

Capítulo 1

Arquivistas, cientistas e dados abertos: uma equação complexa?

Maria Cristina Vieira de Freitas,¹ Luís Miguel Nunes Corujo²

¹0000-0002-8849-8792, Universidade de Coimbra, Portugal, cristina.freitas@fl.uc.pt

²0000-0003-4411-2453, Universidade de Lisboa, Portugal, luiscorujo@campus.ul.pt

Resumo. O meio digital favorece a produção intensiva e a partilha de dados, e obriga ao desenvolvimento de práticas de ciência aberta, de curadoria e de novas competências no âmbito da literacia de informação. O “dilúvio de dados” afeta os mais diversos quadrantes institucionais e governamentais. Concordando com esse enquadramento, são objetivos deste capítulo: compreender algumas das componentes envolvidas nesses novos ambientes, focalizar a atenção no papel que arquivistas podem desempenhar nesse meio, apoiando cientistas em tarefas de gestão de dados de pesquisa e apresentar algumas iniciativas internacionais nessas matérias. Para o efeito, realizamos pesquisas bibliográficas em bases de dados, com especial incidência na recolha de textos publicados nos últimos 10 anos. Os principais resultados e conclusões indicam que apesar da resistência inicial em admitir os dados de pesquisa como objetos de trabalho há uma crescente tendência para que arquivistas aceitem este desafio, engajem-se nas iniciativas que tomam corpo nas instituições e nos governos e encarem a gestão dos dados de pesquisa como uma oportunidade para o crescimento e a visibilidade académica e profissional.

Palavras-chave. gestores de informação, arquivistas, dados abertos, gestão de dados, literacia de dados.

Abstract. The digital environment supports intensive production and data sharing and requires the development of open science and curation practices, and skills in the information literacy field. The “data deluge” affects the most diverse institutional and governmental environments. According to this framework, the aims of this chapter are: to understand some of the components involved in these new environments, to focus attention on the role that archivists can play in this environment, supporting scientists in research data management tasks and to present some

international initiatives concerning these matters. For this purpose, bibliographic searches in databases were carried out, focused on the collection of papers published in the last 10 years. The main results and conclusions suggest that, despite initial resistance to acknowledge research data as a work object, there is a growing tendency for archivists to accept this challenge, to embed in projects that take place in institutions and governments, and view research data management as an opportunity for growth and academic and professional visibility.

Keywords. information managers, archivists, open data, data management, data literacy.

Introdução

Quem controla o passado, controla o futuro:
quem controla o presente, controla o passado.
(George Orwell. 1984)

Há quase uma década, Akmon, Zimmerman, Daniels e Hedstrom (2011) recordavam que as diferentes possibilidades do meio digital favoreciam a produção intensiva de dados e obrigavam o desenvolvimento de novas competências necessárias à sua gestão. O “dilúvio de dados” (“*data deluge*”¹), segundo diziam, afetava os mais diversos quadrantes institucionais (*e.g.*, governos, laboratórios científicos, universidades, empresas, etc.), com especial incidência em atividades e em tarefas relacionadas com a definição de responsabilidades e de prioridades. Num tal contexto, responder a perguntas aparentemente simples (*e.g.*, “o que”, “quem”, “como”, “quando”, “onde”, “com que meios” e “por quanto tempo”), mas suficientemente sensibilizadoras para ir à raiz dos problemas e encaminhar soluções era, e ainda é, uma tarefa absolutamente crucial. Também recordavam o crescente interesse que esses temas vinham recebendo, por parte dos/as cientistas, então tidos/as como um dos/as principais agentes envolvidos/as nas infraestruturas de dados emergentes, ao nível institucional e governamental. Todos esses ingredientes se reuniam em torno das competências reclamadas a uma nova área de estudo, situada na confluência de diversas áreas: a “*data-intensive science*” ou “*e-Science*”².

Num tal cenário, o desenvolvimento da ciência, conforme também argumentaram Hey, Tansley e Tolle (2009), tornar-se-ia cada vez mais dependente da organização e do controlo desses conjuntos cada vez mais expressivos de dados, disponibilizados e partilhados pelas suas comunidades de práticas, cujos usos não mais confinar-se-iam, de um

¹ A expressão “*data deluge*” é a metáfora convencionalmente usada no meio tecnológico para explicar que o volume de dados diariamente produzido nos ambientes digitais, atualmente, ultrapassa largamente as nossas capacidades de gestão e de utilização da sua totalidade.

² De acordo com a definição dada pela *International Conference on eScience*, o termo “*eScience promotes innovation in collaborative, computationally-or data-intensive research across all disciplines, throughout the research lifecycle*”. Disponível em: <https://escience-conference.org/>.

modo exclusivo, a grupos específicos de pesquisadores/as. Daí também derivaria que as possíveis futuras utilizações desses dados deveriam ser antecipadas pelos seus responsáveis (produtores/as e curadores/as), como forma de potenciar e de facilitar essas tarefas³. Com efeito, ao tomarem essas medidas, consciente ou inconscientemente, os/as responsáveis pelo tratamento desses dados estavam a transferir para o seu próprio contexto algumas das clássicas atividades (*e.g.*, reunir, organizar, selecionar, descrever, preservar e disponibilizar documentos/informação) desempenhadas por profissionais de informação, incluindo-se, entre estes/as, os/as arquivistas.

Da confluência desses assuntos, surgem-nos as seguintes perguntas: qual seria a relação a estabelecer entre arquivistas e cientistas na gestão desse conjunto expressivo de dados científicos? Qual seria o contributo da Arquivística, e dos/as profissionais de arquivo, nessas matérias? Haverá exemplos de iniciativas institucionais e suprainstitucionais que, de algum modo, evidenciam essa relação e esse contributo?

Uma equação é uma redução de um problema a pontos simples e claros, de modo a facilitar a obtenção de uma solução. Consistentemente com esta asserção e tentando responder às perguntas antes formuladas, no presente capítulo, estruturamos as nossas ideias em duas secções principais. Assim, na primeira secção, intentamos compreender algumas das componentes envolvidas nessas matérias, focalizando as nossas análises em conteúdos de textos publicados nos últimos 10 anos, maioritariamente artigos científicos, divulgados em bases de dados e em plataformas agregadoras consideradas credíveis e certificadas pela ciência⁴. Na segunda secção, apresentamos algumas das iniciativas que ilustram e corroboram as análises previamente realizadas, finalizando com uma síntese das ideias principais, em jeito de conclusão.

