. E-mail: <carmen.diez@uc3m.es>.

Como citar este artículo/How to cite this article: Díez-Carrera, C.; Cruz-Mundet, J.R. A la búsqueda de un modelo: 25 años calculando los costes de la preservación digital. *Transinformação*, v. 36, e2411333, 2024. <https://doi.org/10.1590/2318-0889202436e2411333>

Resumo

El cálculo de los costes es uno de los aspectos de la preservación digital permanente que aún está a la espera de solución. Saber cuánto cuesta preservar un paquete de información de archivo a largo plazo, ha sido una aspiración que surge desde que la comunidad interesada a finales del siglo pasado, se empezó a plantear el binomio preservación permanente aplicado al ecosistema digital. El objetivo del trabajo es analizar los diferentes modelos para el cálculo de los costes desarrollados en los 25 años transcurridos entre 1998, fecha del primer planteamiento teórico y 2023. Para ello se realizó una revisión exhaustiva de la literatura, junto con un estudio de los sitios web de los proyectos. El resultado ha sido la obtención de un diagnóstico sobre la viabilidad de dichos modelos, las últimas propuestas aún en desarrollo, lo que nos permite poner de relieve los aspectos pendientes de resolver para lograr despejar el horizonte de la preservación digital permanente.

Palabras clave: Cálculo de costes. Modelización. Preservación digital.

Abstract

Cost calculation is one of the aspects of long-term digital preservation that is still awaiting resolution. Knowing how much it costs to preserve a package of archival information in the long term is an aspiration that has arisen since the interested community, at the end of the last century, began to consider the binomial permanent preservation applied to the digital ecosystem. The objective of the work is to analyze the different models for calculating the costs developed in the 25 years between 1998, date of the first theoretical approach and 2023. To do this, an exhaustive review of the literature has been carried out, together with the study of the project websites. The result has been obtaining a diagnosis on the viability of these models, the latest proposals still in development, which allows us to highlight the aspects pending resolution to clear the horizon of long-term digital preservation.

Keywords: Cost Calculation. Modeling. Digital Preservation.

Introducción

El presente trabajo tiene por objetivo conocer el estado del arte y el grado de madurez alcanzado en la modelización de los costes de la preservación. El cálculo de

los costos es uno de los aspectos de la preservación digital permanente que aún está a la espera de solución. Saber cuánto cuesta conservar un paquete de información de archivo a largo plazo, empleando la terminología Open Archival Information System (OAIS) (Sistemas de Transferencia de Datos e Información Espaciales, 2015), es una aspiración que surgió desde que la comunidad interesada a finales del siglo pasado, se empezó a plantear el binomio preservación permanente aplicado al ecosistema digital. La preocupación de los especialistas y las primeras voces de alarma surgieron al inicio de la revolución tecnológica. Cuando los usuarios de Internet apenas representaban el 2% de la población mundial y las conexiones eran vía modem, Terry Kuny, consultor tecnológico canadiense, acuñó la expresión Digital Dark Age (Edad Oscura Digital) en su alocución dirigida a la conferencia de la Federación Internacional de Asociaciones Bibliotecarias (IFLA) (Kuny, 1997). En ella sostenía que estábamos –y continuamos– en una era en la que se valoran más el cambio y la velocidad, que la conservación y la longevidad. En ese mismo momento empezaba a plantearse la necesidad de establecer un modelo para el cálculo de los costes, una preocupación que ha dado resultados, pero a la que le queda recorrido hasta alcanzar sus fines. Y así, pasados 25 años desde los primeros planteamientos (Beagrie; Greenstein, 1998) está lejos de resolverse.

Es paradójico y, hasta cierto punto contradictorio, que aún seamos incapaces de contestar a la sencilla pregunta de ¿cuánto cuesta? Así es difícil lograr que las instituciones de memoria obtengan los fondos suficientes para llevar a cabo la conservación del patrimonio digital, y que las organizaciones sin obligaciones sociales en la materia asuman la responsabilidad de preservar. La toma de decisiones sin un modelo de cálculo de costes es casi impensable.

El desarrollo de un modelo de costos es una tarea que requiere información, investigación rigurosa y una buena comprensión de los desafíos. En la preservación digital estos desafíos son de cuatro tipos: de carácter tecnológico, informativo, metodológico y de actividad (Badawy, 2017). Los desafíos tecnológicos derivan de la naturaleza de la propia tecnología, de su obsolescencia y de la incertidumbre inherente a los formatos de ficheros existentes (Erkoyuncu *et al.*, 2009; Romero Rojo, 2011); los costos de migración de la información (Russell, 2000) y los costos de las tecnologías utilizadas en nuevos proyectos, como la computación en la nube (Baker *et al.*, 2006; Rosenthal *et al.*, 2010). La información de costos es un gran desafío, ya que no está fácilmente disponible y es difícil de recabar. La metodología para la modelización de costos está insuficientemente elaborada y el desarrollo de la preservación como actividad, por qué no como negocio, es un desafío carente de un control preciso de los costes, lo que dificulta su lanzamiento como actividad empresarial.

Procedimientos Metodológicos

Para lograr el objetivo de este trabajo vamos a profundizar en la economía de la digitalización y de los costos asociados a la preservación digital, analizando los diferentes modelos que se han desarrollado en el tiempo. Asimismo, nos proponemos fijar la tasa de supervivencia de estos modelos, como una forma de medir su éxito.

Las fuentes de información serán de dos tipos: en primer lugar, la producción bibliográfica especializada en el tema, desde sus inicios en los años finales del siglo XX, hasta la actualidad a finales de 2023. En segundo lugar, los sitios web de los proyectos de investigación, de las empresas desarrolladoras y de las instituciones que han participado en los mismos. Partiremos de una exhaustiva revisión de la literatura, para determinar el estado actual del conocimiento, el estado del arte. A continuación, analizaremos uno a uno los modelos implementados para conocer su estado y su destino, así como las nuevas propuestas en fase de desarrollo.

