

Curadoria Digital e Custos – Exploração de abordagens e percepções [en]

Digital Curation and its Costs: A Study of Practices and Insights [pt]

Luís Corujo <luiscorujo_at_campus_dot_ul_dot_pt>, Universidade de Lisboa, Faculdade de Letras, Centro de Estudos Clássicos
Jorge Revez <jrevez_at_campus_dot_ul_dot_pt>, Universidade de Lisboa, Faculdade de Letras, Centro de Estudos Clássicos
Carlos Guardado da Silva <carlosguardado_at_campus_dot_ul_dot_pt>, Universidade de Lisboa, Faculdade de Letras, Centro de Estudos Clássicos

Abstract

Introdução – No âmbito das preocupações que estão na origem da curadoria digital, toma-se como exemplo o contexto da produção de grandes volumes de informação científica, que requer abordagens que garantam a sua manutenção, reutilização e valorização, dado o seu elevado custo.

Objetivos – Pretende-se conhecer o pensamento existente referente aos custos da curadoria digital e desenvolver uma proposta de modelo de enquadramento para a análise de custos em curadoria digital, inclusivamente projetos de carácter mais prático. Tal implica abordar a definição deste conceito e a problemática dos custos, baseando-nos nos estudos referentes a modelos de custos.

Metodologia - Procedeu-se a uma revisão da literatura, contextualizando a questão dos custos no seio da curadoria digital, delimitada pelos dados recolhidos nas fontes de pesquisa, a *Biblioteca do Conhecimento Online* (B-On) e o *Repositórios Científicos de Acesso Aberto de Portugal* (RCAAP). Em seguida, desenvolveu-se um modelo de enquadramento esquematizado com base nas categorias identificadas por via do Método da Comparação Constante, algumas relacionadas com o Modelo de Referência *Open Archival Information System* (OAIS) e o ciclo de vida *Digital Curation Centre* (DCC). Tal permitiu desenvolver uma análise de conteúdo das percepções recolhidas em diversos autores, resultando em memorandos, de que este texto é um resumo.

Resultados/Conclusão – Propõe-se um esquema de sistematização das problemáticas da curadoria digital, que constitui um modelo de enquadramento que se considera pertinente para a análise de custos em curadoria digital, inclusivamente projetos de carácter mais prático, e que interliga a visão do ciclo de vida da curadoria do objeto digital do DCC e a abordagem do modelo de referência OAIS, numa lógica transversal apreendida pelos Modelos de Custos e Plano/Políticas de gestão dos dados. Conclui-se que se deteta uma mudança de paradigma, de uma visão de *black-box* para uma abordagem de identificação dos custos e de tentativa de sistematização de modelos de previsão para utilização institucional, como forma de incentivar a transparência e a *accountability* e de captar o interesse de potenciais financiadores.

Introdução

No âmbito do estudo dos custos relativos à curadoria digital, partimos do exemplo do movimento da Ciência Aberta (*Open Science*), que contribuiu para a tomada de consciência da importância dos dados, na medida em que estes fornecem evidências do conjunto do conhecimento científico, constituindo a base do seu desenvolvimento [Molloy 2011]. Não tendo este estudo a intenção de problematizar os conceitos de *Open Science*, *Big Science*, *Big Data*, *Open Data* no âmbito deste exemplo de contextualização, refere-se que os dados científicos permitem lidar com questões, por exemplo, ecológicas e climáticas, de saúde, segurança nacional e nanotecnologia [Poole 2015], sendo, assim, um ativo intelectual de primeira importância, passível de revisão por pares, avaliação de qualidade e reutilização [Heidorn 2011]. Os dados científicos são objetos sociotécnicos complexos que existem dentro das comunidades, e muito mais - e menos - do que simples mercadorias que podem ser exploradas e/ou comercializadas num mercado público [Borgman 2015]. A OCDE define-os como registos factuais, usados como fontes primárias, “que são usualmente considerados na comunidade científica como necessários para validar os resultados de investigação. A definição exclui especificamente cadernos de notas de laboratório, análises preliminares e objetos físicos, tais como amostras de laboratório [OCDE 2007]. Embora os cientistas considerem os dados publicados pertença da comunidade científica, muitas editoras reivindicam direitos sobre esses dados e não permitem que sejam reutilizados sem a sua autorização [Murray-Rust 2008]. No âmbito da Ciência Aberta, urge que esses dados sejam disponibilizados de uma forma útil e em acesso livre, ou seja, sem custo ou barreiras, para benefício da sociedade. Quanto maior for a quantidade de dados abertos – *open data* –, maior será o nível de

transparência e reprodutibilidade, resultando numa maior eficiência do processo científico [Molloy 2011]. Existe pouco consenso sobre o que significa para os dados serem “abertos”. O primeiro enquadramento de *open data*, de Murray-Rust e Rzepa (2004), engloba a maioria das ideias posteriores. Murray-Rust e outros, sob a égide da *Open Knowledge Foundation*, desenvolveram posteriormente uma sucinta definição legal de *open data*: Um pedaço de dados ou conteúdo é aberto se alguém tiver a liberdade para o utilizar, reutilizar e redistribuir – estando somente sujeito, no máximo, ao requisito de atributo e/ou partilha [Open Data Commons 2013]. Em ambiente de negócios, as definições são mais ambíguas: informações legíveis por máquina, particularmente dados governamentais, disponibilizados para outras pessoas [Manyika et al. 2013]. A ideia de Dados Abertos no sentido de haver liberdade para serem reutilizados [Open Data Commons 2013] é uma condição necessária, mas não suficiente, para fins de investigação. Tal requer o cumprimento dos preceitos de flexibilidade, transparência, conformidade legal, proteção da propriedade intelectual, responsabilidade formal, profissionalismo, interoperabilidade, qualidade, segurança, eficiência, responsabilidade/responsabilização e sustentabilidade derivados dos princípios da OCDE [OCDE 2007]. O relatório da *Royal Society* do Reino Unido, referente à Ciência como uma iniciativa aberta [RS PSU 2012] define *open data* como dados que cumprem os critérios de disponibilidade inteligente. Os dados devem ser acessíveis, utilizáveis, avaliáveis e inteligíveis. Para além disso, a disponibilidade pode facilitar a criação de dados, podendo os dados abertos incluir também a capacidade de tratar representações de objetos de investigação como dados, independentemente dos objetos descritos estarem ou não disponíveis [Borgman 2015].

Esta conjugação de fatores conduz a um reforço da produção científica, que é já *data intensive*, na medida em que manipula muitos dados sobre todas as áreas científicas com recurso a novas metodologias de tratamento de dados (*Big Data*), e ao seu desenvolvimento em projetos de colaboração em grande escala, por vezes multidisciplinares, com orçamentos gigantescos suportados por consórcios de instituições públicas e privadas, de carácter nacional e internacional, que Weinberg (1961) apelidou de *Big Science*. De acordo com Price (1963), *Big Science* corresponde às ciências caracterizadas por grandes equipas distribuídas de colégios invisíveis de investigadores, que conhecem e trocam informações de maneira formal e informal, colaborações internacionais, grandes coleções de dados e grandes instalações de instrumentação, que surge no fim da 2.ª Guerra Mundial, contrastando com a *Little* ou *Small Science*, desenvolvida nos últimos trezentos anos de trabalho independente, numa escala menor, e caracterizada por comunidades menores, heterogeneidade de métodos e dados de investigação, e controlo, análise e infraestruturas locais [Borgman 2015] [Borgman 2007] [Cragin et al. 2010] [Taper and Lele 2004]. Como Price observou, alguns campos da *Little* ou *Small Science* podem tornar-se *Big Science*, embora a maioria permaneça pequena.

A *Big Data* é atualmente alvo de grande atenção, tendo sido apelidada de petróleo dos negócios modernos [Mayer-Schonberger and Cukier 2013], a cola das colaborações [Borgman 2007] e uma fonte de atrito entre os investigadores [Edwards 2010] [Edwards et al. 2011]. A *Big Data* é uma questão de escala, não de tamanho absoluto e, à escala, ela possibilita e torna pensáveis novas questões. No entanto, *Big data* não significa necessariamente melhores dados, e a *Little Data* é igualmente essencial para a investigação científica, e pode ser identicamente valiosa [Borgman 2015]. Distinguir entre *Big Data* e *Little Data* é problemático devido às muitas formas pelas quais algo pode ser considerado grande. O Oxford English Dictionary define *Big Data* como “dados de tamanho muito grande, e que tipicamente a sua manipulação e gestão apresentam desafios logísticos significativos; [também] o ramo da computação que envolve esses dados” [OED n.d.]. Outras definições de *Big Data* dizem respeito à escala relativa, e não ao tamanho absoluto. Mayer-Schönberger e Cukier consideram não existir uma definição rigorosa de *Big Data*, dado que, no princípio, a ideia era a de que o volume da informação tinha aumentado tanto que aquela que se analisava já não cabia na memória que os computadores utilizam para a processar, pelo que os engenheiros necessitavam de modernizar as ferramentas para a analisar. Atualmente, a *Big Data* (os dados massivos) refere-se a coisas que se podem fazer a grande escala, mas não a uma escala inferior, para extrair novas percepções ou criar novas formas de valor, de tal forma que transformam os mercados, as organizações, as relações entre os cidadãos e os governos, pondo em questão a forma como vivemos e interagimos com o mundo, pelo que assinalam o princípio de uma transformação assinalável [Mayer-Schonberger and Cukier 2013, 17]. Nesta perspetiva, constitui um recurso e uma ferramenta que serve mais para informar do que para explicar; indica o caminho para a compreensão, mas, mesmo assim, pode induzir em erro, dependendo de os dados serem bem ou mal manejados [Mayer-Schonberger and Cukier 2013, 241]. No âmbito científico, a *Big Data* é a investigação possibilitada pelo uso de dados sobre um fenómeno numa escala ou enfoque sem precedentes [Schroeder 2014]. Os dados podem ser considerados grandes ou pequenos, dependendo do que se pode fazer com eles, que ideias podem revelar e a escala de análise necessária em relação ao fenómeno de interesse. Uma definição inicial que permite distinguir as diferentes formas pelas quais os dados podem ser grandes mantém-se válida: volume, variedade, velocidade ou uma combinação destes [Laney 2001]. Um aumento substancial em qualquer uma dessas dimensões de dados pode levar a mudanças na escala da investigação e do conhecimento.

Outro termo utilizado é o *Long Tail Data*, que referencia a disponibilidade e o uso de dados nas áreas científicas ou em setores económicos. Quando aplicado à investigação académica, um pequeno número de grupos de investigação trabalha com grandes volumes de dados, alguns grupos trabalham com muito poucos dados e a maioria fica numa posição intermédia [Foster et al. 2013]. *Long Tail Data* é uma forma abreviada útil para mostrar a gama de volumes de dados em utilização por qualquer área ou grupo de

investigação. Também permite destacar o facto de que apenas algumas áreas lidam com volumes de dados muito grandes. Em suma, os volumes de dados são distribuídos de forma desigual pelas várias áreas científicas [Borgman 2015]. No entanto, a fraqueza desta metáfora está na sugestão de que as práticas de dados de qualquer área científica ou qualquer indivíduo podem ser posicionadas numa escala bidimensional. As atividades científicas são influenciadas por inúmeros fatores para além do volume de dados manipulados. Geralmente, são as questões de investigação que direcionam a escolha de métodos e dados, mas o inverso também pode ser verdade. A disponibilidade dos dados pode orientar as perguntas de investigação que podem ser feitas e os métodos que podem ser aplicados. As escolhas dos dados também dependem de outros recursos à disposição dos investigadores individuais, incluindo teoria, conhecimento especializado, laboratórios, equipamentos, redes técnicas e sociais, espaços de investigação, recursos humanos e outras formas de investimento de capital [Borgman 2015].

Esta *Big Data* não significa apenas maior quantidade de registos, denotando também que são de diferentes tipos: para além de relatórios escritos, sobressaem, principalmente, conjuntos de dados (*datasets*) não-textuais, em formato digital, que constituem dados brutos a serem trabalhados e dos quais resultam dados progressivamente mais refinados [Buckland 2011].

A noção de dados de produção científica foi enquadrada inicialmente como *data intensive research*, ou investigação intensiva de dados, nas iniciativas de política da primeira década do século XXI, incluindo a *eScience*, *eSocial Science*, *eHumanities*, *eInfrastructure*, e *cyberinfrastructure* [Atkins 2003] [Edwards et al. 2007] [Hey and Trefethen 2005] [Unsworth et al. 2006]. Os três primeiros termos acabaram por se fundir na *eResearch* [Borgman 2015].

No entanto, os dados digitais são muito mais frágeis do que fontes de prova física (analógicas) que sobreviveram durante séculos, e são vulneráveis a perdas; a sustentabilidade é difícil de alcançar. Ao contrário do papel, do papiro e das gravuras, os dados digitais não podem ser interpretados sem os instrumentos técnicos usados na sua criação. O *hardware* e o *software* evoluem rapidamente, tornando ilegíveis os registos digitais, a menos que sejam migrados para novos formatos. Estes registos digitais requerem documentação sobre os procedimentos pelos quais foram obtidos, tal como o requerem os exemplares, amostras e outros materiais que servem de fontes de prova física. A menos que sejam feitos investimentos deliberados para a curadoria de dados para uso futuro, a maioria desaparecerá rapidamente. A preservação de representações digitais de dados é apenas um aspeto no âmbito de sustentar o seu valor. Independentemente de serem digitais ou não, os dados raramente são independentes. Eles são indissociáveis dos métodos, teorias, instrumentos, *software* e contexto de investigação. A manutenção do acesso aos dados científicos requer a curadoria de cada um dos objetos e das relações existentes entre eles. Assumindo que os investigadores até podem estar dispostos a disponibilizar os seus dados, tal vontade pode aumentar se for para os disponibilizar em repositórios confiáveis que irão efetuar a curadoria e a disseminação dos seus dados indefinidamente. Os repositórios podem agregar valor aos dados por meio de metainformação, proveniência, classificação, normas para estruturas de dados e migração. Eles também podem agregar valor, tornando os dados mais visíveis e aumentar a sua usabilidade através de ferramentas e serviços. Estes são investimentos substanciais a serem feitos pelas comunidades, universidades e agências e governos que os financiam [Borgman 2015].