1. Compreendendo a equação: arquivos, ciência aberta e dados abertos

De acordo com Kassen (2014), a investigação em torno dos dados abertos é uma tendência, que se vem propagando pelos vários ecossistemas ligados às tecnologias de informação e comunicação, desde há alguns anos. Por um lado, nota-se o interesse de um número cada vez maior de países em “abrir os seus dados” governamentais ao público, o que muito se deve aos benefícios e aos impactos positivos que se julga poder auferir a

³ Akmon, Zimmerman, Daniels e Hedstrom (2011, p. 333) mencionaram alguns estudos desenvolvidos fora do âmbito arquivístico e que, na altura, abordaram as questões envolvidas na utilização dos dados científicos: a) Piwowar e Chapman (2009), por exemplo, debruçaram-se sobre a adesão dos/as cientistas às políticas de divulgação de dados; b) Zimmerman (2008), por seu turno, analisou os desafios envolvidos na reutilização de dados científicos; e c) Borgman, Wallis e Enyedy (2007) investigaram as dificuldades inerentes ao processo de produção de dados científicos partilháveis pela comunidade científica. Em comum, para além de examinar os obstáculos encontrados, desde a produção até à partilha dos dados, estes estudos tiveram o mérito de revelar como os/as cientistas produziam e geravam quotidianamente os seus dados, e realçaram a importância do envolvimento precoce e concertado, de arquivistas e de cientistas, em todas as fases do seu ciclo de vida.

⁴ Sensivelmente: B-On® (disponível em: <https://www.b-on.pt/>); Scopus® e Web of Science®.

partir desta abertura. Por outro, a transparência e a prestação de contas (“*accountability*”), a participação e a capacitação autónoma de cidadãos/ãs, o crescimento económico e o estímulo à inovação são hoje vantagens competitivas avaliadas como conquistas e como benefícios associados às políticas governamentais de abertura dos dados. Não obstante, e a nível global, nota-se que os impactos positivos auferidos não têm atingido transversalmente os diversos países, permanecendo uma situação de assimetria quer quanto ao grau de maturidade dos dados produzidos e geridos, quer quanto à existência de práticas e de políticas de abertura dos mesmos (Kassen, 2014).

Akmon, Zimmerman, Daniels e Hedstrom (2011) referem que entre os anos que compreenderam as décadas de 60 e de 90 do século XX surgiram estudos longitudinais, financiados por instituições de pesquisa, cujo foco foi compreender as práticas científicas e o registo dessas mesmas práticas, pela ótica dos/as cientistas, considerados/as agentes imprescindíveis nos processos de gestão dos dados e na descoberta de novos conhecimentos. Estes estudos, de um certo modo, contribuíram para apoiar as atividades arquivísticas voltadas à seleção - para a preservação ou para a eliminação -, de documentos e de registos de ações consideradas vitais nesses mesmos processos. Ainda, indicam ter sido realizados outros estudos, em diferentes momentos e tendo como pano de fundo a ótica arquivística⁵, que revelaram atitudes pouco favoráveis de profissionais da área, no que respeita a admitir que os dados produzidos por cientistas em investigações eram algo mais do que simplesmente “dados”, não merecendo, pois, especial atenção. Como justificação para essa negligência destacava-se o facto de os/as cientistas fecharem os seus dados em laboratórios e de esse comportamento, que de resto demonstrava a sua pouca capacidade de abertura, desencorajar os/as arquivistas a intervir na sua avaliação e no seu tratamento⁶.

Com efeito, o problema da admissibilidade dos dados nos arquivos, que é objeto de debate desde o dealbar da era digital, deve-se a razões de variada ordem. Algumas delas terão conexão com a visão arquivística tradicional de que o que comumente se identifica como “dado” não serve aos interesses da pesquisa histórica; portanto, não interessa à

⁵ Na tradição arquivística que acompanha os países de língua inglesa, e que se baseia na visão clássica da gestão dos arquivos pelo ciclo de vida dos documentos, há uma separação profissional entre *records managers* (gestores/as de documentação corrente e intermédia) e *archivists* (conservadores/as de documentação histórica). Nesse texto, não iremos adentrar estes aspetos. Para os efeitos pretendidos, consideraremos que nas ações de gestão documental/informacional interfere e participa o/a arquivista, o que aliás faz parte do alinhamento estratégico que se espera deste/a profissional nos atuais contextos híbridos e multidisciplinares nos quais exerce as suas atividades.

⁶ Nesmith (2015, p. 122) pondera que a pesquisa arquivística está mais atenta à realidade. Como prova, refere o facto de o tema da preservação dos dados científicos fazer parte da agenda arquivística. Bowker (2000 citado por Nesmith 2015, p. 123), igualmente, reforça que o crescimento da investigação sobre essas matérias, entre os arquivistas, sugere que o passado de negligência do meio digital terá sido abandonado, inclusivamente pelos/as próprios/as cientistas, que já perceberam que não é possível prosseguir com a atividade de reunir dados, publicar artigos e, seguidamente, abandonar os seus *datasets* originais. Nem mesmo as agências de financiamento se mostram disponíveis para promover essas atitudes. O autor também adverte para a necessidade de superar o que identifica como uma falta de perceção sobre a “historicidade dos dados”, o que é corroborado pela seguinte afirmação: “we need to historicize our data and its organization so as to create flexible databases that... might really help us gain long-term purchase on questions of planetary management” (Bowker, 2000 citado por Nesmith, 2015, p. 124).

agenda arquivística (Warnow-Blewett, Genuth, & Weart, 2001). Esta concepção é, no entanto, contrariada pelos resultados de estudos que examinam o tratamento dos dados científicos sob a perspectiva arquivística e identificam a sua real relevância para a comunidade científica envolvida.

A título de exemplo, veja-se o importante estudo que Akmon, Zimmerman, Daniels e Hedstrom (2011) desenvolveram num laboratório de pequeno porte. Os resultados obtidos foram ao encontro de outros estudos realizados na mesma época e clarificaram aspectos relevantes do processo de produção e de gestão dos dados científicos, realizados nesses nichos laboratoriais específicos. Segundo as suas conclusões, esses ambientes representavam uma excelente oportunidade para o trabalho e para a investigação arquivística, uma vez que os/as cientistas, que foram objeto do estudo, demonstraram algum interesse pelas soluções arquivísticas e pela sua aplicação no âmbito da curadoria dos seus dados de investigação. Nesses novos contextos, conforme complementam, tomaram corpo e passaram a fazer todo o sentido o uso de expressões tais como “*data archivists*” e “*data curators*”.

Por esta via, atividades já emergentes e relacionadas à “*digital curation*”, de uma forma geral, ou à “*data curation*”, de um modo específico, converter-se-iam numa oportunidade a não desperdiçar. Assim, convencer os/as cientistas a organizar e a descrever os seus dados, de uma forma apropriada e tendente ao uso imediato ou futuro, e a integrá-los numa infraestrutura normalizada, permitindo a sua divulgação, juntamente com os dados produzidos por outros estudos, tornando-os mais facilmente partilháveis entre a comunidade, passariam a ser alguns dos novos desafios a incluir à agenda arquivística (Borgman, Wallis, & Enyedy, 2007; Birnholtz, & Bietz, 2003; Van House, 2003 citados por Akmon, Zimmerman, Daniels e Hedstrom, 2011, p. 330-331).