La metodología se basa en la establecida en su estudio por Cruz Mundet y Díez Carrera (2015), quienes analizaron los modelos desarrollados entre 1995 y 2014, los cuales serán complementados y mejorados con nuevas fuentes, destacadas por la revisión de la literatura, para orientarnos al estudio de los cambios producidos hasta la actualidad. El propósito es descubrir y analizar qué ha sucedido en estos años: cuál ha sido el destino de esos modelos, cómo ha evolucionado el interés de la investigación por la materia, cuantificar, analizar y tipificar los nuevos modelos propuestos. Para concluir, pretendemos establecer la continuidad de los modelos supervivientes y las previsiones para el futuro.

La revisión de la literatura se basó en una búsqueda exhaustiva en los siguientes repositorios: Clarivate *Web of Science*, Scopus, ProQuest Central, Dialnet, Digital CSIC, *eScholarship*, JSTOR y *Academic Search Ultimate* (Ebsco) junto con *Library, Information Science & Technology Abstracts* (LISTA); siendo estas dos últimas en realidad coincidentes al punto que parecen ser la misma base de datos, tanto por la similitud de la interface, como por la similitud de los resultados. Primeramente, se estableció el ámbito temporal, comenzando en 1998, el año de la primera publicación conocida sobre la materia, la ya mencionada de Beagrie y Greenstein, para finalizar con 2023, de modo que comprendiera un arco temporal de 25 años, para dar perspectiva al estudio. En una primera fase se investigó cada repositorio mediante el uso de términos precisos y restringidos al título y a los descriptores de materias, con una selección de voces adecuadas al objeto: costes, preservación digital, curación/curaduría digital, modelos/*Costs*, *Digital Preservation*, *Digital Curation*, *Models*. El resultado fue la obtención de referencias fiables y adecuadas, pero escasas, por lo que se procedió a una segunda fase con búsquedas sencillas y avanzadas en combinación con otros términos relacionados: retorno de la inversión/*Return of Investment/ROI*, precios/*Pricing*, coste del riesgo/*Cost of Risk*. Como resultado se obtuvieron un total de 344 referencias bibliográficas, que fueron sometidas a un cribado para eliminar duplicados (criterio cuantitativo), así como los trabajos reiterativos y sin relación temática (criterio cualitativo), con lo que se obtuvieron 147 referencias; de las cuales pasaron un último cribado de lectura 68, basado en la correspondencia efectiva entre el título y las palabras clave y el contenido.

Una vez identificados los modelos y los proyectos que los desarrollaron, decidimos medir su durabilidad, su permanencia en el tiempo, para lo que hemos buscado en Internet sus páginas web y, cuando han desaparecido, se ha procedido a una nueva búsqueda a través de organismos nacionales e internacionales relacionados con los proyectos originales y sus sucesores o responsables de su conservación (*Joint Information Systems Committee*, *Digital Preservation Coalition*, *Internet Archive Wayback Machine*, *Software Heritage*, entre otros) en busca de información y de instantáneas (snapshots).

Resultados

El análisis de la literatura y de los proyectos que dieron origen a los diferentes modelos nos llevó a estructurar los resultados a partir de una perspectiva dual: cronológica y tipológica. Acompañaremos el desarrollo temporal de las diferentes soluciones y, al mismo tiempo, las vamos a agrupar de acuerdo con las tipologías que hemos podido descubrir como resultado de la investigación: modelos teóricos (centrados en el ciclo de vida y en los factores de costes), modelos bibliotecarios (centrados en la preservación bibliográfica y de datos de la investigación), y los modelos competitivos (con una perspectiva empresarial y de transferencia a la sociedad).

Ciclo de Vida y Factores de Costes

Las primeras contribuciones al tema fueron de naturaleza teórica y orientadas a intentar establecer el ciclo de vida de la preservación digital y a la caracterización de los factores de costes

asociados a cada fase de dicho ciclo, destacando el estudio de Beagrie y Greenstein (1998) y el de Hendley (1998). El primero fijó en siete los factores de costes: creación, adquisición, administración, descripción, uso, preservación y gestión de derechos. El segundo estableció las fases funcionales del ciclo de vida de la preservación: la creación de datos, la selección y evaluación, la gestión, la descripción, el uso, la preservación y la gestión de derechos; determinando los factores de costes particulares asociados a cada una de dichas etapas.

Russell y Weinberger (2000) establecieron la diferencia entre costes de preservación y costes de acceso, ya que preservación y acceso representan procesos tecnológicos y de gestión con diferentes costes asociados. Sin embargo, están vinculados, por el hecho de que el acceso al material, incluso en condiciones de alta restricción, confiere valor a la información almacenada.

Ashley (1998) subraya los costes y ayuda a los conservadores a elaborar sus propios cálculos, hace lo mismo Sanett (2002), mientras que Kenney y Rieger (2000) tratan de los costes de reformatación, Crespo y García Molina (2001) se centran en el diseño del repositorio de archivos, Russell y Weinberger (2000) 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, 2000; Holdsworth, 2001; Lorie, 2001; Rothenberg, 1999; Rothenberg; Bikson, 1999; Wheatley, 2001).