No contexto dos dados científicos digitais, este investimento em recursos e tempo corre perigo por causa do volume, da complexidade, do dinamismo, da proveniência, do armazenamento, da propriedade [Ogburn 2010] e da vulnerabilidade dos dados digitais. Mais ainda pelo facto de os custos tornarem proibitiva a sua reprodutibilidade, pelo que requerem abordagens que garantam a manutenção, a reutilização e a sua valorização. A curadoria digital pretende dar resposta a estas questões.

Encontrando-se o termo *curadoria* já em 1712, a sua origem remonta ao Direito Romano e, mais especificamente, à figura do curador, aquele que detinha a função da *curatoria*, isto é, a responsabilidade de cuidar das pessoas e proteger o património, quando a execução de dívidas. Mais tarde, encontra-se o “cura” católico, cuidador das almas de uma paróquia. O termo foi apropriado pelo campo da comunicação social, investigação científica e artes [Ramos 2012]. Lee & Tibbo (2011) indicam o uso do termo *data curation* nas décadas de 80 e 90 do século XX, relativamente à gestão de dados científicos, enquanto, na primeira década do novo século, um seminário sobre Arquivos, Bibliotecas e Ciência Digitais apresentou o termo *digital curation* [Beagrie and Pothen 2002]. No entanto, os mesmos autores situam a origem de este campo no relatório de Waters & Garrett (1996). O *Digital Data Curation Task Force Report* [Macdonald and Lord 2003], relativo à curadoria de dados científicos, esteve na origem do Digital Curation Centre (DCC) em 2004, tendo contribuído para a divulgação desta disciplina através de projetos, ferramentas, unidades de apoio e formação. Em 2006, Beagrie apresentou a evolução do termo no artigo inaugural do *International Journal of Digital Curation* [Beagrie 2006].

Não pretendendo apresentar uma evolução nem distinguir o conceito de *curadoria digital* do de *preservação digital*, muitos autores identificam esta como um aspeto da curadoria digital [Kejser et al. 2014] (entre outros). Tal adequa-se à forma como o conceito preservação digital é considerado neste trabalho, tal como usado por Corujo [Corujo 2015, 165–166], cujo objetivo é garantir o acesso à informação eletrónica e a sua autenticidade independentemente do formato, do *software* e do *hardware* usados na sua

produção ou utilização para que foi concebida de origem, de modo a que possa ser utilizada a longo prazo, sem qualquer constrangimento físico de plataforma, legal, patrimonial ou outro. Para o *Digital Curation Centre (DCC)*, a curadoria digital envolve a manutenção, a preservação e a valorização dos dados científicos digitais em toda a sua vida útil, tendo teorizado um ciclo de vida que inclui as fases desde a sua conceptualização até à reutilização/transformação. Para Abbot (2008), a curadoria digital inclui a gestão de grandes quantidades de dados para utilização, garantindo a recuperação e a legibilidade, aplicando-se a uma variedade de profissionais do seu ciclo de vida, desde os criadores de conteúdo até aos investidores, definidores de política e gestores de repositórios. É um exercício caro que requer investimentos significativos de tempo e competência, algo problemático para pequenas instituições, uma vez que as maiores vantagens da curadoria digital são visíveis a longo prazo e os investimentos podem demorar a dar frutos.

Higgins (2011) considera que a curadoria digital surge a partir da evolução técnica, da compreensão e do amadurecimento da atividade organizacional e do fluxo de trabalho que enfatiza o acesso e a reutilização de materiais durante o seu ciclo de vida, tendo em conta que a gestão a longo prazo do material digital passou gradualmente da preservação passiva para a curadoria ativa. Considera-se que a curadoria digital parte de uma visão holística do ciclo de vida dos dados, dependente do contexto de produção e da necessidade social de utilização, integrando diversos agentes e atividades interdisciplinares. A curadoria digital requer planificação e políticas de gestão e negócio, incluindo planos de sustentabilidade e gestão de risco - implica transparência e identificação de parceiros para financiamento, custos/investimentos, benefícios tangíveis e intangíveis, e estratégias de mitigação -, possível apenas quando integrada numa instituição comprometida com recursos e competências.

Tal como a *Big Science*, a curadoria digital requer capacidade financeira e orçamentação a longo-prazo para suportar recursos necessários à preservação dos ativos digitais, usando modelos de custo que raramente se adequam à organização, sendo este cálculo um dos principais pontos fracos da sua gestão. Os gastos alteram-se perante mudanças tecnológicas e sociopolíticas, que complexificam a otimização do orçamento disponível [Faria and Ferreira 2015]. Sendo uma atividade recente, ainda está longe de alcançar um consenso e um compromisso que ultrapasse a sua comunidade de prática. As primeiras tentativas surgiram nos anos 90 do século XX, não fornecendo cálculos de custo explícitos, mas começaram a avaliar e a determinar custos dos vários fatores associados às fases do ciclo de vida. Estes estudos deram origem, na década seguinte, ao modelo de referência *Open Archival Information System (OAIS)* e ao ciclo de vida *DCC*, tendo servido de base aos modelos de custo, que têm sido desenvolvidos e são abordados aprofundadamente por Mundet & Carrera (2015), Ferreira et al. (2014) e Keijser et al. (2014). No cenário português, verificou-se também o desenvolvimento de um modelo nacional pensado para uma estrutura partilhada de continuidade digital [Rodrigues et al. 2015].

O Modelo de Referência *OAIS*, surge em 2002 como recomendação do CCSDS, sendo publicada no ano seguinte como a Norma ISO 14721:2003. Em 2012, é publicada uma versão revista pela CCSDS, agora como Prática Recomendada, com o apoio de onze entidades-membro e entidades observadoras, e que dá origem à ISO 14721:2012. Fornece um conjunto de modelos terminologia completa que não se destina a substituir terminologias existentes, mas pretende manter o modelo num nível geral de abstração independente de qualquer contexto de aplicação específico, para torná-lo reutilizável e aplicável a todas as áreas, sectores e tipos de informação digital [Corujo 2015, 65]. Os modelos que inclui são: O Modelo de Ambiente Externo (composto por três intervenientes com os papéis de produtor, administrador ou gestor, e consumidor); O Modelo de Informação (inclui o objeto de dados, o objeto de informação, a informação de representação, a informação de descrição, a informação de conteúdo, a informação de descrição de preservação, a informação de empacotamento, e o pacote de informação, que pode ser um pacote de submissão de informação, pacote de informação de arquivo, ou pacote de disseminação de informação); e o Modelo Funcional (que inclui as entidades funcionais de ingestão, de armazenamento de arquivo, de gestão de dados, de acesso, de administração, de planeamento de preservação, e de serviços comuns, e ainda as interfaces que ligam estas entidades funcionais) [Corujo 2015, 66–85].

O Modelo de ciclo de vida de curadoria do *DCC* fornece uma visão gráfica, de alto nível das fases necessárias para curadoria e a preservação dos dados, desde a conceptualização inicial ou receção, ao longo do ciclo de curadoria iterativo. Este modelo inclui as ações executáveis durante todo o ciclo de vida, entre as quais a descrição e a representação de informação, a planificação de preservação, a observação e participação da comunidade, a curadoria e preservação; as ações sequenciais de conceção, criação ou receção, de avaliação e seleção, de ingestão, ação de preservação, armazenamento, acesso, utilização e reutilização, e transformação; as ações ocasionais de eliminação, reavaliação e migração.

Perceciona-se que a definição de curadoria digital com o qual este estudo se identifica é o fornecido pelo Projeto 4C: A curadoria digital envolve as atividades que compõem as várias entidades funcionais do Modelo de Referência *OAIS* (ingestão, gestão de dados, armazenamento (arquivístico), planeamento de preservação, acesso/disseminação, serviços comuns, administração de repositório), mas também atividades que podem ser consideradas expressões das funções *OAIS*, como atividades de pré-repositório, que incluem a criação de dados e digitalização e pré-ingestão (avaliação, seleção, preparação, gestão de direitos), e

atividades pós-repositório, por exemplo, uso de conjuntos de dados (*datasets*); e atividades de gestão, tais como disponibilização de orçamentos, políticas e análise de risco [4C Project 2013] [Kejser et al. 2014]. Esta escolha prende-se pelo facto de esta ser a definição considerada mais abrangente e versátil em termos das abordagens estudadas, na medida em que permite ter em conta diferentes realidades e contextos relativamente a diversas necessidades sociais das partes interessadas e as características dos objetos digitais e dos conteúdos dos dados que são alvo de curadoria digital. Tal escolha é ainda reforçada por se identificar diretamente com o Modelo de Referência OAIS, que vários estudos identificam como o enquadramento-base da maioria dos modelos existentes relevantes para a curadoria digital [Kejser et al. 2014]. Assim, esta definição, ao incorporar o planeamento de preservação dentro do conceito de curadoria digital, permite escorar a opção, no âmbito deste estudo, de identificar a preservação digital como um dos aspetos integrantes da curadoria digital.

Considera-se que a curadoria digital, na perspetiva em que é considerada por este trabalho, tem de ter em conta as necessidades das partes interessadas e as características dos objetos digitais e dos conteúdos dos dados, sendo estes conteúdos, dependendo do nível de granularidade da análise, um dos aspetos dos objetos digitais [Thibodeau 2002]. Deste modo, cada projeto de curadoria digital, e dentro dessa, a análise de custos, terá de ter esses elementos em consideração, independentemente do contexto do projeto. Por isso, considera-se não haver diferenças de nível de granularidade que justifiquem uma distinção do conceito de curadoria digital em termos temáticos, conforme o contexto. Podem verificar-se diferenças decorrentes das finalidades ou, eventualmente, de tipos de dados/informação, ou da envolvente sociotécnica, mas situam-se na mesma dimensão de estudo.

Qualquer registo digital constitui um objeto digital. Desta forma, e no âmbito do contexto utilizado como exemplo, assume-se que qualquer dado, independentemente da tipologia ou do contexto de origem, tem a potencialidade de ser tratado como objeto científico, ou matéria-prima de uma investigação. Assim, qualquer dado científico, ou com potencialidade para ser utilizado no âmbito de uma investigação científica, que seja potencialmente um alvo de curadoria digital, nomeadamente no que concerne aos custos, terá de ser também considerado em termos de objeto digital. Assim, o contexto de curadoria digital aqui apreendido refere-se somente a dados registados digitalmente, nomeadamente aqueles que estão registados em repositórios digitais, dos quais são exemplos os repositórios digitais com valências científicas-académicas. Usando o exemplo do contexto da investigação científica desenvolvida no âmbito das várias facetas e abordagens da *Big Science* e da *Big Data* (reitera-se não caber neste estudo a intenção de problematizar estes conceitos, que são somente usadas num contexto exemplificativo) têm beneficiado de apoio e financiamento para mitigar os custos da experimentação nesta área. No entanto, as agências de financiamento começam a solicitar compromissos das entidades de investigação para preservar e tornar acessível aquele conteúdo, muitas vezes constituído por grandes quantidades de dados a longo-prazo [Poole 2013]. Importa, assim, que as instituições antecipem e orçamentem os custos da investigação considerando as necessidades de curadoria digital [Walters and Skinner 2011]. Se preservar o digital é significativamente dispendioso [Rodrigues et al. 2015], para Mundet & Carrera (2015) ainda não é possível saber quanto custa, o que dificulta a obtenção de financiamento e o convencimento de decisores e gestores, habituados, que estão, a conhecer e a controlar custos. Embora constatem um crescente interesse, estes autores consideram os modelos de previsão de custos pouco desenvolvidos e prematuros, porque não se adequam à realidade, demonstrando estarmos ainda na infância da curadoria digital. Também, verificam a existência de uma consciência generalizada de que os custos não se podem isolar do meio em que se produzem, estando ligados aos interesses de terceiros, à estrutura organizacional e ao contexto cultural, suscitando questões relativas às perceções dos produtores, dos utilizadores e dos gestores de dados digitais, nomeadamente os investigadores, referentes ao seu entendimento do que é a curadoria digital e às suas perspetivas sobre os custos.

Tendo-se feito anteriormente um estudo [Corujo et al. 2016] com o objetivo de conhecer o pensamento existente referente aos seus custos e de desenvolver uma proposta de modelo de enquadramento para a análise de custos em curadoria digital, inclusivamente projetos de carácter mais prático, interessa agora, passados dois anos, verificar que investigação tem sido feita pela comunidade científica, e se tal investigação permite verificar se os estudos têm permitido sensibilizar as instituições com responsabilidade de produção, gestão e disseminação de dados científicos abertos, para a necessidade de tomar em consideração os custos no seio da curadoria digital. Tal implica abordar a definição da curadoria digital e a problemática dos custos, baseando-nos nos estudos referentes a modelos de custos.

A pesquisa (metodologia e dados)

Efetuiu-se uma pesquisa em 5 de janeiro de 2016 na Biblioteca do Conhecimento Online (B-On) e no Repositório Científico de Acesso Aberto de Portugal (RCAAP) com a pretensão de obter como resultados referências com os termos compostos *Digital Curation* e *Curadoria Digital*, que estivessem relacionados com palavras cujas raízes são “custo” e “cost”, utilizando-se os operadores booleanos OR e AND e a truncatura à direita. No caso do RCAAP, não se recorreu ao uso de parênteses na separação dos conjuntos de termos, por este agregador não interpretar condições de pesquisa que as contenham. Os resultados permitiram obter 11 e 36 referências, respetivamente, tendo-se verificado neste último a duplicação de um resultado, apresentado em dois repositórios diferentes. Destes foram usados, respetivamente, 9 e 30 estudos.

No dia 11 de março de 2018, repetiu-se a mesma pesquisa, adicionando uma delimitação das datas de publicação entre 2016 e 2018, com o objetivo de verificar a existência de novas referências desde a data da investigação anterior. A pesquisa no RCAAP retornou 0 (zero) resultados e a B-On 2 (dois) resultados, dos quais apenas um foi considerado para efeitos da análise.