De acordo com Lee e Tibbo (2011), o termo “*digital curation*” é usado para identificar as práticas adotadas por diferentes comunidades no âmbito da produção, da partilha e da utilização dos dados, especialmente no que respeita às medidas a tomar para mapear e para preservar aqueles conjuntos (de dados) considerados vitais para a compreensão dos muitos processos que lhes deram causa e que, como tais, devem estar disponíveis para a utilização, tanto no presente como no futuro⁷. Ainda, consideram que o termo possui um largo espectro, servindo como um guarda-chuva, que geralmente abriga um rol de atividades desenvolvidas por diferentes profissionais, dos mais diversos quadrantes, tendo, no entanto, em comum, o uso e o aproveitamento dos recursos tecnológicos necessários aos cuidados que todos/as devemos ter para manter, continuamente e pelo tempo que forem requeridos, os nossos dados de pesquisa. Beagrie (2004 citado por Lee, & Tibbo, 2011, p. 126) refere que o termo é largamente utilizado para identificar medidas e práticas necessárias à permanência dos dados produzidos no meio digital, sejam eles provenientes das pesquisas científicas ou de outros processos, respeitando-se o seu ciclo de vida, transmitindo-os às atuais e às futuras gerações. Neste sentido, a cu-

⁷ Esta é, também, a concepção de Yakel (2007) citada por Lee e Tibbo (2011, p. 159), para quem a “*digital curation*” implica o envolvimento intensivo de profissionais da informação em atividades de gestão (incluindo-se aqui a preservação) de dados digitais, tornando-os utilizáveis sempre que solicitados (no presente e no futuro, indiferentemente).

radoria digital alarga os limites da preservação digital, que nela também se inclui, porque implica “all the processes needed for good data creation and management, and the capacity to add value to data to generate new sources of information and knowledge” (Beagrie, 2004 citado por Lee, & Tibbo, 2011, p. 126).

Noonan e Chute (2014) identificam a curadoria de dados como uma das componentes envolventes da curadoria digital, tratando-se, idealmente, do encadeamento e da prossecução de um conjunto de atividades que proporcionam a boa gestão dos dados, no decorrer do seu ciclo de vida, bem como a sua manutenção apropriada, garantindo a sua avaliação, seleção e armazenagem, em condições de segurança, de integridade e de autenticidade, para serem utilizados e reutilizados nos ambientes tecnológicos disponíveis no presente e também no futuro. A nosso parecer, esta declaração de Noonan e Chute (2014), subliminarmente, vai ao encontro do conceito de “continuidade”, que é um referente absolutamente relevante e que alarga, e muito, o significado do conceito de preservação, sobretudo, e principalmente, no movediço meio digital.

Por esta via, a curadoria de dados passa a fazer parte da ementa dos/as profissionais de informação, que são os/as que historicamente se encontram vinculados à custódia e à proteção dos documentos – e da informação –, como sejam: arquivistas e bibliotecários/as. Walters e Skinner (2011) reforçam que a curadoria de conteúdos – termo que agrega ainda uma outra perspetiva, a do significado – passa a ser o foco real e imediato do trabalho a realizar no âmbito das atividades e das tarefas desempenhadas por esta classe de profissionais.

Num sentido mais estrito, a relação entre a Arquivística e a Curadoria Digital, enquanto caminhos que se cruzam, para Dobrevá e Duff (2015), situa-se naquilo que designam como um “continuum bidirecional”. Assim, se por uma via, a Curadoria Digital ampara-se em conceitos sedimentados no meio arquivístico, entre os quais a noção de preservação relacionada ao contexto, ao conteúdo e à estrutura em que os objetos são produzidos e utilizados, bem como às suas características definidoras – autenticidade e integridade –, por uma outra via, traz à Arquivística a possibilidade de alargar os seus limites e de desafiar os seus próprios conceitos e métodos de trabalho nem sempre completamente adequados aos cambiantes meios digitais.

Noonan e Chute (2014) também identificam, nos espaços de circunscrição da curadoria digital, algumas oportunidades e desafios a enfrentar pelos/as arquivistas, adivinhando bons frutos, especialmente no que respeita à consolidação do seu papel social e cultural. Segundo as suas análises, presentemente, encontram-se reunidas condições excelentes para a realização de uma convergência de interesses entre as diversas áreas do conhecimento e as igualmente diversas instituições envolvidas nesses imbrincados processos.

Como se evidencia, as análises anteriores vêm sendo reiteradas em alguma da literatura científica visível e especializada nessas matérias. Essas possibilidades, que se cruzam e que fertilizam reciprocamente as diversas áreas envolvidas, ajudam a clarificar que tipo de relação se pode estabelecer entre a Arquivística, as Tecnologias de Informação e a Curadoria Digital e o que esperar em relação aos contributos multidisciplinares a obter a partir desta confluência de saberes e de fazeres num futuro próximo.

Em boa verdade, desde há décadas os/as profissionais de arquivo constatarem ser impossível desenvolver os meios adequados para manter, indefinidamente, todos os documentos que possuía ao seu cuidado. O meio digital veio sedimentar esta constatação. Por um lado, aspetos tais como a alta densidade e a capacidade de armazenagem dos suportes, bem como a portabilidade, a fluidez, a duplicabilidade e a conectividade, que definem e caracterizam os documentos digitais, exacerbam realidades arquivísticas já de si complexas, das quais, até há pouco tempo, as Massas Documentais Acumuladas – ou MDA – eram o exemplo mais flagrante. O seu análogo mais próximo, atualmente, identificamo-lo no fenómeno do “*big data*”⁸. Para além de tudo isso, a manifesta dependência que as informações de carácter arquivístico têm em relação ao seu contexto de produção e às demais informações contidas nos processos de que fazem parte e com as quais criam indelévels laços, acrescida, ainda, da necessidade de preservação da sua integridade e da sua autenticidade, ao longo do tempo, impelem os/as arquivistas à busca de soluções que facultem a sua boa descrição e que possibilitem a sua boa compreensão, no futuro.

Lauriault, Craig, Taylor e Pulsifer (2007 citados por Akmon, Zimmerman, Daniels, & Hedstrom, 2011, p. 333) destacam o papel crucial que o/a arquivista pode ter na preservação dos dados científicos, desde que compreenda bem as “regras da casa”. No que respeita ao processo de seleção para a preservação, recomendam uma lógica diferente da habitualmente usada no âmbito dos arquivos, baseada sobretudo nas atividades do produtor. Ponderam que, nesses casos, será preferível procurar compreender as diferentes necessidades das comunidades de pesquisa, tendo como referências os/as próprios/as cientistas, que são os/as que produzem e que utilizam os dados sobre os quais deverão incidir as ações de preservação.