Russell y Weinberger (2000) también identificaron 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. Sanet (2002, 2003) estudió el modelo de Beagrie y propuso un modelo basado en 3 categorías de costes: de capital, operativos directos y operativos indirectos. Al mismo tiempo estos primeros estudios reflejan la creciente conciencia sobre la complejidad del análisis de costes en sus componentes, y comenzaron a mostrar la comprensión de las relaciones entre las tecnologías y la amplia gama de actividades necesarias para alcanzar la preservación permanente (Eakin *et al.*, 2008, 4). Con el tiempo la perspectiva se fue enriqueciendo hasta comprender los costes de preservación de un objeto a lo largo de todo su ciclo de vida, como lo ejemplifican muy bien los modelos de costes del proyecto *Life Cycle Information For E-Literature* (LIFE), y el de Beagrie, Chruszcz, Lavoie (Eakin *et al.*, 2008).

El primer proyecto que se propuso llevar a la práctica los modelos teóricos precedentes fue Digital Preservation Testbed (2000-2005) desarrollado por los Archivos Nacionales de Países Bajos, posteriormente unido al proyecto Planets (<https://www.planets-project.eu/>), que en 2010 pasó a formar parte de Open Preservation Foundation (<https://openpreservation.org/>). Se experimentó con cuatro clases de documentos: textuales, hojas de cálculo, correos electrónicos y bases de datos de distinto tamaño, complejidad y naturaleza. El proyecto produjo unos indicadores de costes y un modelo orientado a ponderar los costes de capital y de personal, conscientes de las limitaciones del tema en fechas tan tempranas. Se trataba, en todo caso, de un modelo inicial y estimativo que fuera puesto a prueba y mejorado, en futuros proyectos (Verdegem; Slats, 2004). El modelo, sin embargo, no tuvo continuidad y se encuentra desactualizado y fuera de uso.

La Universidad de Glasgow y el Joint Information Systems Committee (JISC) produjeron un conjunto de recomendaciones estratégicas para elaborar proyectos no orientados a ofrecer beneficios económicos de manera inmediata a una organización, sino beneficios intangibles, conocido como proyecto ESPIDA (2005-2007). Centrado en asegurar que el proceso decisional sea transparente y cuente con toda la información relevante, también incluyó una herramienta para evaluar los costes (Curral; MacKinney, 2007); pero este proyecto no tuvo mayor desarrollo y terminó decayendo.

Los Modelos Bibliotecarios

Una parte considerable de los modelos han sido desarrollados bajo el liderazgo de las bibliotecas nacionales, como resultado de su deber de preservación del patrimonio bibliográfico, así como universitarias vinculadas especialmente al tema de los datos de investigación.

En relación con el compromiso patrimonial, la British Library junto con la University College London lanzaron el proyecto LIFE, con el propósito de elaborar un modelo de cálculo de costes de la preservación basado en el concepto de ciclo de vida digital (Hole *et al.*, 2010). Desarrollado en tres fases entre 2005 y 2010, en la primera (LIFE1) se construyó el Generic Preservation Model (GPM), que servía para estimar los costes de preservación del contenido de una forma básica y supuso una gran contribución para la comprensión de los costes a largo plazo y para ayudar a las instituciones a planificar el futuro de sus colecciones digitales (McLeod; Wheatley; Ayrís, 2006). En la segunda fase (LIFE2) el modelo fue probado con estudios de caso y se aplicó a la preservación de materiales analógicos (publicaciones periódicas), y así poder comparar los ciclos de vida analógico y digital en relación con los costes (Davies, 2008). La tercera (LIFE3) produjo una herramienta de previsión de costes con el objetivo de “mejorar nuestra capacidad de anticipar las necesidades de dotación de recursos de las futuras actividades de la preservación digital, guiar la toma de decisiones sobre si o cuándo adquirir, o cuando hacerlo, cómo o cuándo preservar, y la cantidad de recursos que hay que poner en marcha a largo plazo” (Wheatley; Hole, 2009, p. 208). Se trataba de una hoja de cálculo útil para numerosos escenarios de preservación.

El Ministerio de Cultura de Dinamarca financió el proyecto Cost Model for Digital Preservation (2009-2012) destinado a instituciones nacionales de patrimonio cultural, a fin de aumentar la rentabilidad de las actividades de preservación digital y proporcionar una base para comparar y estimar los requisitos de costes futuros. Un modelo basado en el modelo funcional, también conocido por modelo de la actividad, contabiliza todos los costes económicos, directos e indirectos, proporcionando un mayor nivel de detalle que el alcanzado por los proyectos precedentes (Kejser; Nielsen; Thirifays, 2011). Transcurridos diez años de su finalización, los enlaces están fuera de línea y no es posible encontrar información en los portales web de las instituciones nacionales danesas. Un destino parecido le cupo a California Digital Library, que lanzó el Total Cost of Preservation Analysis (2011-2013), con el que modelar los costes económicos totales de la preservación a largo plazo (Abrams *et al.*, 2012).

La MetaArchive Cooperative (<https://www.metaarchive.org>), una cooperativa de bibliotecas y archivos que trabajan en beneficio mutuo mediante la preservación distribuida de sus materiales empleando el software LOCKSS, creada en 2004 y administrada por Educopia Institute, es una de las redes responsable de preservar el contenido de sus miembros en una red descentralizada y distribuida de cinco copias. En 2015 lanzó la iniciativa Getting to the Bottom Line: 20 Digital Preservation Cost Questions (Schultz; Trehub; Skinner, 2015). “Este recurso fácil de usar fue diseñado para suministrar a las instituciones preguntas que puedan utilizar para identificar la gama de posibles costes asociados a cualquier servicio de preservación digital en particular, ya sea propietario, compartido u otro” (Schultz; Trehub; Skinner, 2015, p. 156). Este conjunto de preguntas funciona al estilo de un modelo de requisitos y fue desarrollado para ayudar a las instituciones evaluar sus opciones en detalle y recopilar la información necesaria para tomar decisiones informadas que conduzcan a soluciones sostenibles. Pero la ausencia de información sobre su uso nos induce a pensar que la iniciativa no tuvo éxito. A decir de la propia cooperativa, en su declaración del 7 de junio de 2023, donde reconocía de algún modo la difícil sostenibilidad del modelo:

A medida que la infraestructura de TI en el mundo académico y las organizaciones de patrimonio cultural se subcontrata en gran medida a la nube, mantener un nodo local para una red como MetaArchive comienza a verse como un servicio “boutique” que es insostenible para algunos de nuestros miembros (MetaArchive, 2023).