20

Em seguida, desenvolveu-se uma proposta de sistematização delimitada pela recolha de dados e elaborada com base nas categorias recolhidas e identificadas por via do Método da Comparação Constante (MCC). O MCC é uma abordagem de análise qualitativa desenvolvida no seio da metodologia da Teoria Fundamentada, e que se foca na comparação de e entre todos os elementos nos dados, que pode ser identificado como um procedimento para interpretar textos, por intermédio da codificação e análise, com o fim de desenvolver teoria [Flick 2018] [Glaser 1965, 437]. Este método pode ser descrito em quatro fases: 1) a comparação de incidentes aplicáveis a cada categoria; 2) a integração das categorias e suas propriedades; 3) a delimitação da teoria; 4) a escrita da teoria [Flick 2018, 51–53] [Glaser 1965, 439–443]. A análise, que se efetuou aos textos recolhidos com base no MCC, foi desenvolvida por meio de um processo de “vai-vem” de codificação, identificação das categorias e saturação dos dados sobre os quais se fundamenta este trabalho, constituindo uma espiral progressiva, cujas reflexões permitiram verificar a relação de várias categorias com o Modelo de Referência *Open Archival Information System* (OAIS) e do ciclo de vida *Digital Curation Centre* (DCC), as quais fundamentaram a construção de um modelo de enquadramento esquematizado, que se considera pertinente para a análise de custos em curadoria digital, inclusivamente projetos de carácter mais prático.

21

O modelo construído permitiu desenvolver uma análise de conteúdo das perceções dos autores recolhidas, originando um conjunto de memorandos por categoria, que este estudo sintetiza. Procurou-se, particularmente, revelar uma dimensão invisível, como explica Bardin: «A partir do momento em que a análise de conteúdo decide codificar o seu material, deve produzir um sistema de categorias. A categorização tem como primeiro objetivo (da mesma maneira que a análise documental) fornecer, por condensação, uma representação simplificada dos dados brutos. (...) A análise de conteúdo assenta implicitamente na crença de que a categorização (passagem de dados brutos a dados organizados) não introduz desvios (por excesso ou por recusa) no material, mas que dá a conhecer índices invisíveis, ao nível dos dados brutos» [Bardin 2011, 148–149].

22

A Curadoria digital e os Custos: O Desenvolvimento do Modelo de Enquadramento de Análise

No âmbito dos trabalhos que resultaram da pesquisa no RCAAP e na B-On, Constantopoulos, et al. (2009) apresentam a curadoria digital como uma nova prática interdisciplinar, que se estende a vários domínios e pretende definir orientações para a gestão coerente da informação, bem como garantir a sua aptidão futura, à medida que evolui o seu contexto de utilização ligada a modelos de ciclo de vida. Compreende diferentes atores, tais como investigadores, conservadores e documentalistas, gestores de informação, académicos, educadores, curadores de exposições e o público em geral.

23

Dallas & Doorn (2009), Donnelly et al. (2010), Sayão & Sales (2012), e Santos (2014) consideram-na um conceito em evolução, que envolve a gestão ativa e a preservação dos dados científicos durante todo o ciclo de vida ligada à necessidade de assegurar a adequação epistémica e a reutilização futura pelas comunidades interessadas, constituindo um termo *guarda-chuva*, que integra ações de produção, organização, armazenamento, seleção, valorização, tratamento e preservação a longo-prazo de ativos digitais, diminuindo o risco da obsolescência. É uma nova área de práticas e de pesquisa interdisciplinar e de muitas áreas profissionais. Ayris (2009) e Whyte & Pryor (2011) centram a curadoria no âmbito do acesso, da partilha e da reutilização de dados científicos, e, ainda, como impulsionadora das políticas de dados. Bernardou et al. (2010) indicam que especialistas das humanidades desenvolvem e valorizam atividades de pesquisa de informação e investigação relacionadas com a curadoria de objetos de informação e produção académica, sendo, nesse sentido, curadores de dados académicos, por excelência. Ferreira et al. (2012) e Bachell & Barr (2014) aludem a ações ligadas à produção de metainformação contextual para garantir a inteligibilidade, o histórico e a reutilização dos dados a longo-prazo. Queiroz (2013) refere-a como uma atividade com vista à obtenção de melhores práticas em relação ao tratamento dos objetos digitais, estando dependente dos contextos tecnológico e de criação, que envolve fatores humanos, socioeconómicos, financeiros, entre outros. Bicarregui et al. (2013) focam a curadoria em termos de planeamento da gestão e da preservação de dados para a *Big Science* para uso continuado, Wilson & Jeffreys (2013) indicam que os repositórios especializados têm experiência consignada nas competências dos seus recursos humanos a nível de curadoria, enquanto Dillon (2013) e Strasser et al. (2014) referem a necessidade de formação referente à gestão de dados e à curadoria digital para os investigadores tomarem conhecimento das boas práticas.

24

Barros (2014), numa perspetiva ligada à linguística aplicada e ao jornalismo, indica que a curadoria digital tem a ver com seleção, organização, apresentação e evolução de conteúdos, sendo uma nova prática a ser assumida pelos profissionais dos *media* e utilizadores voluntários, apenas compreendida quando considerado o contexto cultural em que se insere. Neste sentido, o curador é fundamental para selecionar o conteúdo, através do *gatekeeping* e do *gatematching*, ambos ligados às fontes da comunicação social.

25

Faria & Ferreira (2015) ligam a curadoria digital ao desenvolvimento de repositórios para preservação e acesso atual e futuro, incluindo atividades de seleção, transferência, gestão de dados e armazenamento, preservação, acesso e serviços comuns e administrativos, sendo que Edmond & Garnet (2015) referem que esta prática exige diferentes ferramentas e estratégias para garantir a acessibilidade dos dados. Saraiva & Quaresma (2015), Machado et al. (2015) e Poole (2015) abordam o ponto de vista do ciclo de vida da curadoria e dos princípios e ações desenvolvidos para assegurar a sustentabilidade e a validade dos dados científicos de modo a que possam ser partilhados, acedidos e reutilizados, agrupados ou transformados no futuro, requerendo infraestruturas humanas, como ciberinfraestruturas, comunidades de investigação, colaboração, planificação, políticas, normas e boas-práticas, pelo que os fatores humano e social devem ser situados em contextos institucionais, que podem apoiar a comunicação, a coordenação e a colaboração no âmbito de financiamentos, revisão, preservação a longo-prazo, prática científica, recursos e infraestrutura. Apontam ainda questões prementes: a sustentabilidade, o custo, a política e o planeamento decorrentes da coordenação e da colaboração, da formação prática e teórica, das práticas quotidianas dos investigadores face aos dados e da sensibilização para a questão.

26

No entanto, mais recentemente, Lee et al. (2016) identificam como crenças a defesa de que as bibliotecas já fazem curadoria de dados há muito tempo, que os serviços de curadoria digital e as competências necessárias já são conhecidos, e que é essencial compreender e abordar as práticas de curadoria específicas do domínio em consideração. Estas crenças podem estar a contribuir para o atraso da comunidade do património cultural no desenvolvimento de serviços verdadeiramente inovadores e de programas de e em colaboração com investigação no panorama dos domínios em mutação, fruto das práticas da investigação interdisciplinar. Tal obriga a repensar a identidade profissional dos responsáveis pela recolha, pela organização e pela preservação dos dados para uso no futuro.

27

Ao mesmo tempo, Dallas (2016) chama a atenção para a necessidade de uma reconcetualização formal da curadoria digital, uma representação adequada do conhecimento dos seus objetos e investigação de práticas de curadoria, que sejam baseadas na experiência e no estabelecimento de infraestruturas digitais, que permitam a curadoria continuada. Esta abordagem propõe uma visão pragmática, que obriga a que a curadoria digital se estabeleça como uma área de investigação científica unificada e relevante para as práticas emergentes de curadoria em ambiente digital.

28

Ao procurar desenvolver a proposta de sistematização, foram identificados os critérios (as categorias resultantes do MCC) que se reconhecem com as fases do ciclo de vida *DCC* [DCC 2004] e com a entidades funcionais do Modelo de Referência *OAIS* [CCSDS 2002], e outras categorias, como os modelos de custo, a usabilidade do modelo, o modelo de cálculo de custo/benefício, o valor/rentabilidade, o detalhe da informação financeira do modelo/transparência, os próprios custos (ligados a aspetos como a infraestrutura, os recursos humanos, a sustentabilidade), o plano/política de gestão de dados, as questões legais, o processo de tomada de decisão e a política de gestão de preservação e acesso.

29

É de referir que a identificação com o ciclo de vida do *DCC* se prende com o facto de este abordar as questões na perspetiva do objeto da curadoria digital, enquanto que a identificação com o Modelo de Referência *OAIS* permite enformar o sistema que sustenta a curadoria digital desses objetos. As categorias identificadas com estes dois modelos são a produção de dados, a ingestão, a preservação (digital), o armazenamento, o acesso, o uso e a reutilização, e os próprios modelos do ciclo de vida *DCC* e de Referência *OAIS*. A categoria avaliação e seleção é somente identificada com o modelo de ciclo de vida *DCC*, e as categorias disseminação/divulgação e administração identificam-se com o Modelo de referência *OAIS*. A Tabela 1 permite identificar a origem de cada uma das categorias.

30

Com efeito, o Modelo de Referência *OAIS* é pertinente como enquadramento conceptual de uma resposta integrada do desafio da preservação digital (que, na perspetiva deste estudo, se considera elemento integrante da curadoria digital), já que fornece uma identificação dos atores externos, das entidades funcionais e das interfaces entre eles. Assim, por via do Modelo de Referência *OAIS*, é possível percecionar as interações entre os vários atores externos e a entidade custodiante. Neste sentido, os custos devem ser considerados, tendo em conta as necessidades e as expectativas dos atores externos, mas também a conceção, o desenvolvimento, a validação, o funcionamento e a avaliação das entidades funcionais e das interfaces entre eles. Adicionalmente, existe literatura que permite sustentar a permanência do modelo *OAIS* no âmbito de perceção de custos. A Biblioteca Real e o Arquivo Nacional da Dinamarca desenvolveram conjuntamente uma metodologia de análise de custos baseada nas atividades para preservação digital sustentada no Modelo de Referência *OAIS* [Kejser et al. 2011], que permite enquadrar um Modelo de Custo para Preservação Digital, focando-se na verificação dos custos da entidade funcional de planeamento de preservação das atividades do modelo *OAIS* e, também, da migração digital. Para estimar esses custos, os autores identificam atividades pertinentes, analisando as funções no modelo *OAIS* e os fluxos entre elas. Este Modelo de Custos para Preservação Digital é percecionado como pertinente para o estudo dos custos da Curadoria Digital pela própria iniciativa *Digital Curation Centre*, que o apresenta no seu sítio *web* como recurso [DCC 2014]. O próprio projeto europeu 4C (*Collaboration to Clarify the Costs of Curation*) [Kejser et al. 2014] que é notável que as instituições parecem ter considerado mais fácil desenvolver novos modelos de custo-

31

benefício do que modificá-los e reutilizá-los e que tal resultou num número relativamente grande de modelos diferentes. Embora os modelos se destinem a necessidades específicas, existem algumas semelhanças notórias com o Modelo de Referência OAIS, em que se baseia a maioria dos modelos existentes relevantes para a curadoria digital. Com efeito, o mesmo documento refere, no ponto 1.1, que por uma questão de clareza e normalização no campo da curadoria digital, o Projeto 4C optou pela aplicação do Modelo de Referência OAIS. Assim, define a curadoria digital como o conjunto de atividades que compõe as várias entidades funcionais do Modelo de Referência OAIS, mas também outras atividades que podem ser consideradas expressões das funções OAIS, designadamente de pré-repositório e pós-repositório, identificadas na definição de curadoria digital.

A sistematização das categorias anteriormente referidas deu origem a um modelo de enquadramento para análise de custos em curadoria digital, que propõe uma esquematização que se baseia na lógica do ciclo de vida do DCC e do Modelo de Referência OAIS. Assim, para além da Avaliação e da Seleção, da Disseminação/divulgação e da Administração, percecionam-se as seguintes ligações no âmbito dos dois modelos Produtor <-> Create or Receive; Ingestão <-> Ingest; Armazenamento de Arquivo <-> Store; Planeamento de Preservação <-> Preservation Planning; Acesso <-> Access, Use & Reuse. A estas adicionam-se critérios emanados do MCC, como elementos considerados transversais a estes modelos, em que se percebe que os Modelos de custo têm de ter em conta a usabilidade do modelo, o modelo de cálculo de custo/benefício (valor/rentabilidade), o detalhe da informação financeira do modelo/transparência, e os custos (infraestrutura, recursos humanos, sustentabilidade), e que o plano/política de gestão de dados deve ter em consideração as questões legais, o processo de tomada de decisão, e a política de gestão de preservação e acesso.

Esta proposta é original na medida em que promove uma perspetiva, que não se limita aos objetos digitais e ao seu ciclo de vida, nem aos sistemas com funções de curadoria, assumindo a necessidade de integrar ambos os prismas numa visão transversal apreendida pelos Modelos de Custos e Planos/Políticas de gestão dos dados (Figura 1).

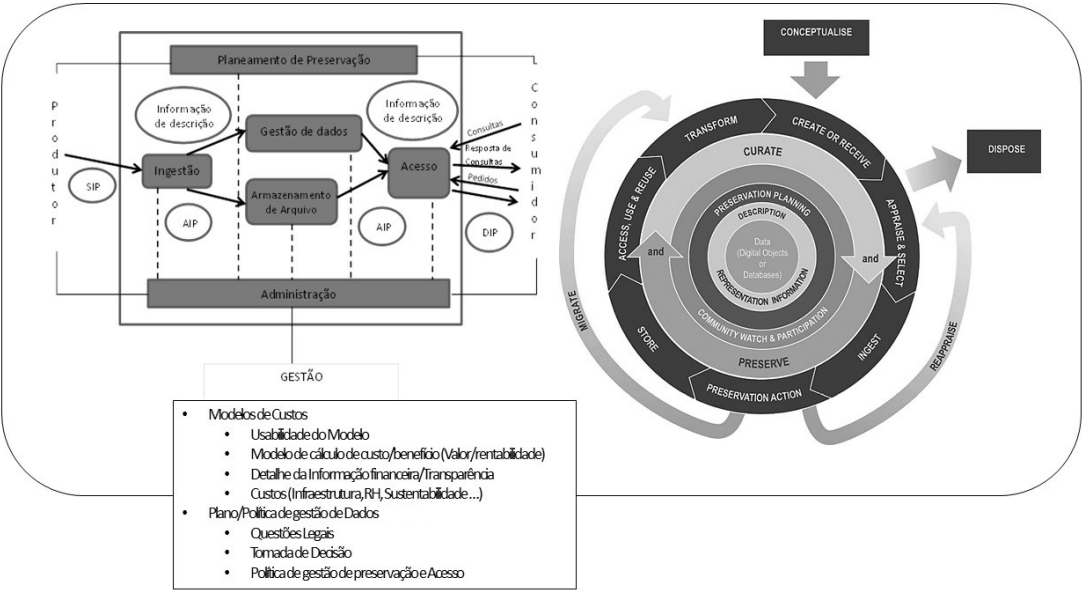


Figure 1. (Figura 1) Proposta de Sistematização das problemáticas no âmbito dos custos da curadoria digital. Por questões de organização de imagem, a ilustração não apresenta as ligações entre o OAIS e o Ciclo de Vida DCC: Produtor <-> Create or Receive; Ingestão <-> Ingest; Armazenamento de Arquivo <-> Store; Planeamento de Preservação <-> Preservation Planning; Acesso <-> Access, Use & Reuse.