A nosso parecer, toda esta proposta de reorientação metodológica depende realmente do grau de integração dos/as arquivistas nesses ambientes. A adoção de uma abordagem naturalista, que parta da realidade vivenciada nos laboratórios, alinhada com interlocutores/as privilegiados/as, torna legítimo o processo decisório e torna as medidas de preservação reconhecíveis, coerentes e consistentes (cf. figura 1.1).

⁸ Ou grandes volumes de dados massivos produzidos a uma velocidade e com uma variedade exponenciais e sujeitos a procedimentos de análise, uso, reutilização e redistribuição.

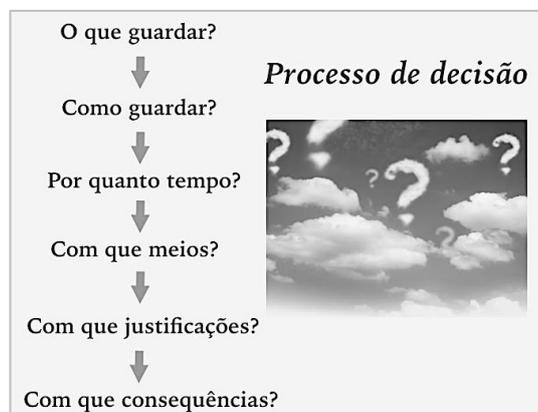


Fig. 1.1. Questões que permeiam as decisões de preservação.

Fonte: elaboração própria.

Todo esse naturalismo que contextualiza devidamente as ações e que responsabiliza os/as interlocutores/as diretos/as, a nosso parecer, foi sinalizado nas teses pós-modernas defendidas por Harris (1998) e resumidas no quadro 1.1.

Quadro 1.1. Pós-Modernismo e avaliação arquivística: sete teses

1. Os documentos resultam de um processo complexo de construção e no ato de avaliação são acrescentadas camadas substanciais a esse processo.
2. O/a arquivista é um coprodutor dos documentos de arquivo.
3. O/a arquivista cria e recria os valores arquivísticos, em cada exercício de avaliação.
4. Os documentos preservados no processo de avaliação dizem muito mais sobre o processo do que sobre o resultado.
5. Os/as arquivistas prestam contas, pelo que a avaliação deve inserir-se num quadro devidamente sustentado por teorias, métodos, políticas e programas.
6. A avaliação deve ser exercida de forma democrática, transparente, reflexiva e baseada em evidências e em justificações, e deve permitir, em qualquer momento, o rastreio das decisões tomadas.
7. A avaliação evidencia o poder exercido pelos/as arquivistas, pelo que as decisões tomadas no presente têm reflexos no futuro.

Fonte: Harris (1998).

A nosso parecer, as teses aduzidas no quadro 1.1, pela sua transversalidade, se quisermos, ajudam a compreender o alcance e as consequências dos princípios norteadores dos processos de avaliação dos dados científicos. Se os/as cientistas têm, evidentemente, muito a dizer aos/as arquivistas acerca do processo de seleção dos seus dados com vista à preservação, também os/as arquivistas, com os seus conhecimentos e a sua experiência acumulada nessas matérias, podem ajudar a compreender o contexto, o conteúdo e as estruturas envolvidas na produção dos dados, proporcionando uma análise mais alargada e justificada do processo. Acresce que a avaliação, sob esta perspetiva, decorre num *continuum* no processo de gestão dos dados, o que reforça ainda mais a sua importância (cf. figura 1.2).

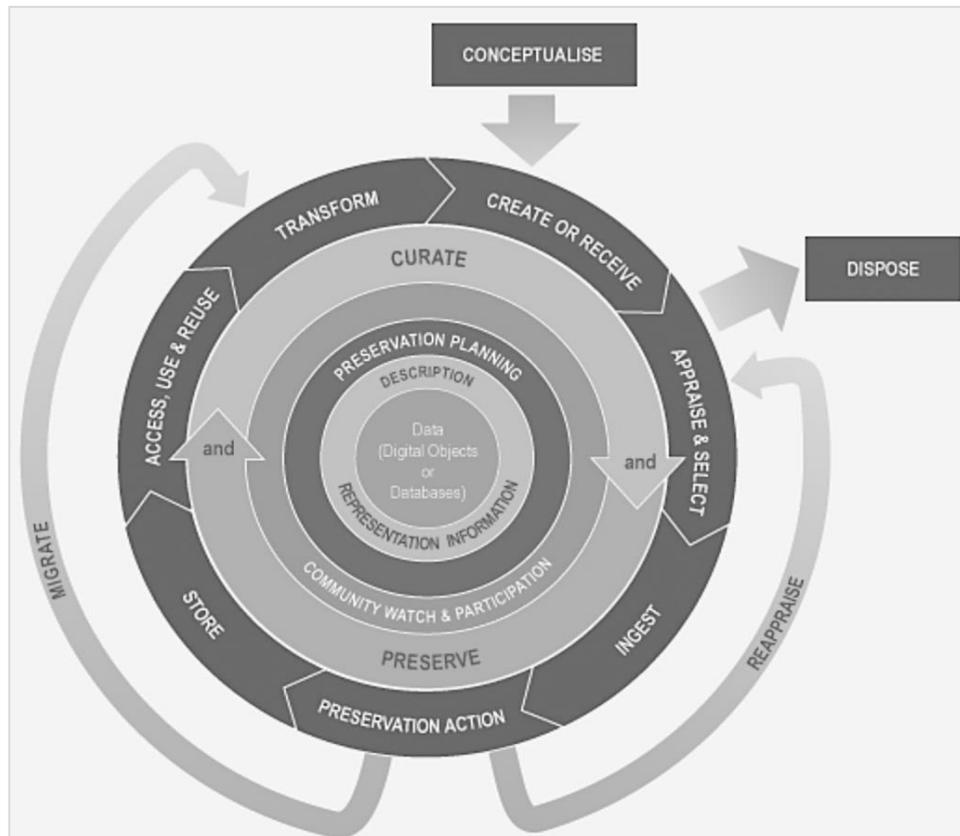


Fig. 1.2. DCC Curation Lifecycle Model.

Fonte: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>.

A noção de *continuum* leva, inevitavelmente, aos recentes desenvolvimentos da história intelectual da Arquivística, que na aceção de Ridener (2009) foi, nos últimos cinquenta anos, profundamente impactada pelas tecnologias de informação, tendo as suas teorias e práticas sido constantemente revisitadas e escrutinadas. Com efeito, as mudanças provocadas pelo advento do mundo digital, muitas vezes abruptas, alteram os cenários de atuação profissional e toldam a boa compreensão e o bom posicionamento dos/as envolvidos/as em relação à sua própria realidade. Essa dificuldade vem posicionado os/as arquivistas entre duas fronteiras aparentemente bem demarcadas e com características bem definidas: a moderna e a pós-moderna (cf. figura 1.3).