La preservación de datos de investigación es otro ámbito bibliotecario llevó a la elaboración de modelos de costes. Así, Keeping Research Data Safe Activity Model (2007-2010) fue desarrollado conjuntamente por las universidades de Cambridge, Southampton y Kings College London, bajo la dirección de Neil Beagrie, Julia Chruszcz y Brian Lavoie. Basado en el ciclo de vida de los datos científicos, divido en fases y éstas a su vez en actividades y subactividades; cada una de las cuales representa una variable de coste (Beagrie; Lavoie; Woollard, 2010). El modelo continuó siendo integrado en el Costs-Benefits Advocacy Toolkit (2017) del Consortium of European Social Science Data Archives (CESSDA), financiado por el programa Horizon 2020. Su principal objetivo fue

desarrollar la madurez de todos los servicios nacionales de archivo de datos en Europa de manera coherente y deliberada hacia la visión de una infraestructura de datos de investigación en ciencias sociales distribuida e integrada, facilitando el acceso de los investigadores a los recursos de datos de ciencias sociales, independientemente de la ubicación del investigador o de los datos (Stokes; Kilbridge; Beagrie, 2016, traducción nuestra).

Continuó con el desarrollo de la metodología y con el conjunto de herramientas, hasta lanzar en abril de 2017 Costs-Benefits Advocacy Toolkit (<https://www.cessda.eu/Tools/Resource-Directory>), bajo licencia Creative Commons contiene: una guía de usuario, tres fichas informativas (Beneficios, Costes y Retorno de la Inversión), cuatro estudios de caso de archivos de datos de ciencias sociales (Eslovenia, Finlandia, Lituania y Reino Unido), dos hojas de trabajo (el cuadro de desarrollo del archivo, y el resumen de beneficios para un archivo de datos). Diseñada para ayudar a los archivos de datos de ciencias sociales con la promoción de sus políticas de coste-beneficio es más un conjunto de recomendaciones que una herramienta ejecutable. Tampoco hay datos ni constancia acerca de su uso o estudios de caso, además del proyecto.

Data Archiving and Networked Services (DANS) de los Países Bajos desarrolló Cost Model for Digital Archiving (2011), un modelo para el cálculo de costes para la preservación y difusión de datos de investigación. Parte de la constatación de que hay más información que recursos para la preservación digital permanente, y se orienta a lograr la sostenibilidad económica del archivo de datos (Palailogk *et al.*, 2012). Otro proyecto desaparecido e ilocalizable.

Los Modelos Competitivos

Además de las instituciones de memoria, especializadas en la conservación del patrimonio documental, también han surgido iniciativas desde el entorno empresarial o de la investigación científica con vocación de transferencia a la sociedad. El primero de los proyectos fue desarrollado en la NASA que, en una fecha tan temprana como 2001, diseñó Cost Estimation Toolkit (CET), para estimar el ciclo de vida completo de los sistemas de datos científicos provenientes de las operaciones terrestres y del espacio, lo que comprendía su introducción, uso y conservación. La herramienta contempla 96 factores de costes y aunque concebida como un instrumento multipropósito, se viene aplicando con éxito a la preservación a largo plazo (Xue *et al.*, 2011). Hasta tal punto que es por hoy el único modelo para el cálculo de costes de la preservación digital en uso, bien que sólo se utiliza en el campo de los datos científicos aeroespaciales.

El 7º Programa Marco de la Unión Europea financió el proyecto 4C-Collaboration to Clarify Costs of Curation (2013-2015), que tiene como objetivo analizar los modelos previos, elaborar nuevas propuestas y favorecer la conciencia y el intercambio de información y de experiencias para el desarrollo de un modelo y una herramienta robustos para el cálculo de costes (<https://4cproject.eu/>). Como resultado, se creó el espacio Curation Costs Exchange (<http://www.curationexchange.org/>), 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 confiables.

Curation Costs Exchange fue concebido como la aplicación práctica de los resultados de la investigación, buscando consolidar una comunidad de interesados en los costes de la preservación digital y al mismo tiempo ofrecer una herramienta de comparación de costes (CCT), que les permitiera compartir sus datos, realizar análisis y hacer comparaciones con los de sus pares. Sin embargo, fue diseñado como una comunidad cerrada, no publicó datos, y esto probablemente favoreció que se le prestara poca atención y que decayera. Al momento de la redacción este trabajo, o la página web está parcialmente inaccesible, el área interactiva de comparación de costes (el CCT) fuera de línea, y el marco web y los módulos subyacentes sin actualizarse para restaurar su funcionalidad. Tampoco admite la inscripción de nuevos usuarios. El 3 de marzo de 2023, la Digital Preservation Coalition (DPC) lanzó una encuesta entre sus usuarios para conocer su interés en la herramienta y decidir su destino. El plazo para contestar finalizó el 30 de abril y terminado el año, los resultados son desconocidos.