A curadoria digital e os Custos na perspetiva dos estudos: Exemplo da Aplicação do Modelo de Enquadramento de Análise

Com o modelo de enquadramento estabilizado, procurou-se efetuar a sua validação, testando-o acerca das perceções expressas pelos dos autores recolhidas na literatura obtida, no sentido de encontrar as suas preocupações no âmbito dos custos relativos à curadoria digital. Esta análise está sistematizada na Tabela 2 (em apêndice), que exemplifica a aplicação do modelo de enquadramento de análise relativa aos custos no âmbito da curadoria digital, bem como as perceções dos autores citados no âmbito de cada categoria analisada.

Em termos de:

Produção de dados: Bicarregui et al. (2013) e Santos (2014) referem-se aos custos da investigação científica na *Big Science*, que lida com volumes de dados dificilmente esgotáveis em artigos publicados, apresentando problemas de gestão, que são qualitativamente diferentes dos de outras disciplinas, daí serem inevitavelmente projetos de larga escala, familiarizados com estimativas de custos. Para Lee et al. (2016), muitos planos ainda falham na abordagem dos custos relacionados com a criação de dados abertos.

36

Avaliação e Seleção: devem considerar o custo dos materiais, o tipo, o estado, a quantidade, a acessibilidade e a singularidade, assim como possibilidades de uso futuro [Poole 2015].

37

Ingestão: Os custos não devem ser descurados, tendo o controlo de qualidade um papel importante nesta fase [Dürr et al. 2008].

38

Preservação (Digital): Ayris (2009) refere a necessidade de um modelo genérico para identificar elementos-chave das atividades de preservação, fatores que contribuem para as despesas, de modo a poder identificar-se e a reduzir a concentração de custos em determinados momentos e a periodicidade dessas ações. Para Keijser et al. (2011), são custos recorrentes, dependentes da gama de serviços que uma instituição oferece, sendo difícil separá-los de outros custos do ciclo de vida, como a produção, o acesso e a divulgação. Para Bicarregui et al. (2013), muitos dos custos de curadoria vêm da preservação, pelo que da sua boa gestão depende o desaparecimento de um conjunto significativo de problemas práticos relativos ao desbloqueamento de dados. Dentro desta fase, incluem-se as estratégias de migração e emulação, tratadas por vários estudos.

39

Armazenamento: Wright et al. (2009) e Santos (2014) alertam para o modo como a constante queda do preço destes dispositivos promove a utilização de maior espaço de armazenamento, mas os custos de energia, espaço, refrigeração e gestão são crescentes, enquanto Whyte & Pryor (2011) integram estes nos custos de reutilização planeados por um repositório. Assim, Rice et al. (2013) indicam que o armazenamento deve cumprir normas abertas, ter escalabilidade e permitir a utilização de hardware de vários fornecedores, com mecanismos de acesso de dados flexíveis. Rosenthal & Vargas (2013), Suchodoletz et al. (2013) e Saraiva & Quaresma (2015) contrapõem o armazenamento local com serviços de computação em nuvem para preservação partilhada, mas enquanto os primeiros verificam que o armazenamento em nuvem não é competitivo a longo-prazo, os últimos consideram haver uma redução de custos de armazenamento e interoperabilidade com outros sistemas e serviços.

40

Acesso, uso e reutilização: Rusbridge & Ross (2007), Queiroz (2013) e Machado (2015) abordam o aumento dos custos das assinaturas de jornais científicos, estudo que serviu de incentivo ao surgimento do movimento de acesso aberto. Esta questão liga-se à lógica do acesso livre aos dados científicos (Open Science) em que Whyte & Pryor (2011) e Saraiva & Quaresma (2015) defendem poupanças na aquisição, no acesso e na gestão, e menos barreiras à participação e à colaboração ativa de pessoas externas a uma comunidade científica, de que é exemplo a lógica que parece estar por detrás da obrigação de disponibilizar gratuitamente os dados da investigação científica financiada publicamente, como forma de retorno para a sociedade pelo financiamento de público [Lee et al. 2016]. Bicarregui et al. (2013) defendem a preponderância da questão dos dados abertos relativamente à preservação, mas referem a necessidade de selecionar os dados a preservar. Evans & Moore (2014) e Poole (2015) realçam o papel da partilha e da reutilização de dados na diminuição de custos face à produção de novos dados numa época de austeridade económica, tendo o cuidado de permitir a reconstituição em formas e formatos diferentes do original. Sayão & Sales (2012) reiteram que o valor dos dados se relacione com reprodutibilidade da pesquisa que o originou para análise de custo-benefício relativa ao acesso e à capacidade de reutilização dos dados. Dürr et al. (2008) apresentam um sistema projetado para garantir a reutilização de dados primários e produzir dados secundários, similar aos apelos de Minor et al. (2010), para que assegurem a proteção de coleções de dados críticos para além do tempo de vida dos projetos e esforços que os geram. O uso de dados secundários tem consequências para a redução dos custos de recolha e duplicação de dados, a partilha dos custos diretos e indiretos da recolha, e os novos usos não previstos no momento da recolha (ex. mineração de dados), mas são portadores de custos em termos de armazenamento e preparação de dados para a curadoria digital e partilha [Whyte and Pryor 2011]. Rice et al. (2013) consideram que compete às instituições de investigação o desenvolvimento de infraestruturas que suportem a variedade de dados e a sua reutilização, principalmente quando, de acordo com Wilson & Jeffreys (2013), são dados de investigação financiada publicamente, daí que Delasalle (2013) indique como um dos desafios a superação do ceticismo entre os investigadores. Para Donnelly & North (2011), os centros de investigação e prestadores de serviços de informação devem orientar as políticas e as estratégias para as necessidades das comunidades de investigação para otimização da utilização e da troca de informação. Bicarregui et al. (2013) referem a seleção de produtos de um projeto, que representa para o cientista um compromisso que envolve a quantidade de tempo que pode investir na compreensão desses dados, o grau de apoio que recebe dos colegas e os proprietários de dados, e a subtilidade da questão que desejam tratar. Machado (2015) reporta-se aos menores custos de novas investigações desenvolvidas a partir dos dados científicos partilhados, algo que Whyte & Pryor (2011) identificam com a poupança nos custos indiretos de investigação através da utilização de recursos agrupados (evitando a fadiga da recolha), e da “troca de presentes” disponíveis para acelerar a investigação da ciência derivada, que originaram aumentos da eficiência e trouxeram uma mudança de patamar. Para Poole (2015), qualquer modelo que defenda a eficiência e a eficácia económica da partilha de dados

41

não espelha a complexidade das relações humanas imbricadas na sua produção. Os benefícios da colaboração incluem o acesso a uma ampla gama de experiências, partilha de custos e recursos, acesso a novas ferramentas, desenvolvimento de normas e boas-práticas e consciencialização para o problema, tendo Akers & Green (2014) defendido um papel de promoção da partilha e da preservação de dados científicos para as bibliotecas académicas. Whyte & Pryor (2011) baseiam-se nas Diretrizes da OCDE para defender o menor custo de acesso possível, de preferência não superior ao custo marginal de disseminação, defendendo que menores custos de acesso aos conhecimentos científicos dependem do financiamento público dos investigadores para aceder aos mesmos.

Disseminação/divulgação: Bicarregui et al. (2013) argumentam contra a divulgação generalizada de dados, por não ser gratuita e ter custos significativos, identificados por Whyte & Pryor (2011) como ligados à compreensão dos dados (preparação e documentação, usando esquemas de metainformação reconhecidos, custos de revisão), concluindo que a partilha deve aumentar o seu valor em vez de diminuir. No entanto, vários governos obrigam as instituições de investigação científica financiada pelo governo a disponibilizar os dados gratuitamente para acesso público [Lee et al. 2016]. Dürr et al. (2008) recomendam investigação sobre ferramentas de visualização para pesquisa e recuperação de conjuntos de dados, enquanto Wilson & Jeffreys (2013) versam os custos de divulgação de serviços de gestão de dados. Para Lee et al. (2016), muitos planos ainda apresentam insuficiências na abordagem aos custos de modo a garantir a recuperação dos dados.

42

Administração: A monitorização é assumida por Kejser et al. (2011) como vigilância da comunidade de interesse (custos variam inversamente consoante a influência da entidade curadora na produção e utilização de formatos) e tecnologia (custos dependem do desenvolvimento e complexidade). Por isso, Poole (2015) refere a necessidade de equilíbrio para que os custos de vigilância proactiva não sejam desproporcionados.

43

Modelo do ciclo de vida DCC & Modelo de referência OAIS: Vários estudos tecem comentários sobre estes modelos no âmbito de projetos dedicados aos modelos de custos relacionados com a gestão e a preservação de dados. Ambos os modelos referem que a sua simples leitura e implementação, embora aconselhável, não tem um carácter obrigatório na gestão e preservação de dados, a menos que se considere a validação, a auditoria e a modelagem de custos de acordo com estas especificações.

44

Modelos de custo: Alguns estudos apresentam projetos como o LIFE, o ESPIDA, o CMDP e o 4C. Para Currall et al. (2007), compreender e comunicar o custo e o valor das atividades de curadoria são atividades importantes para assegurar a sobrevivência a longo prazo de ativos digitais, mas existem problemas para exprimir compreensivelmente o seu valor a todas as partes interessadas, especialmente aos potenciais financiadores. Davies et al. (2007) concluem que os custos podem surgir em diferentes fases do ciclo de vida, ser recorrentes e com diferentes frequências. Wright et al. (2009) apresentam uma abordagem ao risco, que combina dimensões de custo, incerteza e benefício, considerando os modelos de custo e valor como parte da atividade mais ampla da modelação económica, que por sua vez é apenas parte da realização de preservação digital e acesso sustentáveis. Ayris (2009) relata uma série de recomendações, tais como a necessidade de se efetuar o cálculo de custos ajustados à inflação e de se considerarem os custos externos ao ciclo de vida. Bicarregui et al. (2013) referem que estes projetos não produziram consenso e que o acesso aos dados pode já não ser possível, dada a variedade de contextos de preservação. No caso dos dados não acedidos, há ainda menos apoio em termos de modelagem de custos.

45

Kilbride & Norris (2014) defendem que, para garantir que a modelação de custos não se torne um fim em si, aquela tem de abranger conceitos relacionados como *risco*, *valor*, *qualidade* e *sustentabilidade*. Estes autores afirmam, ainda, que a investigação da modelação de custos de curadoria tem sido muito ativa, tendendo a valorizar os custos de preservação, pela dificuldade em articular os benefícios da preservação e a complexidade da tarefa, algo que Evans & Moore (2014) exemplificam com o facto de uma decisão aparentemente simples, como o formato do ficheiro, ter consequências na preservação e na reutilização. Deste modo, concluem que, embora os modelos de custos afirmem ser genéricos, tendem na prática a ser específicos de determinadas instituições, ao mesmo tempo que as entidades protegem as informações acerca dos custos considerados sensíveis. Outras lacunas identificadas têm a ver com a má usabilidade e a falta de consenso sobre a forma de estruturar os dados de custo. Faria & Ferreira (2012) consideram que este domínio evoluiu bastante derivado de uma maior compreensão dos custos e de que estes modelos permitem que os processos de decisão estratégica sejam mais eficientes e precisos, embora não exista um modelo único aplicável a todos os casos.

46

Usabilidade do Modelo: Kejser et al. (2011) defendem que os modelos de custos são imprecisos para estimar o custo futuro devido aos desafios colocados pela manipulação do elemento de previsão, que influencia vários aspetos do modelo. Paralelamente, Kilbride & Norris (2014) apontam a pouca absorção das ferramentas e dos métodos desenvolvidos com estes modelos.

47

Modelo de cálculo de custo/benefício: Currall et al. (2007) apontam a necessidade de melhor informação sobre custos e benefícios para aperfeiçoar o processo de tomada de decisão por parte dos gestores e investidores. Estes pretendem tomar conhecimento dos potenciais benefícios que a gestão cuidada da informação digital pode trazer para os clientes externos da organização (talvez

48

através de um maior acesso) e saber se melhora o desempenho dos processos da instituição, se ajuda a desenvolver o seu negócio e a expandir o seu conhecimento, a noção do impacto que terá sobre as suas finanças, e se o custo das ações necessárias é menor ou maior que o valor de uma coima por não-conformidade. Bicarregui et al. (2013) consideram os custos razoavelmente objetivos, e embora seja difícil estimá-los para lá de uma certa ordem de magnitude, eles podem ser estimados. Em contrapartida, a valorização é muitas vezes difusa, envolvendo benefícios educativos e de assistência, que são reais, mas apenas concretizáveis numa análise formal de custo-benefício. Kilbride & Norris (2014) consideram necessário o fornecimento de modelos de custo-benefício claros, incluindo a sua descrição conceptual e vocabulário normalizado.

Valor/rentabilidade: O valor foi perspectivado por Currall et al. (2007), a nível de geração de renda através de venda de ativos, licenciamento e/ou direitos de bens, ensino e de investigação, contratos, subvenções, taxas, doações, bem como a nível de redução de custos de trabalho, tempo espaço e despesas diretas. Evans & Moore (2014) concluem que a avaliação do valor e do impacto dos dados tornaram-se cada vez mais importantes num clima de recessão e de diminuição da atividade económica.

49

Detalhe da informação Financeira do Modelo/transparência: A informação financeira é referida por Kilbride & Norris (2014) que discutem se os modelos de custo-benefício atuais vão ao encontro das necessidades dos interessados no cálculo e na comparação de informação financeira, enquanto Ayris (2009) refere a inadequação dos modelos de financiamento a termo (bolsas de investigação ou contratos) para responder às necessidades de acesso e preservação a longo-prazo.