DEZ PONTOS DE DIVERGÊNCIA NAS ABORDAGENS ARQUIVÍSTICAS	
Moderna(custodial)	Pós-moderna (pós-custodial)
Guarda	Acesso
Recors management	Records continuum
Evidência	Informação
Imparcialidade	Parcialidade
Objetividade	Subjetividade
Custodiador/a (arquivista)	Facilitador/a (arquivista)
Orgânico (contexto)	Funcional (contexto)
Ênfase no resultado	Ênfase no processo
Exclusivismo	Transversalidade
Visão reducionista (micro-realidades)	Visão holística (macro-realidades)

Fig. 1.3. Dez pontos de divergência nas abordagens arquivísticas.

Fonte: Elaboração própria.

Compreender esses aparentes pontos de divergência patentes nessas duas formas de abordagem – a moderna e a pós-moderna –, que a nosso ver se complementam, porque a hibridez dos contextos arquivísticos atuais impele a diferentes tipos de lógicas e de esforços a envidar na prática corrente de quem lida diariamente com o documento, com a informação e com o conhecimento, é condição indispensável para se chegar a algum consenso sobre quem é, na atualidade, o/a arquivista e sobre com que realidade/s ele/ela próprio/a se debate no seu meio.

De acordo com Curry (2011), Feijen (2011) e King (2011), também citados no estudo seminal de Akmon, Zimmerman, Daniels e Hedstrom (2011, p. 330-331), ao descrever a sua experiência de uso e de gestão de dados, não raro, os/as cientistas mencionam explicitamente que necessitam dos/as arquivistas para incrementar essa sua experiência, o que quer dizer que eles/as se encontram atentos/as aos problemas gerados pela produção e pelo crescimento, muitas vezes exponencial, dos seus dados de pesquisa. A partilha desses dados, conforme recordam, envolve mais do que simplesmente torná-los acessíveis por meio de um *website* institucional. O investimento em profissionais experientes, em padrões e em formatos normalizados e internacionalmente aceites torna-se algo imprescindível num contexto em que os próprios dados se transformam rapidamente. Todo o conhecimento imprescindível de que é preciso acionar para equacionar esse problema compreende-se em torno do que envolve a literacia de dados (*data literacy*).

Considerada como um tipo especial de literacia de informação, a literacia de dados, de acordo com o abrangente estudo realizado por Calzada Prado e Marzal (2013, p. 126), desvela-se no conjunto das habilidades e das competências a somar para aceder, interpretar, analisar, criticar, gerir, manusear e utilizar eticamente os dados, devendo, quanto a nós, ser dado o indispensável realce a esta última componente da equação. Assim sendo, envolve um leque variado de conceitos e de estratégias, bem como o domínio de uma série de conhecimentos a avaliar nas mais diversas circunstâncias. Lidar

com os dados implica a boa compreensão dos meios adequados para os seleccionar e sistematizar, e para combiná-los com outras fontes de informação e com outros conhecimentos prévios existentes (cf. quadro 1.2).

Quadro 1.2. Literacia de dados: competências requeridas

- 1) Identificar o contexto no qual os dados são produzidos e reutilizados (ciclo de vida dos dados);
- 2) Reconhecer o valor, os tipos e os formatos dos dados;
- 3) Determinar em que momento os dados são necessários;
- 4) Aceder às fontes de dados apropriadas às necessidades de informação verificadas;
- 5) Avaliar, de um modo crítico, os dados e as fontes de dados;
- 6) Identificar e empregar métodos de pesquisa adequados;
- 7) Manipular e analisar os dados;
- 8) Apresentar convenientemente informações quantitativas (dados específicos, tabelas, gráficos, relatórios e similares);
- 9) Aplicar os resultados à aprendizagem, à tomada de decisão e à resolução de problemas;
- 10) Planear, organizar e autoavaliar no decorrer do processo.

Fonte: Calzada Prado e Marzal (2013, p. 126) [tradução nossa].

Como nota final, nesta primeira secção, notamos que provavelmente o mais interessante nessa nova experiência que estão a ter os/as profissionais de arquivo no apoio à gestão dos dados científicos é a que se revela nas mudanças já observadas no seu próprio contexto de trabalho. Os/as arquivistas têm vindo a sair daquela zona de conforto identificada com os ambientes burocráticos em que se encontravam e estão a ser atraídos/as para os contextos nos quais se deparam com vastos contingentes de dados a gerir e que se governam por um lógica diferente. Toda essa consciência de que há um novo papel profissional a ser desempenhado, em muitos sentidos inclusivamente mais flexível e mais exigente, tem levado os/as arquivistas mais bem posicionados/as a percecionarem já algum incremento da sua relevância na, e para a, comunidade.

2. Resolvendo a equação: algumas iniciativas no âmbito dos dados abertos

Um dos aspetos que se verifica em relação à maioria das iniciativas de dados abertos é que são patrocinadas pelos governos, preocupados em disponibilizar os dados que resultam da investigação científica financiada publicamente, o que constitui, aliás, um dos argumentos para que se considere que os/as contribuintes não devem ter de pagar mais do que uma vez pelos dados produzidos com os seus impostos, algo que é interpretado como uma defesa da utilização de licenças abertas apropriadas e que permitam a distribuição gratuita dos dados. Assim, pode constatar-se uma primazia das iniciativas em dois âmbitos: a Governamental/Administração Pública e a Académica/Investigação Científica (McDonald, & Léveillé, 2014).

Nesta perspectiva, autores como Vancauwenberghe e Cromptvoets (2018, p. 94) apresentam as principais tendências e os desenvolvimentos na governança de iniciativas de dados abertos. Assim, são dignos de nota:

- a) no âmbito da Gestão Estratégica: as estratégias, os planos de ação e as avaliações de dados abertos, operacionalizados por meio de documentos de política, anunciando a necessidade de dados abertos, o posterior desenvolvimento de estratégias e planos de ação de dados abertos, muitas vezes impulsionados por iniciativas internacionais, e as avaliações para monitorização da disponibilidade, utilização e/ou benefícios de dados abertos;
- b) no âmbito da tomada de decisão coletiva: o conselho de governança, os grupos de utilizadores/as de dados abertos, as consultas públicas, que incluem vários tipos de estruturas coletivas de tomada de decisão, o estabelecimento de grupos de utilizadores/as de dados abertos para os envolver ativamente na tomada de decisão, o estabelecimento de consultas públicas para envolver os/as cidadãos/ãs nas tomadas de decisão sobre iniciativas de dados abertos;
- c) no âmbito da atribuição de tarefas e responsabilidades: o estabelecimento de entidade coordenadora, a identificação de dados e a definição de conjuntos de dados de alto valor, assumidos pelo estabelecimento ou designação de instituição coordenadora central, a identificação e o inventário de dados do governo a serem abertos, e a definição de prioridade de conjuntos de dados de alto valor;
- d) no âmbito da criação de mercados: as plataformas de dados e os catálogos nacionais, as estruturas nacionais de licenciamento e os incentivos financeiros para dados abertos, exemplificados com o lançamento e a manutenção de plataformas e de portais de dados abertos, que fornecem acesso a dados governamentais abertos, a harmonização das condições de acesso, o uso através de estruturas de licenciamento, e o fornecimento de incentivos financeiros para estimular o fornecimento e o uso de dados abertos;
- e) no âmbito da cultura interorganizacional e da gestão do conhecimento: as diretrizes e os manuais de dados abertos, a formação sobre dados abertos, a sensibilização e o trabalho em rede, que inclui a facilitação de documentos de orientação e as ferramentas para apoiar os/as proprietários/as de dados na adoção de dados abertos, a formação presencial e à distância sobre competências e habilitações em dados abertos, as ações de sensibilização e de ligação em rede para as diferentes partes interessadas;
- f) e, no âmbito da regulação e da formalização: as leis e as determinações sobre dados abertos, consubstanciadas em políticas e em legislação de dados abertos, que formalizam os princípios-chave da governança de dados abertos num enquadramento vinculativo, o estabelecimento de regras e de regulamentos sobre dados abertos de diversas formas.