M. Badawy planteó en su tesis doctoral (2017) el Proyecto ENSURE: Enabling kNowledge Sustainability Usability and Recovery for Economic value, una evolución del proyecto europeo del mismo nombre en el que participara, desarrollado entre 2011 y 2014 (<https://cordis.europa.eu/project/id/270000>). La idea original era estudiar la gestión del ciclo de vida de la preservación de datos producidos o controlados por organizaciones con intereses comerciales en contraposición la tendencia dominante centrada en datos del patrimonio cultural relativamente homogéneos. De modo que es el primer proyecto pensado específicamente para el ámbito empresarial, a partir del cual Badawy se centra en los costes. Su análisis parte de una constatación: quienes toma decisiones en el mundo empresarial necesitan controlar el alcance de los compromisos financieros previsibles que van a asumir y, como las inversiones en preservación permanente se deben sostener a largo plazo, es acuciante la necesidad de comprender con claridad los costes y su comportamiento futuro; de modo que la selección de los activos digitales a preservar sea una decisión informada. Esta premisa destaca la dificultad de prever los costes, dado el impacto de la obsolescencia y de las incertidumbres, que son al tiempo la razón por la que la preservación es necesaria.

El proyecto tiene como objetivo proporcionar una solución total de preservación digital permanente para un nuevo sector en el mercado de preservación en constante crecimiento. Las empresas de la sanidad, finanzas y ensayos clínicos están ahora interesadas en preservar sus datos, debido a las obligaciones legales y al coste creciente de la regeneración de datos, especialmente en el sector de los ensayos clínicos. Junto con los nuevos sectores comerciales, ENSURE aspira a liderar la computación en la nube en sus métodos de almacenamiento y aprovechar la potencia informática de los proveedores de la nube (Badawy, 2017, p. 19, traducción nuestra).

Para este fin, desarrolló un modelo de costes de punto único (Single-Point Cost Model), que produce un valor único que representa el coste estimado.

Así como el modelo de Badawy se centra en las empresas del sector de la sanidad que operan sus datos en la nube comercial, el último de los ensayos es una sugerente propuesta elaborada por

J. L. de la Rosa, A. El-Fakdi, J. A. de Olvera, J. A. y P. N. Carrillo, investigadores del área de ingeniería de sistemas de la Universidad de Girona. *Cost-Aware Digital Objects (2015/2022)* plantea un nuevo modelo de preservación digital al que denominan paradigma *Self Preserved Digital Object (SPDO)*, en el cual los objetos digitales son actores activos en la gestión de su preservación permanente, a través de la responsabilidad sobre el presupuesto disponible (El-Fakdi; de la Rosa, 2022).

Estos objetos digitales responsables de su propia preservación digital permanente se denominan *Cost Aware Digital Object (CADO)*, cuentan con su propio presupuesto para la contratación de servicios de preservación, como almacenamiento, prevención pérdida de datos, evaluación de riesgos, recuperación de datos, extracción de metadatos, migración y nuevos formatos de almacenamiento, asegurando que permanezcan accesibles, legibles, íntegros y auténticos. En este paradigma de preservación autosostenida, los objetos digitales están programados de forma inteligente con habilidades de detección conscientes de los costes, gestionando un presupuesto determinado para asegurar su supervivencia a largo plazo, negociando con repositorios de almacenamiento y servicios web mediante mecanismos de subasta; una opción cada vez más frecuente tanto en los repositorios descentralizados comerciales, como en los de código abierto, que están comenzando a ganar visibilidad impulsados por el auge de las tecnologías blockchain y las estructuras descentralizadas. Buenos ejemplos de ello son Filecoin (<https://filecoin.io/>), Storj (<https://www.storj.io/how-it-works>), y Sia (<https://sia.tech/>).

Más que un modelo para el cálculo de costes, se trata de un nuevo contexto donde los objetos digitales comparten la responsabilidad de la preservación y que en cierta medida resuelve el problema del cálculo de costes derivándolo al terreno de la competitividad, de la oferta y la demanda del mercado.

Discusiones y Conclusiones

El cálculo de los costes de la preservación digital permanente se revela como un tema abierto, lejos de haber logrado un modelo de uso general, posiblemente debido a la propia naturaleza diversa de los materiales y de las entidades interesadas. Un tema sobre el que se han arrojado algunas luces, pero que alberga muchas zonas de sombra a pesar del tiempo transcurrido. Así, en la literatura, ha aumentado la atención, pero con un exceso de reiteración y pocas contribuciones, lo cual también ocurre con cuanto tiene que ver con la preservación digital. Otra característica de la literatura es que se centra en el lado de los costes, y comparativamente presta poca atención a la articulación de los beneficios; en cambio, a menudo se toman por conocidos 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 (Cruz Mundet; Díez Carrera, 2015).

En los veinticinco años que comprende el período de análisis (1998-2023) se han sucedido catorce proyectos de amplio propósito, buena muestra de la preocupación existente en la comunidad interesada en la preservación. El éxito medido en términos de continuidad parece indicar que nos encontramos en una fase inicial del camino hacia la parametrización de los costes. El balance final es de doce proyectos desaparecidos, dos sin perspectivas claras de continuidad, aunque su cronología es demasiado reciente para confirmarlo (ENSURE y CADO), y sólo uno, *Costs Estimation Toolkit* de la NASA, ha sobrevivido y continúa utilizándose, bien que restringido a una tipología (datos científicos).