50

Custos: Os estudos identificam custos diretos fixos, variáveis e indiretos, tendo Wilson & Jeffreys (2013) notado que os serviços institucionais têm, muitas vezes, custos fixos elevados relativamente aos custos variáveis (necessidade de desenvolvimento constante e recursos humanos), enquanto Whyte & Pryor (2011) defendem que a redução de custos indiretos requer o uso eficiente de recursos escassos na recolha de dados, incluindo temas de investigação e instrumentação. A eficiência também poderia ser atingida pela cooperação interinstitucional como forma de reduzir os custos de conformidade [Lee et al. 2016]. Assim, Kejser et al. (2011) defendem a ideia de que os modelos devem representar custos económicos completos, quer diretos quer indiretos. Dentro destas categorias de custos, os estudos abordam os custos de descrição ligados à metainformação, os custos crescentes de processamento e os custos associados a aquisições em geral e a recursos humanos.

51

Infraestrutura: A infraestrutura inclui plataformas, componentes informáticos/escalabilidade, sendo abordada na maioria dos estudos. Wright et al. (2009) chamam a atenção que os modelos *Total Cost of Ownership* das instituições são muitas vezes modelados erradamente como um valor anual. Obviamente, tais custos são substanciais a nível institucional, sendo preciso ter em conta a dimensão do serviço para assegurar que responde às necessidades. Suchodoletz et al. (2013) referem a alternativa de usar uma nuvem pública, para as instituições evitarem a manutenção de servidores dispendiosos e subutilizados.

52

Sustentabilidade: Ayris (2009) identifica um conjunto de elementos-chave para alcançar a sustentabilidade das coleções digitais a longo-prazo, enquanto Poole (2015) considera que as instituições tendem a dedicar uma ínfima parte dos seus orçamentos à curadoria e que o financiamento não acompanha o crescimento dos dados, pelo que os planos devem demonstrar liderança empresarial, propostas de valor preciso, minimização de custos e exploração de fontes de receitas variadas, assim como compromissos mensuráveis para prestação de contas. A nível de planos de financiamento, Ayris (2009) constata a falta de incentivos para impor modelos económicos sustentáveis, pelo que o financiamento para a curadoria digital continua a basear-se em projetos de curto prazo e que, embora os financiadores identifiquem os resultados sustentáveis, muitas vezes negligenciam a definição de os resultados acordados serem alcançados ou medidos [Poole 2015]. De entre os estudos que se debruçam sobre modelos de gestão e negócio, Wilson & Jeffreys (2013) concluem que os modelos de negócios são melhor produzidos ao nível dos componentes individuais, do que em termos da totalidade da infraestrutura. Poole (2015) afirma que o desenvolvimento de modelos de negócio, bem como a colaboração através de redes de preservação constituem preocupações cruciais no cálculo de custos, estando intimamente ligados à questão da partilha de custos, como exemplificam Suchodoletz et al. (2013). Lee et al. (2016) mencionam a partilha de custos pela cooperação interinstitucional de modo a poupar no investimento feito no âmbito da conformidade legal. Wilson & Jeffreys (2013) introduzem uma questão pouco desenvolvida e que se reflete na utilidade da avaliação do potencial de absorção de um serviço pelo público, relativamente à capacidade de planeamento e de desenvolvimento do modelo de negócio. Wright et al. (2009) abordam a modelação de custos de riscos ligados à incerteza, à imprevisibilidade e à ameaça de ocorrência, que, para Ayris (2009), podem ser considerados em planos e orçamentos de contingência, reservas estratégicas ou seguros para mitigar as perdas. A explicitação dos riscos permite o desenvolvimento de uma abordagem específica, que justapõe risco com custos e benefícios [Poole 2015]. Mas, para Lee et al. (2016), ainda existem várias instituições de investigação científica que não têm os custos em consideração nos seus planos para disponibilizar os dados de forma sustentável.

53

Plano/Política de gestão de dados científicos: Lee et al. (2016) indicam que algumas instituições de investigação científica continuam a mencionar de forma breve ou até a silenciar nos seus planos os custos financeiros e administrativos. Poucas instituições sugerem que os investigadores deveriam consagrar os custos da gestão de dados nos seus orçamentos. Ayris (2009)

54

advoga a escolha da dimensão adequada das atividades de preservação: precisam de ser suficientemente grandes para conseguirem responder aos desafios, mas não tão grandes de modo a que o seu fracasso possa ser catastrófico. Ayrís considera não haver substituto para uma organização flexível, empenhada e dedicada a preservar um *corpus* de material. Cada organização deve ser capaz de fazer escolhas e traçar estratégias para lidar com problemas inesperados de todos os tipos. Se necessário, pode realizar triagens e estabelecer compromissos. Donnelly & North (2011) observam que a existência de uma abordagem única para o futuro das ciências ou uma política de dados não específica não será produtiva e eficaz, e que as políticas e a prestação de serviços de informação precisam de se aproximar das comunidades de investigação. Wilson & Jeffreys (2013) afirmam que os investigadores não estão expostos aos custos reais das soluções de gestão de dados (frequentemente subaproveitados), ao passo que, quando estes serviços estão centralizados, os custos totais são visíveis e desanimadores, pelo que Delasalle (2013) defende que os planos de gestão de dados devem corresponder às expectativas dos cientistas e à sua área de estudo e, também, contemplar os custos associados aos planos do projeto. Rice et al. (2013) consideram que um serviço gratuito no momento da utilização elimina uma das principais barreiras para a boa curadoria e permite que os dados possam ser guardados na infraestrutura adequada, em vez de uma infraestrutura escolhida em função do menor custo.

Questões legais: Abordadas as questões legais por alguns estudos, as suas exigências devem ser equilibradas pelas partes interessadas, no âmbito da política [Poole 2015]. Podemos incluir aqui os direitos de propriedade intelectual em termos de videojogos [Bachell and Barr 2014] e conteúdo jornalístico [Barros 2014]. Lee et al. (2016) referem a existência de obrigações legais por parte das instituições de investigação científica financiada pelo governo dos EUA, que exigem a disponibilização gratuita dos dados em acesso público, que devem ser detetáveis e utilizáveis.

Tomada de decisão: Whyte & Pryor (2011) identificam ainda deliberações para os decisores políticos, investigadores e comunidades de utilizadores sobre os riscos, os custos e os benefícios de uma participação mais ampla, antes e depois dos resultados serem produzidos, enquanto Delasalle (2013) defende o envolvimento dos investigadores no processo de tomada de decisão sobre o que é mantido, onde e por quanto tempo.

Política de gestão, de preservação e acesso: Os estudos tecem considerações sobre a política de gestão, de preservação e acesso intimamente ligadas à certificação, que para Bicarregui et al. (2013) deve ser apoiada por um conjunto de ferramentas metodológicas, incluindo descritivos, estudos de caso e modelos de custo para fornecerem orientações sobre o desenvolvimento das melhores práticas em política de gestão e preservação de dados e infraestrutura para esses projetos.

Conclusão

Este estudo permitiu apresentar uma proposta de sistematização original das problemáticas, no âmbito dos custos da curadoria digital, e uma visão que não se limita aos objetos digitais e ao seu ciclo de vida, nem aos sistemas com funções de curadoria, na medida em que assume a necessidade de integrar ambos os prismas numa perspetiva dotada de transversalidade apreendida pelos modelos de custos e planos/políticas de gestão dos dados. Embora esta sistematização seja delimitada pela recolha de dados, constitui um modelo de enquadramento, que se considera pertinente para a análise de custos em curadoria digital, inclusivamente projetos de carácter mais prático.

Assume-se que esta proposta requer mais contributos para o seu desenvolvimento e aprimoramento. Porém, com base no modelo de enquadramento esquematizado, desenvolveu-se uma análise das perceções dos autores, que permitiu identificar a existência de uma apropriação do conceito de curadoria digital de dados científicos na perspetiva de ciclo de vida, enfatizando as questões da preservação digital, questão difícil de separar dos demais elementos deste ciclo. Este posicionamento reflete-se na modelação de custos, apesar de faltar um modelo funcional bem definido e um consenso sobre os princípios contabilísticos aceites. Os custos sempre foram abordados, mas considera-se que se assiste a uma mudança de paradigma, de uma visão de black-box para uma abordagem de identificação dos custos e de tentativa de sistematização de modelos de previsão para a sua utilização institucional, como forma de incentivar a transparência e a accountability e de captar o interesse de potenciais financiadores. Este esforço de clarificação ganhou maior ímpeto com a recente crise financeira mundial.

Os estudos revelam a consciência generalizada de que os custos não podem ser isolados do meio em que ocorrem, inserindo-se num quadro mais amplo, que inclui os interesses de terceiros, a estrutura organizacional e o ambiente cultural. No entanto, esta abordagem ainda tem de ser aprimorada, porque os modelos contêm muitas lacunas. Por exemplo, a investigação internacional relativa à curadoria digital é acompanhada no Brasil, principalmente sob a forma de dissertações académicas, enquanto em Portugal se revela pela participação ativa em equipas internacionais de investigação e desenvolvimento de modelação de custos nesta área de estudo.

Também se verifica que, entre o primeiro estudo publicado por Corujo, Silva & Revez (2016) e o que se apresenta agora, por um lado, a urgência do assunto parece não ter resultado em mais produção académica-científica e, por outro, as instituições continuam

a não apresentar uma progressão na adoção de práticas, que tenham em vista abordar a questão dos custos nos seus planos para a criação, a manutenção e a disseminação dos dados científicos abertos.

Apêndice

Origem da Categoria	Designação da Categoria
OA/IS e o Ciclo de Vida DCC	Produção de dados
Ciclo de Vida DCC	Avaliação e Seleção
OA/IS e o Ciclo de Vida DCC	Ingestão
OA/IS e o Ciclo de Vida DCC	Preservação (Digital)
OA/IS e o Ciclo de Vida DCC	Armazenamento
OA/IS e o Ciclo de Vida DCC	Acesso, uso e reutilização
OA/IS	Disseminação/divulgação
OA/IS	Administração
Modelo do ciclo de vida DCC & Modelo de referência OA/IS	Modelo do ciclo de vida DCC & Modelo de referência OA/IS
Análise própria	Modelos de custo
Análise própria	Usabilidade do Modelo
Análise própria	Modelo de cálculo de custo/benefício
Análise própria	Valor/rentabilidade
Análise própria	Detalhe da informação Financeira do Modelo/transparência
Análise própria	Custos
Análise própria	Infraestrutura
Análise própria	Sustentabilidade
Análise própria	Plano/Política de gestão de Dados
Análise própria	Questões legais
Análise própria	Tomada de decisão
Análise própria	Política de gestão de preservação e acesso

Table 1. (Tabela 1) Relação das categorias do modelo de enquadramento com a origem das mesmas FONTE: Produção própria, com base em Akers & Green (2014); Ayris (2009); Bachell & Barr (2014); Barros (2014); Bicarregui et al. (2013); CCSDS (2002); Corujo, Silva & Revez (2016); Currall et al. (2007); Davies et al. (2007); DCC (2004); Delasalle (2013); Donnelly & North (2011); Dürr et al. (2008); Evans & Moore (2014); Faria & Ferreira (2012); Kejser et al. (2011); Kilbride & Norris (2014); Lee et al. (2016); Machado (2015); Minor et al. (2010); Poole (2015); Queiroz (2013); Rice et al. (2013); Rosenthal & Vargas (2013); Rusbridge & Ross (2007); Santos (2014); Saraiva & Quaresma (2015); Sayão & Sales (2012); Suchodoletz et al. (2013); Whyte & Pryor (2011); Wilson & Jeffreys (2013); Wright et al. (2009)

Designação da Categoria	Autores	Comentários e ideias expressas pelos autores
Produção de dados	Bicarregui et al. (2013) e Santos (2014)	Referem-se aos custos da investigação científica na <i>Big Science</i> , que lida com volumes de dados dificilmente esgotáveis em artigos publicados, apresentando problemas de gestão, que são qualitativamente diferentes de outras disciplinas, daí serem inevitavelmente projetos de larga escala, familiarizados com estimativas de custos.
	Lee et al. (2016)	Detetam que muitos planos ainda falham na abordagem aos custos relacionados com a criação de dados abertos.
Avaliação e Seleção	Poole (2015)	Defende que se considerar o custo dos materiais, o tipo, o estado, a quantidade, a acessibilidade e a singularidade, e possibilidades de uso futuro
Ingestão	Dürr et al. (2008)	Referem a importância do controlo de qualidade.
Preservação (Digital)	Ayris (2009)	Aponta para a necessidade de um modelo genérico para identificar elementos-chave das atividades de preservação, fatores que contribuíram para as despesas, para identificação e redução de momentos de concentração de custo e a periodicidade dessas ações.
	Kejser et al. (2011)	Defendem que os custos de preservação são custos recorrentes, dependentes da gama de serviços que uma instituição oferece, sendo difícil separá-los dos outros custos de ciclo de vida.