No **meio científico, académico e universitário**, do ponto de vista das iniciativas de dados abertos, verifica-se a existência de múltiplas entidades de ensino superior, que desenvolveram iniciativas disciplinares e interdisciplinares de *big data*, perçecionadas com as questões ligadas aos dados abertos científicos. Para Schwab e Zhang (2018),

instituições como as Universidades de Michigan, Berkeley, John Hopkins, Irvine e Iowa, reconheceram o enorme potencial dessas atividades e o seu contributo para o avanço da Ciência a longo prazo e a sua reputação como Universidades de Investigação pioneiras. Para estes autores, a busca por este reconhecimento levará provavelmente a que mais escolas de gestão e departamentos de empreendedorismo adiram a iniciativas de dados abertos, no futuro.

Um outro exemplo é apresentado por McLeod (2012), que participou em dois projetos de dados abertos, no âmbito do seu trabalho no JISC⁹, referentes à gestão de dados científicos na Universidade de Northumbria, com vista à melhoria dos sistemas e das práticas de gestão e dos resultados/produzidos da atividade científica. Pretendiam melhorar as infraestruturas, o planeamento e desenvolver ferramentas (modelos de negócio, análise e planeamento de custo-benefício) de gestão de dados de investigação; incentivar a publicação de dados por meio de melhores citações, ligações e integração; estimular a aquisição de ferramentas adequadas, entre académicos/as e pessoal de apoio à investigação nas universidades; e aumentar as competências e capacidade de gestão de dados de investigação através de formação focada na disciplina.

McLeod (2012, pp. 95-96) considera que os dados abertos e a partilha de dados oferecem grandes oportunidades para os/as profissionais de arquivo, seja pela necessidade de os manter por períodos de tempo cada vez maiores, seja pelo processo de criação e de gestão de dados e registos de informação. Estes/as profissionais podem trabalhar com planeadores/as de sistemas e com produtores/as de dados, para que os mesmos sejam capturados de maneira que permita torná-los acessíveis e utilizáveis, de forma legítima e apropriada, recorrendo ao uso de metadados pertinentes e de sistemas interoperáveis. Isto permite evitar as situações causadas pelos produtores/as de informação que não têm uma visão tão abrangente do ciclo de vida ou *continuum* dos registos (*i.e.*, dados/informações). Existem também questões referentes à conservação definitiva e à eliminação desses dados. As agências financiadoras da investigação, pelo menos no Reino Unido, estão a exigir a conservação a longo prazo de dados científicos, aparentando ser esta uma abordagem “geral”, envolvendo prazos de conservação aparentemente arbitrários. As perguntas que se fazem são, pois: será esta abordagem apropriada ou necessária para a publicação/partilha e conservação definitiva de dados? “Que registos [dados] precisam de ser conservados?”. No contexto do acesso aberto, o corolário é que os dados conservados devem ser partilhados para uso/reutilização. Além dos juízos filosóficos, teóricos e de valor, há questões relativas aos recursos a empregar na conservação e na preservação de dados. Assim: quem financiará a infraestrutura humana e física? Os órgãos de financiamento? As organizações de investigação? Os/as investigadores/as? O plano para o acesso aberto e para os dados abertos considera o acesso à informação para além das questões de transparência e de responsabilidade, incluindo o campo da inovação, da iniciativa, do crescimento económico e do retorno do investimento. Neste âmbito, esse plano abre perspectivas interessantes para os/as profissionais de arquivo.

⁹ Disponível em: <https://www.jisc.ac.uk/>.

Por seu lado, Walford (2018), ao investigar a intensificação das práticas relativas aos dados que ocorreram nas últimas décadas nas ciências ambientais, aborda uma recente iniciativa internacional relativa a dados de Observações Globais da Terra, partindo da análise crítica na perspectiva dos arquivos. Esta abordagem considera, no entanto, que um foco dessas práticas de dados infraestruturais que se cinja apenas a uma lógica arquivística corre o risco de perder de vista outros elementos importantes dos cenários científicos emergentes baseados em dados. Um desses elementos é a recolha de dados, que a autora argumenta dever ser também estudada como um processo de apropriação e de transformação crucial, para entender as relações que constituem o conhecimento científico contemporâneo. Este estudo, feito com base no trabalho de campo etnográfico conduzido no âmbito de um projeto de Sistemas Terrestres de grande escala, na Amazônia brasileira, em comparação com uma análise histórica das coleções coloniais britânicas nos séculos XVIII e XIX, permite estabelecer linhas entre a investigação sobre as iniciativas de dados abertos e os arquivos definitivos.

Dentro das iniciativas dos dados abertos relativos à **governança, ligadas à administração pública**, considera-se digna de nota uma iniciativa de 64 países que constitui uma Parceria de Governo Aberto, para demonstrar o seu compromisso em tornar abertos os seus dados para uso, reutilização e redistribuição, de acordo com os princípios norteadores dos dados abertos. Esta iniciativa pretende garantir compromissos dos governos para promover a transparência, a responsabilização, capacitar os/as cidadãos/ãs e explorar tecnologias para fortalecer a governança. Para participar nesta iniciativa, os países (e seus respetivos governos) devem cumprir com os critérios de elegibilidade e demonstrar o seu compromisso com os princípios do governo aberto nas áreas de transparência fiscal, acesso à informação, divulgação de rendimentos e bens, e envolvimento do cidadão. Tal implica que os governos e organismos da Administração Pública destes países desenvolvam iniciativas de Informação Aberta Governamental (Attard, Orlandi, Scerri, & Auer, 2015, p. 6).