Los factores que influyeron en la falta de continuidad y de éxito son diversos. Los modelos se construyeron mayoritariamente sobre situaciones individuales, sin posibilidad de ser extrapolados,

salvo excepciones, y estas se refieren a categorías de instituciones (sobre todo las de educación superior), a tipos de ficheros (datos, documentos...) o tipologías de materiales (audiovisuales, textuales...); son difíciles de comparación y de ser aplicados a la realidad. Ya lo señalaron en su estudio Kejsler, Nielsen y Thirifays (2011, p. 256): “Otro tema sin resolver es el desarrollo de fórmulas para crear modelos de costes operativos.” En cualquier caso, los costes son componentes de los modelos de sostenibilidad, que también deben considerar el valor y el contexto en el que se alcanza la comprensión de valor. Es una dimensión cultural o institucional por la que se atribuye valor a un objeto; lo que impide que prevalezca la dimensión técnica, así como las técnicas económicas y financieras que pueden ser usadas para cuantificar la noción de valor. Como señaló Bradley (2007, p. 157): “Es claramente imposible ni preservar información digital sin una infraestructura organizacional, económica, social, estructural y técnica sostenible, ni es sensato preservar material sin valor sostenido”.

Los modelos de costes tienen un enfoque de ciclo de vida, como ya se demostró hace tiempo 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; Weinberger, 2000). Y que los costes de preservación dependen en gran medida del rango de servicios que ofrece una institución. En lo que no hay acuerdo todavía es sobre cómo debería ser estructurado el ciclo de vida de los costes de la preservación digital; o cómo se puede desglosar y detallar las fases individuales del ciclo de vida. Al respecto, Sanet sugirió desarrollar un marco de costes y su asignación en un modelo funcional bien definido, así como aplicar principios contables generalmente aceptados (Sanett, 2002).

Los proyectos no sólo deberían responder a la pregunta de cuánto cuesta, sino también a la de por qué necesitamos hacerlo, y por qué debemos gastar dinero en esto, 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 (Currell; MacKinney, 2007). Una de esas preguntas, esencial, es el riesgo de la inactividad y sus consecuencias, planteada con anterioridad por Lavoie y Dempsey (2004); un aspecto sobre el que insistió años después Beagrie (2017), los costes de la inacción (Costs of Inaction), no hacer nada y medir sus consecuencias, lo cual “podría contribuir exitosamente a la promoción y a los estudios de caso de preservación digital”.

Cada día es mayor la conciencia sobre los costes medioambientales derivados de la digitalización, como consecuencia del consumo intensivo de energía, lo que ha llevado a constatar la necesidad de seleccionar aquello que vale la pena conservar, y concluir que no todos los objetos digitales son activos digitales, sino sólo aquellos que tienen un valor presente y futuro y han sido seleccionados. Es necesario identificar el valor de los activos digitales y tomar decisiones. Mientras que otros proponen un cambio de paradigma respecto a la permanencia, abogando por determinar un nivel de pérdida aceptable frente a la conservación redundante de copias: “Dada la inevitabilidad de las pérdidas y el alto coste ambiental de la preservación digital, los administradores de los sistemas de preservación digital deben determinar niveles aceptables de pérdida en los programas de preservación digital” (Pendergrass *et al.*, 2019).

La creación de una comunidad de preservación que comparta de forma abierta y transparente sus experiencias y datos pormenorizados de costes distribuidos por factores continúa siendo una aspiración puesta sobre la mesa con insistencia (Dohe; Durden, 2018; Schmitt, 2016). Tal ha sido el propósito de algunos de los proyectos examinados en este trabajo, pero fracasaron en el intento, lo que lleva a algunos autores a creer que es imposible predecir el coste de conservarlo para siempre

“*keeping it forever*” (Dohe; Durden, 2018) o, dicho de otro modo, que no hay técnicas contrastadas para estimar los costes de preservar información digital a largo plazo (Ismail; Affandi, 2018). Siendo esto así, sería necesario buscar las causas, cosa que no hemos detectado o no hemos sido capaces de hallar en nuestra investigación, algo que nos proponemos hacer en un próximo proyecto.