	Bicarregui, et al. (2013)	Mencionam que muitos dos custos de curadoria vêm da preservação, pelo que da sua boa gestão depende o desaparecimento de um conjunto significativo de problemas práticos relativos ao desbloqueamento de dados. Dentro de esta fase, incluem-se as estratégias de migração e emulação.
Armazenamento	Wright et al. (2009) e Santos (2014)	Identificam que a constante queda do preço destes dispositivos promove a utilização de maior espaço de armazenamento, mas os custos de energia, espaço, refrigeração e gestão são crescentes.
	Whyte & Pryor (2011)	Integram-nos nos custos de reutilização encarados por um repositório.
	Rice et al. (2013)	Advogam que o armazenamento deve cumprir normas abertas, ter escalabilidade e permitir utilizar <i>hardware</i> de vários vendedores, com mecanismos de acesso de dados flexíveis.
	Rosenthal & Vargas (2013), Suchodoletz et al. (2013) e Saraiva & Quaresma (2015)	Contrapõem o armazenamento local com serviços de computação em nuvem para preservação partilhada.
Acesso, uso e reutilização	Rusbridge & Ross (2007), Queiroz (2013) e Machado (2015)	Concluem que o aumento dos custos das assinaturas de jornais científicos serviu de incentivo ao surgimento do movimento de acesso aberto.
	Whyte & Pryor (2011) e Saraiva & Quaresma (2015)	Referem que o acesso livre aos dados científicos (<i>Open Science</i>) traz poupanças na aquisição, no acesso e na gestão, e menos barreiras à participação e à colaboração ativa de pessoas de fora numa comunidade científica.
	Lee et al. (2016)	Referem que a lógica que parece estar por detrás da obrigação de disponibilizar gratuitamente os dados da investigação científica financiada publicamente, refere-se ao retorno feito à Sociedade do investimento feito com os dinheiros públicos.
	Bicarregui et al. (2013)	Defendem a preponderância da questão dos dados abertos relativamente à preservação, assim como a necessidade de selecionar os dados a preservar.
	Evans & Moore (2014) e Poole (2015)	Realçam o papel da partilha e da reutilização de dados na diminuição de custos face à produção de novos dados numa época de austeridade económica, tendo o cuidado de permitir a reconstituição em formas e formatos diferentes do original.
	Sayão & Sales (2012)	Mencionam que o valor dos dados relaciona-se com reprodutibilidade da pesquisa que o originou para análise de custo-benefício relativa ao acesso e à capacidade de reutilização dos dados.
	Dürr et al. (2008)	Apresentam um sistema projetado para garantir a reutilização de dados primários e produzir dados secundários.
	Minor et al. (2010)	Apelam para que se assegure a proteção de coleções de dados críticos para além do tempo de vida dos projetos e esforços que os geram.
	Whyte & Pryor (2011)	O uso de dados secundários tem consequências para a redução dos custos de colheita e duplicação de dados, a partilha dos custos diretos e indiretos da colheita, e os novos usos não previstos no momento da colheita (ex. Mineração de dados), mas são portadores de custos em termos de armazenamento e preparação de dados para a curadoria e partilha.
	Rice et al. (2013), Wilson & Jeffreys (2013), Delasalle (2013)	Compete às instituições de investigação que desenvolvam infraestruturas que suportem a variedade de dados e a sua reutilização, por serem dados de investigação financiada publicamente, sendo este um dos desafios a superação do ceticismo entre os investigadores.

Acesso, uso e reutilização (continuação)	Donnelly & North (2011)	Defendem que os centros de investigação e prestadores de serviços de informação devem orientar as políticas e as estratégias para as necessidades das comunidades de investigação para otimização da utilização e troca de informação.
	Bicarregui et al. (2013)	Referem a escolha de produtos de um projeto, que representa para o cientista um compromisso que envolve a quantidade de tempo que pode investir na compreensão desses dados, o grau de apoio que recebe dos colegas e os proprietários de dados, e a subtilidade da questão que desejam tratar.
	Machado (2015)	Reporta-se aos menores custos de novas investigações desenvolvidas a partir dos dados científicos partilhados.
	Whyte & Pryor (2011)	Identificam os dados científicos partilhados com a poupança nos custos indiretos de investigação através da utilização de recursos agrupados (evitando a fadiga da colheita) e a “troca de presentes” disponíveis para acelerar a investigação da ciência derivada, que originaram aumentos da eficiência e trouxeram uma mudança de patamar.
	Poole (2015)	Defende que qualquer modelo que defenda a eficiência e a eficácia económica da partilha de dados não espelha a complexidade das relações humanas imbricadas na sua produção.
	Akers & Green (2014)	Referem que os benefícios da colaboração incluem o acesso a uma ampla gama de experiências, partilha de custos e recursos, acesso a novas ferramentas, desenvolvimento de normas e boas-práticas e consciencialização para o problema, tendo as bibliotecas académicas um papel de promoção da partilha e da preservação de dados científicos.
	Whyte & Pryor (2011)	Baseiam-se nas Diretrizes da OCDE para defender o menor custo de acesso possível, de preferência não superior ao custo marginal de disseminação, indicando que menores custos de acesso aos conhecimentos científicos anteriores dependem do financiamento público dos investigadores para o aceder.
Disseminação/divulgação	Bicarregui et al. (2013)	Argumentam contra a divulgação generalizada de dados, por não ser gratuita e ter custos significativos.
	Whyte & Pryor (2011)	Concluem que, com a partilha, devem aumentar os custos ligados à compreensão dos dados (preparação e documentação, usando esquemas de metainformação reconhecidos, custos de revisão), em vez de diminuir.
	Lee et al. (2016)	Mencionam que vários governos obrigam as instituições de investigação científica financiada por verbas públicas a disponibilizar os dados gratuitamente para acesso público.
	Dürr et al. (2008)	Recomendam investigação sobre ferramentas de visualização para pesquisa e recuperação de conjuntos de dados.
	Wilson & Jeffreys (2013)	Abordam os custos de divulgação de serviços de gestão de dados.
	Lee et al. (2016)	Afirmam que muitos planos ainda falham de forma significativa na abordagem aos custos para garantir a recuperação dos dados.
Administração	Kejser et al. (2011)	Assumem a monitorização como a vigilância da comunidade de interesse (custos variam inversamente consoante a influência da entidade curadora na produção e utilização de formatos) e tecnologia (custos dependem do desenvolvimento e complexidade).
	Poole (2015)	Refere a necessidade de equilíbrio para que os custos de vigilância proactiva não sejam desproporcionados.
Modelo do ciclo de vida DCC & Modelo de referência OA/IS	A generalidade dos autores	Tece comentários sobre estes modelos no âmbito de projetos dedicados aos modelos de custos da gestão e preservação de dados. Refere que a sua simples leitura e implementação, embora aconselhável, não tem um carácter obrigatório na gestão e na preservação de dados, a menos que se considere a validação, a auditoria e a modelagem de custos de acordo com estas especificações.
Modelos de custo	Currall et al. (2007)	Compreender e comunicar o custo e o valor das atividades de curadoria são fatores importantes para assegurar a sobrevivência de ativos digitais a longo-prazo, mas existem problemas para exprimir compreensivelmente o seu valor a todas as partes interessadas, especialmente aos potenciais financiadores.
	Davies et al.	Concluem que os custos podem surgir em diferentes fases do ciclo de vida,

	(2007)	ser recorrentes e com diferentes frequências.
	Wright et al. (2009)	Apresentam uma abordagem ao risco, que combina dimensões de custo, incerteza e benefício, considerando os modelos de custo e valor como uma parte da atividade mais ampla da modelação económica, que por sua vez é apenas parte da realização de preservação digital e do acesso sustentáveis.
	Ayris (2009)	Efetua diversas recomendações, tais como a necessidade de efetuar o cálculo de custos ajustados à inflação e de considerar os custos externos ao ciclo de vida.
	Bicarregui et al. (2013)	Referem que os projetos de modelos de custo não produziram consenso e que tal pode nem ser possível dada a variedade de contextos de preservação. No caso dos dados não acedidos, há ainda menos apoio em termos de modelagem de custos.
	Kilbride & Norris (2014)	Defendem que, para garantir que a modelação de custos, esta não se torne um fim em si, tendo de abranger conceitos relacionados como <i>risco</i> , <i>valor</i> , <i>qualidade</i> e <i>sustentabilidade</i> . Afirmam que a investigação da modelação de custos de curadoria tem sido muito ativa, tendendo a valorizar os custos de preservação, pela dificuldade em articular os benefícios da preservação e a complexidade da tarefa.
Modelos de custo (continuação)	62 63 Evans & Moore (2014)	Exemplificam a complexidade da tarefa com o facto de uma decisão aparentemente simples, como formato do ficheiro, ter sérias consequências na preservação e na reutilização. Concluem que, embora os modelos de custos afirmem ser genéricos, tendem na prática a ser específicos de determinadas instituições, ao mesmo tempo que as entidades protegem justamente as informações acerca dos custos considerados sensíveis. Outras lacunas identificadas foram a má usabilidade e a falta de consenso sobre a forma de estruturar os dados de custo.
	Faria & Ferreira (2012)	Consideram que este domínio evoluiu bastante derivado de uma maior compreensão dos custos e do facto de estes modelos permitirem que os processos de decisão estratégica sejam mais eficientes e precisos, embora não exista um modelo único aplicável a todos os casos.
Usabilidade do Modelo	Kejser et al. (2011)	Defendem que os modelos de custos são imprecisos para estimar o custo futuro devido aos desafios colocados pela manipulação do elemento de previsão, que influencia vários aspetos do modelo.
	Kilbride & Norris (2014)	Apontam para a pouca absorção das ferramentas e dos métodos desenvolvidos com os modelos de custos.
Modelo de cálculo de custo/benefício	Currall et al. (2007)	Apontam a necessidade de melhor informação sobre custos e benefícios para aperfeiçoar o processo de tomada de decisão por parte dos gestores e investidores. Estes pretendem tomar conhecimento dos potenciais benefícios que a gestão cuidada da informação digital pode trazer para os clientes externos da organização (talvez através de um maior acesso), e se melhorará o desempenho dos processos da instituição, ou se ajudará a desenvolver o seu negócio e a expandir seu conhecimento, assim como a ter noção do impacto que terá sobre as suas finanças, e se o custo das ações necessárias é menor ou maior que o valor de uma coima por não-conformidade.
	Bicarregui et al. (2013)	Consideram os custos razoavelmente objetivos, e embora seja difícil estimá-los para lá de uma certa ordem de magnitude, podem ser estimados. Em contrapartida, a valorização é, muitas vezes, difusa, envolvendo benefícios educativos e de assistência, que são reais, mas apenas concretizáveis em uma análise formal de custo-benefício.
	Kilbride & Norris (2014)	Consideram necessário o fornecimento de modelos de custo-benefício claros, incluindo a sua descrição conceptual e vocabulário normalizado.
Valor/rentabilidade	Currall et al. (2007)	Perspetivam o valor a nível de geração de renda através de venda de ativos, licenciamento e/ou direitos de bens, ensino e de investigação, contratos, subvenções, taxas, doações, e a nível de redução de custos de trabalho, tempo espaço e despesas diretas.
	Evans & Moore (2014)	Concluem que a avaliação do valor e do impacto dos dados tornaram-se cada vez mais importantes num clima de recessão e de diminuição da atividade económica.
Detalhe da informação Financeira do Modelo/transparência	Kilbride & Norris (2014)	Discutem se os modelos de custo-benefício atuais vão ao encontro das necessidades dos interessados no cálculo e na comparação de informação financeira.

		Ayris (2009)	Refere a inadequação dos modelos de financiamento a termo (bolsas de investigação ou contratos) de modo a responder às necessidades de acesso e preservação a longo-prazo.	
Custos		Wilson & Jeffreys (2013)	Referem que os serviços institucionais têm muitas vezes custos fixos elevados relativamente aos custos variáveis (necessidade de desenvolvimento constante e recursos humanos).	
		Whyte & Pryor (2011)	Defendem que a redução de custos indiretos requer o uso eficiente de recursos escassos na recolha de dados, incluindo temas de investigação e instrumentação.	
		Lee et al. (2016)	Defendem que a eficiência também poderia ser atingida pela cooperação interinstitucional como forma de reduzir os custos de conformidade.	
		Kejser et al. (2011)	Defendem a ideia de que os modelos devem representar custos económicos completos, sejam eles diretos ou indiretos. Dentro de estas categorias de custos, os estudos abordam os custos de descrição ligados à metainformação, os custos crescentes de processamento e os custos associados a aquisições em geral e a recursos humanos.	
Infraestrutura		Wright et al. (2009)	Referem que os modelos <i>Total Cost of Ownership</i> das instituições são muitas vezes modelados erradamente como um valor anual. Tais custos são substanciais a nível institucional, sendo preciso ter em conta a dimensão do serviço para assegurar que responde às necessidades.	
		Suchodoletz et al. (2013)	Referem a alternativa de usar uma nuvem pública, de modo a que as instituições evitem a manutenção de servidores dispendiosos e subutilizados.	
Sustentabilidade		Ayris (2009)	Identifica um conjunto de elementos-chave para alcançar a sustentabilidade das coleções digitais a longo-prazo.	64
			Constata, a nível de planos de financiamento, a falta de incentivos para impor modelos económicos sustentáveis, pelo que o financiamento da curadoria digital continua a basear-se em projetos de curto prazo.	65
			Defende que, em termos de custos de risco, a incerteza, a imprevisibilidade e a ameaça de ocorrência podem ser tratadas por planos e orçamentos de contingência, reservas estratégicas ou seguros para mitigar as perdas.	66
Sustentabilidade (continuação)	67	Poole (2015)	Considera que as instituições tendem a dedicar uma ínfima parte dos seus orçamentos à curadoria e que o financiamento não acompanha o crescimento dos dados, pelo que os planos devem demonstrar liderança empresarial, propostas de valor preciso, minimização de custos e exploração de fontes de receitas variadas, assim como compromissos mensuráveis para prestação de contas.	68
			A nível de planos de financiamento constata que embora os financiadores identifiquem os resultados sustentáveis, muitas vezes negligenciam a definição de os resultados acordados serem alcançados ou medidos.	69
			Afirma que o desenvolvimento de modelos de negócio, bem como a colaboração através de redes de preservação constituem preocupações cruciais no cálculo de custos.	70
			A explicitação dos riscos permite o desenvolvimento de uma abordagem que justapõe risco com custos e benefícios.	71
		Wilson & Jeffreys (2013)	Concluem que os modelos de negócios são melhor produzidos ao nível dos componentes individuais, em vez de em termos de totalidade da infraestrutura.	
		Suchodoletz et al. (2013)	Exemplificam a ligação íntima da questão da partilha de custos com o	72