Este tipo de iniciativas permitiu o desenvolvimento de investigação e reflexão sobre aspetos que estão intimamente ligados à gestão documental/arquivística. Essa reflexão permite que autores/as como Attard, Orlandi, Scerri e Auer (2015, p. 5-6) apresentem uma proposta de ciclo de vida, específica para os dados abertos governamentais, cujos processos dividem-se em três secções: a) pré-processamento ou preparação dos dados para publicação, que inclui os processos de criação, seleção, harmonização (com as normas) e publicação dos dados; b) exploração ou utilização dos dados, que inclui os processos de interligação (para contexto e interpretação), descoberta (por parte dos utilizadores), análise (utilização passiva) e exploração (utilização proativa); e c) manutenção dos dados para garantir a sua sustentabilidade, efetuada por processos de curadoria dos dados.

Attard, Orlandi, Scerri e Auer (2015, p. 17) identificam um conjunto de elementos ou atributos para efetuar a avaliação das iniciativas de governo aberto. No âmbito dos desafios técnicos, incluem aspetos ligados a: formatos dos dados, ambiguidade dos dados, detetabilidade/descoberta dos dados, representação dos dados, capacitação da instituição. Do ponto de vista legal e de política, mencionam: conflitos regulamentares, privacidade e proteção de dados, direitos de autor e de licenciamento, e riscos derivados da

responsabilização. Do ponto de vista económico-financeiro, existem as questões de orçamentação. A nível organizacional, referem: institucionalização, sobreposição de enfoques e apoio técnico. Os aspetos culturais incluem: motivação, sensibilização, participação do público e competição. Para além destes, referem as abordagens de classificação de publicação de dados (tecnológica, organizacional), orientações de publicação (Oito Princípios dos Dados Abertos Governamentais, Esquema de Cinco Estrelas para Dados Abertos Ligados, etapas de publicação de Dados Abertos Governamentais do W3C eGov Interest Group) (Attard, Orlandi, Scerri, & Auer, 2015, pp. 10-12). Finalmente, no âmbito da avaliação da qualidade dos dados, referem os aspetos ligados a: usabilidade, precisão, completude, consistência, atualidade, acessibilidade e abertura (Attard, Orlandi, Scerri, & Auer, 2015, pp. 12-13).

A investigação sobre o contributo da gestão documental em projetos de dados abertos também é feita ao nível da Administração Local, como é exemplo o projeto de dados abertos da edilidade de Girona, no âmbito do governo aberto (Casellas i Serra, 2013). O projeto é dirigido pelo gabinete do presidente da Câmara, coordenado pela unidade de análise territorial, em colaboração com os Departamentos de Tecnologias de Informação e de Gestão Documental e Arquivos. O objetivo refere-se ao cumprimento da legislação europeia e espanhola sobre a reutilização da informação e também as solicitações de maior transparência administrativa e de boa governança, pelo que o projeto teve em conta os indicadores definidos pelo Índice de Transparência dos Municípios de Espanha (International Transparency España, 2012). Dado que a legislação e as normas em questão somente estabelecem princípios gerais de ação e como os registos ou recursos de informação devem ser tornados públicos, verificou-se uma omissão quanto à indicação da informação que deve ser tornada pública, ou que metodologia ou procedimentos a seguir. Tal levou ao desenvolvimento de um modelo corporativo de projeto de dados abertos que integrasse explicitamente a perspetiva da gestão documental. Também permitiu que a unidade orgânica responsável pela função de gestão documental e arquivística reforçasse a sua posição no seio de uma organização de pequena ou média dimensão, e que diminuísse a perceção de que a sua colaboração se limitava à informação não-estruturada, nomeadamente de arquivo histórico. Este projeto permitiu chegar a um conjunto de conclusões de teor técnico-prático, úteis para outros projetos: a) o Quadro de classificação fornece informações valiosas para detetar conjuntos de dados suscetíveis de serem publicados, especialmente se identificarem os sistemas aplicativos específicos que produzem registos, devendo ser o elo entre a função, o registo de informação de arquivo e os dados relacionados com a sua produção; b) o Catálogo dos Sistemas Aplicativos deve identificar as relações e/ou ligações destes com funções e atividades identificadas no Quadro de Classificação, incluindo a estrutura e a descrição dos dados essenciais; c) o papel dos arquivos históricos tradicionais pode ser reforçado pela definição dos prazos de conservação dos conjuntos de dados a serem publicados e pela função de conservar os arquivos de dados abertos; d) o aumento anual de produtos *Open Data* criados poderia supor uma dificuldade para dar mais transparência às organizações, pelo que é necessário distinguir entre conjuntos de dados novos e mais antigos, orientando a gestão destes últimos para os arquivos; e) as estratégias de preservação de registos de informação de arquivo em bases de dados facilitam a seleção de

dados para projetos *Open Data*; f) certos processos de extração podem ser complexos e, muitas vezes, a organização pode considerar que não é crítico desenvolver estratégias para a preservação de registos de informação de arquivo em bases de dados; a inclusão da perspetiva de dados abertos na análise para implementar soluções de fluxos de trabalho pode facilitar a extração de dados, no futuro; g) a avaliação de dados significativos da tramitação dos processos documentais deve ser útil para fortalecer a criação e a formalização de registos de informação administrativa, podendo ser integrados nas estratégias de preservação da informação de arquivo nas bases de dados; h) a qualidade dos dados é um dos problemas mais relevantes em projetos de dados abertos, e só é possível resolvê-los no momento da criação e da captura da informação; i) em alguns casos, a dissociação de dados pode ser difícil, pelo que a gestão documental pode contribuir para a análise específica desses processos documentais e facilitar a introdução de diretrizes sobre entradas de informações, facilitando a dissociação automática; j) a descrição precisa do contexto de criação de conjuntos de dados permite a sua compreensão por parte dos/as utilizadores/as, além de contribuir para o reforço da sua autenticidade; l) o fornecimento de informação contextual para cada recurso de informação permite que os/as utilizadores/as avaliem sua precisão ou questionem sua credibilidade (Casellas i Serra, 2013, pp. 5-6).

Outro exemplo de investigação sobre esta temática surge por intermédio da conhecida iniciativa InterPARES Trust (ITrust)¹⁰ e apresenta-se como uma colaboração entre investigadores/as internacionais que tem como finalidade o desenvolvimento de um enquadramento para redes locais, nacionais e internacionais de políticas, procedimentos, regulamentos, normas e legislação relativas aos registos de informação de arquivo, eletrónicos e digitais, com o objetivo de garantir a confiabilidade pública baseada em evidências de boa governança, uma forte economia digital e memória digital persistente. Apresenta uma preocupação relativamente ao impacto das iniciativas de gestão documental/arquivos. Esta iniciativa tem desenvolvido um projeto de investigação, que pretende dar resposta a questões relativas ao desenvolvimento de especificações de seleção e de avaliação de informação documental e ao estabelecimento formal de tabelas de seleção de documentos. Essas questões prendem-se com os raciocínios e com os critérios que influenciam as especificações de seleção e avaliação desses tipos de iniciativas, com o nível de relevância que as abordagens atuais, em termos de desenvolvimento, aprovação e aplicação de tabelas temporais de seleção têm no universo dos dados abertos e do *big data*, e sobre as mudanças necessárias. Uma das abordagens utilizadas baseia-se especificamente na análise dos processos de negócio e fluxos de trabalho (McDonald & Léveillé, 2014, pp. 99-100). É no âmbito desta iniciativa que surge um ensaio relativo às questões da avaliação de informação arquivística no seio das iniciativas de dados abertos, baseado na experiência de conhecimento dos/as investigadores/as sobre numerosas instituições governamentais e do setor público canadiano, para ilustrar os problemas que muitas organizações enfrentam quando pretendem desenvolver iniciativas de dados abertos e *big data* (McDonald, & Léveillé, 2014, p. 106). Partindo do exemplo de um Ministério fictício, descrevem os processos de negócio desenvol-