Referencias

- Abrams, S. *et al.* *Total cost of preservation: Cost modeling for sustainable services*. California: California Digital Library, 2012. Disponible en: <https://wiki.diglib.org/images/archive/0/0a/201605031729141TCP-total-cost-of-preservation.pdf>. Acceso en: 14 sept. 2023.
- Ashley, K. Digital archive costs: Facts and fallacies. In: DLM Forum on Electronic Records (DLM '99), 1998, Bruselas. *Proceedings* [...]. Bruselas: DLM, 1998. p. 121-128. Disponible en: http://ec.europa.eu/archives/ISPO/dlm/fulltext/full_ashl_en.htm. Acceso en: 6 jun. 2023.
- Badawy, M. *A framework for whole lifecycle cost of long-term digital preservation*. 2017. Dissertation (Doctoral in Engineering) - Cranfield University, Bedfordshire, United Kingdom, 2017. Disponible en: <http://dspace.lib.cranfield.ac.uk/handle/1826/14315>. Acceso en: 23 nov. 2023.
- Baker, M. *et al.* A fresh look at the reliability of long-term digital storage. In: EuroSys European Conference on Computer Systems, 2006, Leuven, Bélgica. *Proceedings* [...]. New York: ACM, 2006. p. 221-234. <https://doi.org/10.1145/1217935.1217957>.
- Beagrie, N. The Costs of Inaction: advocating for digital preservation. [S. l.: s. n.], 2017. Disponible en: <https://blog.beagrie.com/2017/12/04/the-costs-of-inaction-advocating-for-digital-preservation/>. Acceso en: 23 nov. 2023.
- Beagrie, N.; Greenstein, D. *A strategic policy framework for creating and preserving digital collections: A report to the digital archiving working group*. London: British Library Research and Innovation Centre, 1998. Disponible en: <https://www.ukoln.ac.uk/services/elib/papers/supporting/pdf/framework.pdf>. Acceso en: 17 nov. 2023.
- Beagrie, N.; Lavoie, B.; Woollard, M. *Keeping Research Data Safe 2: Final report*. London: Joint Information Systems Committee, 2010. Disponible en: <https://repository.essex.ac.uk/2147/>. Acceso en: 21 nov. 2023.
- Bradley, K. Defining digital sustainability. *Library Trends*, v. 56, n. 1, p. 148-163, 2007. Disponible en: <https://hdl.handle.net/2142/3772>. Acceso en: 15 sept. 2023.
- Crespo, A.; Garcia-Molina, H. Cost-Driven design for archival repositories. In: Joint Conference on Digital Libraries 2001 (JCDL'01), 2001, Roanoke, Virginia, USA. New York: ACM, 2001. p. 363-372. Disponible en: <https://doi.org/10.1145/379437.379729>. Acceso en: 15 sept. 2023.
- Cruz Mundet, J.R.; Díez Carrera, C. *Los costes de la preservación digital permanente*. Gijón: Ediciones Trea, 2015.
- Currall, J.; McKinney, P. *Espida handbook: Expressing project costs and benefits in a systematic way for investment in information and IT*. Glasgow: University of Glasgow: JISC, 2007. Disponible en: <https://eprints.gla.ac.uk/140314/1/140314.pdf>. Acceso en: 18 sept. 2023.
- Davies, R. (ed.). *The LIFE2 Final Project Report*. Bristol: Joint Information Systems Committee, 2008. Disponible en: <https://discovery.ucl.ac.uk/id/eprint/11758>. Acceso en: 12 sept. 2023.
- Dohe, K.; Durden, D. *The cost of keeping it: Towards effective cost-modeling for digital preservation at the University of Maryland*. Boston, Massachusetts: iPRES, 2018. <https://doi.org/10.13016/M2CZ3283Q>.
- Dollar, C. M. *Authentic electronic records: Strategies for long-term access*. Minnesota: Cohasset Associates, 2000. Disponible en: <http://localhost:8080/xmlui/handle/123456789/2860>. Acceso en: 1º sept. 2023.
- Eakin, L. *et al.* *A selective literature review on digital preservation sustainability model*. [S. l.: s. n.], 2008. Disponible en: https://www.academia.edu/3708793/A_Selective_Literature_Review_on_Digital_Preservation_Sustainability. Acceso en: 23 sept. 2023.
- El-Fakdi, A.; de la Rosa, J. evaluating auction mechanisms for the preservation of cost-aware digital objects under constrained digital preservation budgets. *Mathematics*, v. 10, n. 1, p. 92, 2022. <https://doi.org/10.3390/math10010092>.

Erkoyuncu, J. A. *et al.* Uncertainty challenges in service cost estimation for product-service systems in the aerospace and defense industries, *In: Ng, I. et al. Complex engineering service systems concepts and research*. London: Springer, 2009. p. 129-147.

Granger, S.; Russell K.; Weinberger, E. *Cost elements of Digital Preservation (version 4)*. [S. l.: s. n.], 2000.

Hendley, T. *Comparison of methods & costs of digital preservation*. London: British Library Research and Innovation Centre, 1998. British Library Research and Innovation Report 106. Disponible en: <http://localhost:8080/xmlui/handle/123456789/11904>. Acceso en: 12 sept. 2023

Holdsworth, D.; Wheatley, P. Emulation, preservation, and abstraction. *RLG Digi News*, v. 5, n. 4, 2001. Disponible en: <https://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file3149.html#feature2>. Acceso en: 2 sept. 2023.

Hole, B. *et al.* The LIFE3 predictive costing tool for digital collections. *New Review of Information Networking*, v. 15, n. 2, 2010. <https://doi.org/10.1080/13614576.2010.526014>.

Ismail, A.; Affandi, H. B. Conceptual paper: Digital preservation strategies in archival institution. *MATEC Web of Conferences*, v. 150, 05052, 2018. <https://doi.org/10.1051/mateconf/20181500505.2>.

Kejser, U. B.; Nielsen, A. B.; Thirifays, A. Cost model for digital preservation: Cost of digital migration. *The International Journal of Digital Curation*, v. 1, n. 6, p. 255-267, 2011. <https://doi.org/10.2218/ijdc.v6i1.186>.

Kenney, A. R.; Rieger, O. Y. *Moving theory into practice: Digital imaging for libraries and archives*. Mountain View, CA: Research Libraries Group, 2000. Disponible en: https://archive.org/details/movingtheoryinto0000unse_q5h2. Acceso en: 7 sept. 2023.

Kuny, T. A Digital Dark Ages? Challenges in the Preservation of Electronic Information. *In: IFLA Council and General Conference, 63., Workshop Audiovisual and Multimedia joint with Preservation and Conservation, Information Technology, Library Buildings and Equipment, and the PAC Core Programme, 1997, Copenhagen. Proceedings [...]. Copenhagen: IFLA, 1997*. Disponible en: http://www.academia.edu/19711051/A_Digital_Dark_Ages_Challenges_in_the_Preservation_of_Electronic_Information. Acceso en: 14 sept. 2023.

Lavoie, B. F.; Dempsey, L. Thirteen ways of looking at digital preservation. *RLG DigiNews*, v. 8, n. 2, 2004. Disponible en: <https://dlib.org/dlib/july04/lavoie/07lavoie.html>. Acceso en: 14 sept. 2023.

Lorie, R. A. Long term preservation of digital information. *In: ACM/IEEE-CS Joint Conference on Digital Libraries, 2001, Roanoke, Virginia. Proceedings [...]. New York: ACM, 2001*. p. 46-352. <https://doi.org/10.1145/379437.379726>.

McLeod, R.; Wheatley, P.; Ayris, P. *Lifecycle information for e-literature: Full report from the LIFE project*. London: UCL, 2006. Disponible en: <https://discovery.ucl.ac.uk/id/eprint/1854>. Acceso en: 22 jun. 2023.