		<p>desenvolvimento de modelos de negócio, e a colaboração através de redes de preservação no cálculo de custos.</p> <p>Abordam a questão da utilidade da avaliação do potencial de absorção de um serviço pelo público, relativamente à capacidade de planeamento e desenvolvimento do modelo de negócio.</p>	73
	Lee et al. (2016)	<p>Mencionam a partilha de custos pela cooperação interinstitucional para poupar no investimento feito no âmbito da conformidade legal.</p> <p>Ainda existem várias instituições de investigação científica que não têm em consideração os custos nos seus planos para disponibilizar os dados de forma sustentável.</p>	74 75
	Wright et al. (2009)	Abordam a modelação de custos de risco ligada à incerteza, à imprevisibilidade e à ameaça de ocorrência.	
Plano/Política de gestão de dados	Lee et al. (2016)	Indicam que algumas instituições de investigação científica continuam a mencionar de forma breve ou até a silenciar os custos monetários e administrativos nos seus planos. Poucas instituições sugerem que os investigadores deveriam consagrar os custos da gestão de dados nos seus orçamentos.	
	Ayris (2009)	Advoga a escolha da dimensão adequada das atividades de preservação, que precisam de ser suficientemente grandes para conseguirem responder aos desafios, mas não tão grandes que modo a que o seu fracasso possa ser catastrófico. Considera não haver substituto para uma organização flexível, empenhada e dedicada em preservar um <i>corpus</i> de material. Esta organização deve ser capaz de fazer escolhas e traçar estratégias para lidar com problemas inesperados de todo o tipo. Se necessário, pode realizar triagens e estabelecer compromissos.	
	Donnelly & North (2011)	Observam que a existência de uma abordagem única para o futuro das ciências ou uma política de dados não específica não será produtiva e eficaz, e que as políticas e a prestação de serviços de informação precisam de se aproximar das comunidades de investigação.	
	Wilson & Jeffreys (2013)	Afirmam que os investigadores não estão expostos aos custos reais das soluções de gestão de dados (frequentemente subaproveitados), ao passo que, quando estes serviços estão centralizados, os custos totais são visíveis e desanimadores.	
	Delasalle (2013)	Defende que os planos de gestão de dados devem corresponder às expectativas dos cientistas e à sua área de estudo e também servir para calcular os custos associados aos planos do projeto.	
	Rice et al. (2013)	Consideram que um serviço gratuito no momento da utilização elimina uma das principais barreiras para a boa curadoria e permite que os dados possam ser guardados na infraestrutura adequada, em vez de uma infraestrutura escolhida em função do menor custo.	
Questões legais	Poole (2015)	Defende que as suas exigências devem ser equilibradas pelas partes interessadas, no âmbito da política.	
	Bachell & Barr (2014)	Mencionam os direitos de propriedade intelectual em termos de videojogos.	
	Barros (2014)	Mencionam os direitos de propriedade intelectual em termos de conteúdo jornalístico.	
	Lee et al. (2016)	Referem a existência de obrigações legais por parte das instituições de investigação científica financiada pelo governo dos EUA, devendo disponibilizar os dados em acesso público gratuito, os quais devem ser detetáveis e utilizáveis.	
Tomada de decisão	Whyte & Pryor (2011)	Identificam deliberações para os decisores políticos, investigadores e comunidades de utilizadores sobre os riscos, os custos e os benefícios de uma participação mais ampla antes e depois dos resultados serem produzidos.	
	Delasalle (2013)	Defende o envolvimento dos investigadores no processo de tomada de decisão sobre o que é mantido, onde e por quanto tempo.	
Política de gestão de	Bicarregui et	Tecem considerações sobre as políticas de gestão de preservação e acesso	

preservação e acesso	al. (2013)	ligadas à certificação, que devem ser apoiada por um conjunto de ferramentas metodológicas, incluindo descritivos, estudos de caso e modelos de custo para fornecer orientações sobre o desenvolvimento das melhores práticas em política de gestão e preservação de dados e infraestrutura para esses projetos.
----------------------	------------	--

Table 2. (Tabela 2) Exemplo de aplicação do modelo de enquadramento de análise às percepções dos autores citados. Sistematização das ideias expressas pelos autores no âmbito de cada categoria analisada. FONTE: Elaboração dos autores, com base em Akers & Green (2014); Ayris (2009); Bachell & Barr (2014); Barros (2014); Bicarregui et al. (2013); CCSDS (2002); Corujo, Silva & Revez (2016); Currall et al. (2007); Davies et al. (2007); DCC (2004); Delasalle (2013); Donnelly & North (2011); Dürr et al. (2008); Evans & Moore (2014); Faria & Ferreira (2012); Kejser et al. (2011); Kilbride & Norris (2014); Lee et al. (2016); Machado (2015); Minor et al. (2010); Poole (2015); Queiroz (2013); Rice et al. (2013); Rosenthal & Vargas (2013); Rusbridge & Ross (2007); Santos (2014); Saraiva & Quaresma (2015); Sayão & Sales (2012); Suchodoletz et al. (2013); Whyte & Pryor (2011); Wilson & Jeffreys (2013); Wright et al. (2009)

FONTE: Elaboração dos autores, com base em Akers & Green (2014); Ayris (2009); Bachell & Barr (2014); Barros (2014); Bicarregui et al. (2013); CCSDS (2002); Corujo, Silva & Revez (2016); Currall et al. (2007); Davies et al. (2007); DCC (2004); Delasalle (2013); Donnelly & North (2011); Dürr et al. (2008); Evans & Moore (2014); Faria & Ferreira (2012); Kejser et al. (2011); Kilbride & Norris (2014); Lee et al. (2016); Machado (2015); Minor et al. (2010); Poole (2015); Queiroz (2013); Rice et al. (2013); Rosenthal & Vargas (2013); Rusbridge & Ross (2007); Santos (2014); Saraiva & Quaresma (2015); Sayão & Sales (2012); Suchodoletz et al. (2013); Whyte & Pryor (2011); Wilson & Jeffreys (2013); Wright et al. (2009).

76

Abstract

Introduction - Regarding the concerns that prompted the emergence of digital curation, this study takes as an example the context of the production of large volumes of scientific information, which requires approaches that ensure its maintenance, reuse and valorization, considering its cost.

Objectives - It is intended to discuss the relevant issues regarding the costs of digital curation and to develop a framework model proposal for the analysis of costs in digital curation, including projects of a more practical nature. This means addressing the concept definition and the issue of costs, based on studies related to cost models.

Methodology - A literature review was carried out, contextualizing the issue of costs within digital curation, with data collected from *Biblioteca do Conhecimento Online* – B-On (Online Knowledge Library) and *Repositórios Científicos de Acesso Aberto de Portugal* – RCAAP (Scientific Open Access Repository of Portugal) as research literature sources. Then, a framework model scheme was developed based on the categories identified using the Constant Comparative Method, some related to the *Open Archival Information System* (OAIS) Reference Model and to the *Digital Curation Centre* (DCC) Life Cycle. This allowed the development of a content analysis of the perceptions collected from different authors, resulting in memos, of which this text is a summary.

Results / Conclusion - It is proposed a systematization of digital curation issues, which constitutes a framework model considered relevant for the analysis of costs in digital curation, including practical projects, which interconnects the DCC Life Cycle view of the digital object curation to the OAIS Reference Model approach, in a transversal logic seized by the cost models and plan/data management policies. It is concluded that a paradigm shift is occurring from a black-box perspective to another one that identifies costs and attempts to standardise predictive models for institutional use with the aim of promoting transparency, accountability, and capturing the interest of potential funding providers.

Note on Translation

For articles in languages other than English, DHQ provides an English-language abstract to support searching and discovery, and to enable those not fluent in the article's original language to get a basic understanding of its contents. In many cases, machine translation may be helpful for those seeking more detailed access. While DHQ does not typically have the resources to translate articles in full, we welcome contributions of effort from readers. If you are interested in translating any article into another language, please contact us at editors@digitalhumanities.org and we will be happy to work with you.

1

Works Cited

4C Project 2013 4C Project 2013, *4C Project Glossary*, 4C Project - Collaboration to Clarify the Costs of Curation. Glossary, viewed 4 March 2019, <http://www.4cproject.eu/community-resources/glossary/>.

- Abbott 2008** Abbott, D 2008, *DCC Briefing Paper: What is digital curation?*, Digital Curation Centre, Edinburgh.
- Akers and Green 2014** Akers, K & Green, J 2014, "Towards a Symbiotic Relationship Between Academic Libraries and Disciplinary Data Repositories: A Dryad and University of Michigan Case Study", *International Journal of Digital Curation*, vol. 9, no. 1, pp. 119–131.
- Atkins 2003** Atkins, DE 2003, *Revolutionizing science and engineering through cyberinfrastructure: report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*, National Science Foundation, Arlington, VA.
- Ayris 2009** Ayris, P 2009, "LIBER's Involvement in Supporting Digital Preservation in Member Libraries", *Liber Quarterly: The Journal of European Research Libraries*, vol. 19, no. 1, pp. 22–43.
- Bachell and Barr 2014** Bachell, A & Barr, M 2014, "Video Game Preservation in the UK: A Survey of Records Management Practices", *International Journal of Digital Curation Volume*, vol. 9, no. 2, pp. 139–170.
- Bardin 2011** Bardin, L 2011, *Análise de Conteúdo*, Edições 70, São Paulo.
- Barros 2014** Barros, N 2014, "Apropriação da curadoria na web por uma empresa de mídia tradicional : um caso de convergência entre narrativa e banco de dados", Unicamp, Campinas.
- Beagrie 2006** Beagrie, N 2006, "Digital Curation for Science, Digital Libraries, and Individuals", *International Journal of Digital Curation*, Vol 1, Iss 1, Pp 3-16 (2006), no. 1, p. 3.
- Beagrie and Pothen 2002** Beagrie, N & Pothen, P 2002, "Digital Curation: Digital Archives, Libraries and e-Science Seminar", *Ariadne*, no. 30, viewed 14 March 2018, <http://www.ariadne.ac.uk/issue30/digital-curation>.
- Benardou et al. 2010** Benardou, A, Constantopoulos, P, Dallas, C & Gavrilis, D 2010, "Understanding the Information Requirements of Arts and Humanities Scholarship", *International Journal of Digital Curation*, vol. 5, no. 1, pp. 18–33.
- Bicarregui et al. 2013** Bicarregui, J, Gray, N, Henderson, R, Jones, R, Lambert, S & Matthews, B 2013, "Data Management and Preservation Planning for Big Science", *International Journal of Digital Curation*, vol. 8, no. 1, pp. 29–41.
- Borgman 2007** Borgman, CL 2007, *Scholarship in the digital age: information, infrastructure, and the Internet*, MIT Press, Cambridge, Mass.
- Borgman 2015** — 2015, *Big data, little data, no data: scholarship in the networked world*, The MIT Press, Cambridge, Massachusetts.
- Borgman et al. 2007** Borgman, CL, Wallis, JC & Enyedy, N 2007, "Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries", *International Journal on Digital Libraries*, vol. 7, no. 1–2, pp. 17–30.
- Buckland 2011** Buckland, M 2011, "Data Management as Bibliography", *Bulletin of the American Society for Information Science & Technology*, vol. 37, no. 6, pp. 34–37.
- CCSDS 2002** CCSDS 2002, "Reference Model for an Open Archival Information System (OAIS)", viewed 16 June 2016, http://nssdc.gsfc.nasa.gov/nost/isoas/us19/650x0_20010226rl.pdf.
- Constantopoulos et al. 2009** Constantopoulos, P, Dallas, C, Androutsopoulos, I, Angelis, S, Deligiannakis, A, Gavrilis, D, Kotidis, Y & Papatheodorou, C 2009, 'DCC&U: An Extended Digital Curation Lifecycle Model', *International Journal of Digital Curation*, vol. 4, no. 1, pp. 34–45.
- Corujo 2015** Corujo, L 2015, "Repositórios digitais e confiança: um exemplo de prevenção digital: o RODA", Dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa, Lisboa, viewed 25 January 2018, <http://repositorio.ul.pt/handle/10451/18109>.
- Corujo et al. 2016** Corujo, L, Silva, CG da & Revez, J 2016, "Digital Curation and Costs: Approaches and Perceptions", in *TEEM'16 Proceedings*, ACM, pp. 277–284.
- Cragin et al. 2010** Cragin, MH, Palmer, CL, Carlson, JR & Witt, M 2010, "Data sharing, small science and institutional repositories", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1926, pp. 4023–4038.
- Cruz Mundet and Díez Carrera 2015** Cruz Mundet, JR & Díez Carrera, C 2015, "El cálculo de costes de la preservación digital: un análisis de modelos", *Anales de Documentación*, vol. 18, no. 2, viewed 16 June 2016, <http://revistas.um.es/analesdoc/article/view/228411>.
- Currall et al. 2007** Currall, J, Johnson, C & McKinney, P 2007, "The world is all grown digital.... How shall a man persuade management what to do in such times?", *International Journal of Digital Curation*, vol. 2, no. 1, pp. 12–28.
- DCC 2004** DCC 2004, *DCC Curation Lifecycle Model*, DCC Curation Lifecycle Model, viewed 12 March 2018, <http://www.dcc.ac.uk/resources/curation-lifecycle-model>.
- DCC 2014** — 2014, *Cost Model for Digital Preservation*, DCC Resources - Cost Model for Digital Preservation, viewed 5 March 2019, <http://www.dcc.ac.uk/resources/external/cost-model-digital-preservation>.
- Dallas 2016** Dallas, C 2016, "Digital Curation beyond the "wild frontier": a Pragmatic Approach", *Archival Science*, no. 16, pp. 421–457.