MetaArchive. MetaArchive's Community Research Task Force is Planning for the Future of the Community. *News & Events*, 2023. Disponible en: <https://metaarchive.org/category/news-events/>. Acceso en: 6 sept. 2023.

Palaiologk, A. S. *et al.* An activity-based costing model for long-term preservation and dissemination of digital research data: The case of DANS. *International Journal of Digital Libraries*, v. 12, n. 4, p. 195-214, 2012. <https://doi.org/10.1007/s00799-012-0092-1>.

Pendergrass, K. L. *et al.* Toward environmentally sustainable digital preservation. *The American Archivist*, v. 82, n. 1, p. 165-206, 2019. <https://doi.org/10.17723/0360-9081-82.1.165>.

Romero Rojo, F. J. *Development of a framework for obsolescence resolution cost estimation*. 2011. Doctoral dissertation (Doctoral in Engineering) - Cranfield University, Bedfordshire, United Kingdom, 2011. Disponible en: <http://dspace.lib.cranfield.ac.uk/handle/1826/6854>. Acceso en: 12 jun. 2023.

Rosenthal, A. *et al.* Cloud computing: A new business paradigm for biomedical information sharing, *Journal of Biomedical Informatics*, v. 43, n. 2, p. 342-353, 2010. <https://doi.org/10.1016/j.jbi.2009.08.014>.

Rothenberg, J. *Avoiding technological quicksand: Finding a viable technical foundation for digital preservation*. Washington: Council on Library and Information Resources, 1999. Disponible en: <https://eric.ed.gov/?id=ED426715>. Acceso en: 1º Sept. 2023.

Rothenberg, J.; Bikson, T. K. *Carrying authentic, understandable and usable digital records through time*. Santa Monica, CA: RAND Corporation, 1999. Disponible en: https://www.rand.org/pubs/rand_europe/RE99-016.html. Acceso en: 21 jun. 2023.

- Russell, K. Digital preservation and the CEDARS project experience. *New Review of Academic Librarianship*, v. 6, 139-154, 2000. <https://doi.org/10.1080/13614530009516805>.
- Russell, K.; Weinberger, E. *Cost Elements of Digital Preservation*. Leeds: Cedars Project, 2000. Disponible en: <https://www.stewardshipgap.net/node/414>. Acceso en: 11 sept. 2023.
- Sanett, S. The cost to preserve authentic electronic records in perpetuity: comparing costs across cost models and cost frameworks. *RLG Diginews*, v. 7, n. 4, p. 8, 2003. Disponible en: <http://worldcat.org/arcviewer/2/OCC/2009/08/11/H1250009815353/viewer/file2.html>. Acceso en: 6 jun. 2023.
- Sanett, S. Toward developing a framework of cost elements for preserving authentic electronic records into perpetuity. *College and Research Libraries*, v. 63, n. 5, p. 388-404, 2002. <https://doi.org/10.5860/crl.63.5.388>.
- Schmitt, K. Kostenmodelle in der digitalen Bestandserhaltung. *Bibliotheksdienst*, v. 50, n. 1, p. 49-61, 2016. <https://doi.org/10.1515/bd-2016-0005>.
- Schultz, M.; Trehub, A.; Skinner, K. Getting to the bottom Line: 20 digital preservation cost questions. In: International Conference on Digital Preservation, 12., 2015, Chapel Hill, North Carolina. *Proceedings* [...]. Chapel Hill, North Carolina: iPRES, 2015. p. 156-158. Disponible en: <http://hdl.handle.net/11353/10.429558>. Acceso: 21 jun. 2023.
- Sistemas de Transferencia de Datos e Información Espaciales. *UNE-ISO 14721:2015: Modelo de referencia*. Madrid: Aenor, 2015.
- Stokes, P.; Kilbride, W.; Beagrie, N. Building the business case and funding advocacy for data management services. In: International Conference on Digital Preservation, 12., 2016, Bern. *Proceedings* [...]. Bern: iPRES, 2016. Disponible en: <https://phaidra.univie.ac.at/o:502825>. Acceso: 6 jun. 2023.
- Verdegem, R.; Slats, J. Practical experiences of the Dutch digital preservation test-bed, *VINE*, v. 34, n. 2, p. 56-65, 2004. <https://doi.org/10.1108/03055720410531004>.
- Wheatley, P. Migration: A CAMiLEON discussion paper. *Ariadne*, v. 29, 2001. Disponible en: <https://www.ariadne.ac.uk/issue/29/camileon/>. Acceso: 12 jun. 2023.
- Wheatley, P.; Hole, B. LIFE3. Predicting long term digital preservation costs. In: International Conference on Preservation of Digital Objects, California Digital Library, 6., 2009, San Francisco, California. *Proceedings* [...]. San Francisco, California: iPRES, 2009. p. 206-210. Disponible en: <https://escholarship.org/uc/item/23b3225n>. Acceso: 12 jun. 2023.
- Xue, P. *et al.* Cost modelling for long-term digital preservation: Challenges and issues. In: International Conference on Manufacturing Research, 9., 2011, Atlanta. *Proceedings* [...]. Atlanta: ICMR, 2011. p. 187-192. Disponible en: https://www.academia.edu/1215797/COST_MODELING_FOR_LONG_TERM_DIGITAL_PRESERVATION_CHALLENGES_AND_ISSUES. Acceso: 2 sept. 2023.

Colaboración

Conceptualización: C. Díez y J.R. Cruz; Investigación: C. Díez y J.R. Cruz; Metodología: C. Díez; Redacción – borrador original: C. Díez y J.R. Cruz; Redacción – revisión y edición: C. Díez y J.R. Cruz.