- Dallas and Doorn 2009** Dallas, C & Doorn, P 2009, "Report on the Workshop on Digital Curation in the Human Sciences at ECDL 2009", *D-Lib Magazine*, vol. 15, no. 11/12.
- Davies et al. 2007** Davies, R, Ayris, P, McLeod, R, Shenton, H & Wheatly, P 2007, "How much does it cost? The LIFE Project - Costing Models for Digital Curation and Preservation", *Liber Quarterly: The Journal of European Research Libraries*, vol. 17, no. 3/4.
- Delasalle 2013** Delasalle, J 2013, "Research Data Management at the University of Warwick: recent steps towards a joined-up approach at a UK university.", *Libreas Library Ideas*, vol. 9, no. 2, pp. 97–105.
- Dillon 2013** Dillon, C 2013, "The Research Library as Digital Curator at Virginia Tech.", *College Undergraduate Libraries*, vol. 20, no. 2, pp. 232–238.
- Donnelly and North 2011** Donnelly, M & North, R 2011, "The Milieu and the MESSAGE: Talking to Researchers about Data Curation Issues in a Large and Diverse e-Science Project", *International Journal of Digital Curation*, vol. 6, no. 1, pp. 32–44.
- Donnelly et al. 2010** Donnelly, M, Jones, S & Pattenden-Fail, J 2010, "DMP Online: The Digital Curation Centre's Web-based Tool for Creating, Maintaining and Exporting Data Management Plans", *International Journal of Digital Curation*, vol. 5, no. 1, pp. 187–193.
- Dürr et al. 2008** Dürr, R, Meer, K, Luxemburg, W & Dekker, R 2008, "Dataset Preservation for the Long Term: Results of the DareLux Project", *International Journal of Digital Curation*, vol. 3, no. 1, pp. 29–43.
- Edmond and Garnet 2015** Edmond, J & Garnet, V 2015, "APIs and Researchers: The Emperor's New Clothes?", *International Journal of Digital Curation*, vol. 10, no. 1, pp. 287–297.
- Edwards 2010** Edwards, PN 2010, *A vast machine: computer models, climate data, and the politics of global warming*, MIT Press, Cambridge, Mass.
- Edwards et al. 2007** Edwards, PN, Jackson, SJ, Bowker, GC & Knobel, CP 2007, *Understanding Infrastructure: Dynamics, Tensions, and Design*, University of Michigan, Ann Arbor, viewed 9 March 2019, <http://hdl.handle.net/2027.42/49353>.
- Edwards et al. 2011** Edwards, PN, Mayernik, MS, Batcheller, AL, Bowker, GC & Borgman, CL 2011, "Science friction: Data, metadata, and collaboration", *Social Studies of Science*, vol. 41, no. 5, pp. 667–690.
- Evans and Moore 2014** Evans, T & Moore, R 2014, "The Use of PDF/A in Digital Archives: Study from Archaeology", *International Journal of Digital Curation*, vol. 9, no. 2, pp. 123–138.
- Faria and Ferreira 2012** Faria, L & Ferreira, M 2012, "Plataforma de colaboração para custear a curadoria digital", paper presented at 12.o Congresso Nacional BAD, Évora, APBAD.
- Faria and Ferreira 2015** Faria, L & Ferreira, M 2015, "Plataforma de colaboração para custear a curadoria digital", in *12o Congresso Nacional BAD, 2015*, Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas (APBAD), viewed 27 January 2018, <http://repositorium.sdum.uminho.pt/handle/1822/37858>.
- Ferreira et al. 2012** Ferreira, M, Saraiva, R & Rodrigues, E 2012, *Estado da arte em preservação digital*, Universidade do Minho, Braga.
- Ferreira et al. 2014** Ferreira, M, Faria, L & Silva, H 2014, *D2. 1: baseline study of stakeholder & stakeholder initiatives*, 4C Project, viewed 16 June 2016, <http://repositorium.sdum.uminho.pt/handle/1822/30903>.
- Flick 2018** Flick, U 2018, *Doing grounded theory*, U Flick (ed.), 2nd edition, The SAGE qualitative research kit, SAGE, Los Angeles London New Delhi Singapore Washington, DC Melbourne.
- Foster et al. 2013** Foster, I, Borgman, CL & Heidorn, PB 2013, *Empowering Long Tail Research*, viewed 9 March 2019, <https://sites.google.com/site/ielttrconcept/>.
- Glaser 1965** Glaser, BG 1965, "The Constant Comparative Method of Qualitative Analysis", *Social Problems*, vol. 12, no. 4, pp. 436–445.
- Heidorn 2011** Heidorn, PB 2011, "The Emerging Role of Libraries in Data Curation and E-science", *Journal of Library Administration*, vol. 51, no. 7/8, pp. 662–672.
- Hey and Trefethen 2005** Hey, T & Trefethen, AE 2005, "Cyberinfrastructure for e-Science", *Science*, vol. 308, no. 5723, pp. 817–821.
- Higgins 2011** Higgins, S 2011, "Digital Curation: The Emergence of a New Discipline", *The International Journal of Digital Curation*, vol. 6, no. 2, pp. 78–88.
- Kejser et al. 2011** Kejser, U, Nielsen, A & Thirifays, A 2011, "Cost Model for Digital Preservation: Cost of Digital Migration", *International Journal of Digital Curation*, vol. 6, no. 1, pp. 255–267.
- Kejser et al. 2014** Kejser, UB, Johansen, KHE, Thirifays, A, Nielsen, A, Wang, D, Strodl, S, Miksa, T, Davidson, J, McCann, P, Krupp, J & Tjalsma, H 2014, *D3.1 - Evaluation of Cost Models and Needs & Gaps Analysis. 4C Project*, report, 30 June, viewed 27 January 2018, http://www.4cproject.eu/documents/D3.1_final_report_10May2014-v1.02.pdf.
- Kilbride and Norris 2014** Kilbride, W & Norris, S 2014, "Collaborating to Clarify the Cost of Curation", *New Review of Information Networking*, vol. 19, no. 1, pp. 44–49.

- Laney 2001** Laney, D 2001, *3D Data Management: Controlling Data Volume, Velocity, and Variety*, February, META Group, viewed 9 March 2019, <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
- Lee and Tibbo 2011** Lee, CA & Tibbo, H 2011, "Where's the archivist in digital curation? Exploring the possibilities through a matrix of knowledge and skills", *Archivaria*, vol. 72, no. Fall 2011, pp. 123–168.
- Lee et al. 2016** Lee, CA, Allard, S, McGovern, N & Bishop, A 2016, "Open Data Meets Digital Curation: An Investigation of Practices and Needs", *International Journal of Digital Curation*, vol. 11, no. 2, pp. 115–125.
- Macdonald and Lord 2003** Macdonald, A & Lord, P 2003, *Digital Data Curation Task Force Report of the Task Force Strategy Discussion Day*, Digital Archiving Consultancy, London, p. 18, viewed 14 March 2018, https://www.webarchive.org.uk/wayback/archive/20140616043831/http://www.jisc.ac.uk/uploaded_documents/CurationTaskForceFinal1.pdf.
- Machado 2015** Machado, D 2015, "Dados de pesquisa em repositório institucional: o caso do Edinburgh DataShare", Universidade Federal do Rio Grande do Sul, Porto Alegre.
- Machado et al. 2015** Machado, H, Soares, P & Silva, T 2015, "Em busca duma anamnese universitária: a materialização do arquivo do ISSSL na Internet", paper presented at 12.o Congresso Nacional BAD, Évora, APBAD.
- Manyika et al. 2013** Manyika, J, Chui, M, Farrell, D, Kuiken, SV, Groves, P & Doshi, EA 2013, *Open data: Unlocking innovation and performance with liquid information*, McKinsey & Company, viewed 9 March 2019, <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>.
- Mayer-Schonberger and Cukier 2013** Mayer-Schonberger, V & Cukier, K 2013, *Big data: la revolución de los datos masivos*, Turner, Madrid.
- Minor et al. 2010** Minor, D, Sutton, D, Kozbial, A, Westbrook, B, Burek, M & Smorul, M 2010, "Chronopolis Digital Preservation Network", *International Journal of Digital Curation*, vol. 5, no. 1, pp. 119–133.
- Molloy 2011** Molloy, JC 2011, "The Open Knowledge Foundation: Open Data Means Better Science", *PLoS Biology*, vol. 9, no. 12, pp. 1–4.
- Mundet and Carrera 2015** Mundet, JRC & Carrera, CD 2015, "El cálculo de costes de la preservación digital: un análisis de modelos", *Anales de Documentación*, vol. 18, no. 2, viewed 27 January 2018, <http://revistas.um.es/analesdoc/article/view/228411>.
- Murray-Rust 2008** Murray-Rust, P 2008, "Open Data in Science", *Serials Review*, vol. 34, no. 1, pp. 52–64.
- Murray-Rust and Rzepa 2004** Murray-Rust, P & Rzepa, HS 2004, "The Next Big Thing: From Hypermedia to Datuments", *Journal of Digital Information*, vol. 5, no. 1, viewed 9 March 2019, <https://journals.tdl.org/jodi/index.php/jodi/article/view/130>.
- OCDE 2007** Organização para a Cooperação e Desenvolvimento Económico 2007, *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OCDE, viewed 9 March 2019, <http://www.oecd.org/sti/inno/38500813.pdf>.
- OED n.d.** *Oxford English Dictionary* n.d., viewed 9 March 2019, <http://www.oed.com/>.
- Ogburn 2010** Ogburn, JL 2010, "The Imperative for Data Curation", *portal: Libraries and the Academy*, vol. 10, no. 2, pp. 241–246.
- Open Data Commons 2013** *Open Data Commons* 2013, Open Data Commons, viewed 9 March 2019, <https://opendatacommons.org/>.
- Poole 2013** Poole, AH 2013, "Now is the Future Now? The Urgency of Digital Curation in the Digital Humanities", *Digital Humanities Quarterly*, vol. 007, no. 2.
- Poole 2015** — 2015, "How has your science data grown? Digital curation and the human factor: a critical literature review", *Archival Science*, vol. 15, no. 1, pp. 101–139.
- Price 1963** Price, DJ de S 1963, *Little science, big science*, George B. Pegram lectures, Columbia University Press, New York.
- Queiroz 2013** Queiroz, B 2013, "A preservação da informação na Universidade Federal de Goiás: uma proposta de curadoria digital", Universidade Federal de Goiás, Goiânia.
- RS PSU 2012** Royal Society (Reino Unido) & Policy Studies Unit 2012, *Science as an open enterprise.*, viewed 9 March 2019, https://royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf.
- Ramos 2012** Ramos, D 2012, "Anotações para a compreensão da atividade do "curador de informação digital", in E Correa (ed.), *Curadoria Digital e o Campo da Comunicação*, ECA-USP, São Paulo, pp. 11–21, viewed 16 June 2016, https://issuu.com/grupo-ecausp.com/docs/ebook_curadoria_digital_usp.
- Rice et al. 2013** Rice, R, Ekmekcioglu, Ç, Haywood, J, Jones, S, Lewis, S, Macdonald, S & Weir, T 2013, "Implementing the Research Data Management Policy: University of Edinburgh Roadmap", *International Journal of Digital Curation*, vol. 8, no. 2, pp. 194–204.
- Rodrigues et al. 2015** Rodrigues, A, Barbedo, F, Runa, L & Sant'Ana, M 2015, *Continuidade digital: relatório final do projecto*, DGLAB, Lisboa.
- Rosenthal and Vargas 2013** Rosenthal, D & Vargas, D 2013, "Distributed Digital Preservation in the Cloud", *International Journal of*

- Rusbridge and Ross 2007** Rusbridge, A & Ross, S 2007, "The UK LOCKSS Pilot Programme: A Perspective from the LOCKSS Technical Support Service", *International Journal of Digital Curation*, vol. 2, no. 2, pp. 111–122.
- Santos 2014** Santos, T 2014, "Curadoria Digital: o conceito no período de 2000 a 2013", Universidade de Brasília, Brasília.
- Saraiva and Quaresma 2015** Saraiva, P & Quaresma, P 2015, "Bibliotecas Universitárias: tendências, modelos e competências", paper presented at 12.o Congresso Nacional BAD, Évora, APBAD.
- Sayão and Sales 2012** Sayão, L & Sales, L 2012, "Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa", *Informação & Sociedade: Estudos (I&S)*, vol. 22, no. 3, pp. 179–191.
- Schroeder 2014** Schroeder, R 2014, "Big Data: Towards a More Scientific Social Science and Humanities?", in M Graham, WH Dutton & M Castells (eds), *Society and the internet: how networks of information and communication are changing our lives*, First edition, Oxford University Press, Oxford ; New York, NY.
- Strasser et al. 2014** Strasser, C, Abrams, S & Cruse, P 2014, "DMPTool 2: Expanding Functionality for Better Data Management Planning", *International Journal of Digital Curation*, vol. 9, no. 1, pp. 324–330.
- Suchodoletz et al. 2013** Suchodoletz, D, Rechert, K & Valizada, I 2013, "Towards Emulation-as-a-Service: Cloud Services for Versatile Digital Object Access", *International Journal of Digital Curation*, vol. 8, no. 1, pp. 131–142.
- Taper and Lele 2004** Taper, ML & Lele, S (eds) 2004, "Models of Scientific Inquiry and Statistical Practice: Implications for the Structure of Scientific Knowledge", in *The nature of scientific evidence: statistical, philosophical, and empirical considerations*, University of Chicago Press, Chicago, pp. 17–51.
- Thibodeau 2002** Thibodeau, K 2002, "Overview of technological approaches to digital preservation and challenges in coming years", in Council on Library and Information Resources & Institute for Information Science (eds), *The state of digital preservation: an international perspective ; conference proceedings, Documentation Abstracts, Inc. Institutes for Information Science, Washington, D.C., April 24-25, 2002*, CLIR, Washington, D.C, viewed 5 February 2019, <http://www.clir.org/pubs/reports/reports/pub107/pub107.pdf>.
- Unsworth et al. 2006** Unsworth, J, Courant, P, Fraser, S, Goodchild, M, Hedstrom, M, Henry, C, Kaufman, PB, McGann, J, Rosenzweig, R & Zuckerman, B 2006, *Our Cultural Commonwealth: The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*, text, 13 December, ACLS: New York, viewed 9 March 2019, <https://www.ideals.illinois.edu/handle/2142/189>.
- Walters and Skinner 2011** Walters, T & Skinner, K 2011, *New roles for new times: Digital curation for preservation*, Association of Research Libraries, viewed 16 June 2016, <https://vtechworks.lib.vt.edu/handle/10919/10183>.
- Waters and Garrett 1996** Waters, D & Garrett, J 1996, *Preserving Digital Information. Report of the Task Force on Archiving of Digital Information.*, CPA/RLG, viewed 16 June 2016, <http://eric.ed.gov/?id=ED395602>.
- Weinberg 1961** Weinberg, AM 1961, "Impact of Large-Scale Science on the United States: Big science is here to stay, but we have yet to make the hard financial and educational choices it imposes", *Science (New York, N.Y.)*, vol. 134, no. 3473, pp. 161–164.
- Whyte and Pryor 2011** Whyte, A & Pryor, G 2011, "Open Science in Practice: Researcher Perspectives and Participation", *International Journal of Digital Curation Volume pp*, vol. 6, no. 1, pp. 199–213.
- Wilson and Jeffreys 2013** Wilson, J & Jeffreys, P 2013, "Towards a Unified University Infrastructure: The Data Management Roll-Out at the University of Oxford", *International Journal of Digital Curation*, vol. 8, no. 1, pp. 235–246.
- Wright et al. 2009** Wright, R, Miller, A & Addis, M 2009, "The Significance of Storage in the "Cost of Risk" of Digital Preservation", *International Journal of Digital Curation*, vol. 4, no. 3, pp. 104–122.