¹⁰ Disponível em: <https://interparestrust.org/trust>.

vidos pela instituição, incluindo as funções, as atividades e as tarefas, e registos de informação (de arquivo) produzidos e que se constituem num largo conjunto de dados, com o fim de desenvolver uma iniciativa de *big data*, para a exploração de um elevado volume de dados que foram digitalizados, e cujos esquemas e normas de metadados de gestão permitem o seu relacionamento entre si e com outras fontes de dados de outros sistemas de informação. Todo este processo, desde a extração e o processamento inicial dos dados, até à produção dos conjuntos de dados, para uso interno e externo, é monitorizado por meio de um conjunto de procedimentos de controlo de integridade dos dados. No entanto, verificaram-se questões ligadas à seleção e à avaliação de registos de informação de arquivo, sobre o que deveria ser conservado, como deveria ser conservado e por quanto tempo. No âmbito desta iniciativa, os/as profissionais de arquivo/gestão documental tentaram resolver estas questões em conjunto, por meio da análise da situação e do desenvolvimento de recomendações, para que fosse possível identificar as relações entre os vários registos de informação de arquivo e verificar que estes apresentavam as qualidades de fiabilidade, rigor e autenticidade, de forma a que possam servir de evidência, apoiar a tomada de decisão, facilitar a responsabilização de pessoas de uma organização, respeitar as obrigações legais e os direitos adquiridos, e servir como ferramentas valiosas para circunstâncias e situações mais abrangentes que as razões que originaram a sua produção (McDonald, & Léveillé, 2014, pp. 107-110). A tentação para conservar todos os conjuntos de dados e os registos de informação gerados pelas iniciativas de *big data* e dados abertos trouxe questões relativamente ao que significa manter “tudo”. Assim, para prover a necessidade de formulação de especificações de seleção e avaliação, verificou-se a necessidade de analisar os processos de negócio que suportavam e estruturavam a iniciativa, para apoiar as decisões sobre os registos a conservar, em que forma e formato, como e por quanto tempo, e o seu destino após o fim do seu valor administrativo para a organização. Também se verificou que o foco deveria estar nos dados que deram origem à iniciativa, e que tinham sido gerados pelos processos de negócio centrais e, potencialmente, outros processos de negócio de suporte, sendo esta uma questão importante para fazer prova da integridade dos conjuntos de dados disponibilizados no portal. Em seguida, optou-se por definir que a metodologia a ser utilizada deveria basear-se na norma *ISO/TR 26122: Information and documentation – Work process analysis for records*, que permitiria efetuar uma análise funcional e sequencial de cada transação e processo como um todo, servindo como cenário para identificar os registos de informação de arquivo, definir requisitos para sua produção, captura e controlo e definir as especificações de seleção e de avaliação relevantes para o planeamento do processo e as necessidades e obrigações da função apoiada pelo processo de negócio (McDonald, & Léveillé, 2014, pp. 111-112). A definição de critérios para seleção e avaliação centrou-se não nos documentos, mas nas transações que constituem um determinado processo de negócio e que incluem transações informáticas e administrativas. Concluiu-se que as especificações de seleção e de avaliação precisavam incluir não só os prazos de conservação (abrangendo as formas de contagem dos prazos), mas também os formatos e outra informação (incluindo procedimentos), necessários para garantir a precisão e a autenticidade dos dados durante o período de tempo designado e que possam ser descartadas (eliminadas ou transferidas para um arquivo) de acordo com as

normas estabelecidas. Para além disso, o processo de desenvolvimento destas especificações deve ser incorporado em todas as etapas associadas ao estabelecimento de iniciativas de *big data*, dados abertos, e do próprio ciclo de vida de desenvolvimento de sistemas (planeamento, análise de requisitos, projeto, implementação, manutenção e revisão) associados ao desenvolvimento e à gestão dos processos de negócio centrais e de suporte. Assim, eram necessários metadados de qualidade usados para descrever, facilitar o acesso e a recuperação, proteger, monitorizar e controlar as ações tomadas, preservar pelo tempo necessário, e eliminar os dados. Neste aspeto, considerou-se que os metadados e o respetivo esquema de metadados utilizado nas iniciativas deveriam ser os instrumentos utilizados para garantir o registo e a justificação das especificações de seleção e de avaliação a serem adotadas. Definiu-se também que estas especificações de seleção e de avaliação deveriam estar incluídas na documentação dessas iniciativas e dos sistemas, que deveriam ser assinadas no âmbito dos processos de aprovação estabelecidos para o planeamento e para a gestão das iniciativas de *big data* e de dados abertos (incluindo os sistemas aplicativos), e que os processos de aprovação, monitorização e auditoria dessas iniciativas deveriam incluir as especificações de seleção e de avaliação, por forma a garantir que fossem respeitados os requisitos de preservação de dados a longo prazo. Paralelamente, verificou ser necessário definir a integração das especificações num enquadramento da gestão documental, que garantisse a captura, a seleção e avaliação dos registos de informação a que se referiam as especificações (McDonald, & Lévilé, 2014, pp. 112-115).

Conclusão

As abordagens verificadas na primeira secção – “Compreendendo a equação” – e que procuraram responder às questões que dizem respeito às relações a estabelecer entre arquivistas e cientistas no âmbito da gestão dos dados científicos, bem como qual é o contributo da Arquivística e dos/as profissionais de arquivo nessas matérias, evidenciam que, a partir do momento em que emerge o interesse pela curadoria dos dados científicos, também emerge a consciência da necessidade de envidar esforços para não perder, já no princípio, o controlo sobre o ciclo de vida dos dados e da informação, tornando-os disponíveis para novos usos, alterando-se, assim, a dinâmica de produção da ciência e com ela a dinâmica de utilização e preservação dos dados científicos. Igualmente, percebe-se que há um rol de funções arquivísticas clássicas que podem ser sujeitas a uma leitura adequada aos novos cenários desenhados pelos dados abertos, entre elas, muito provavelmente a mais crucial, pela sua transversalidade, a Avaliação. Ao engajarem-se nesses contextos, procurando compreender quais são os processos de produção, de uso e reutilização, bem como de distribuição e redistribuição dos dados, os/as arquivistas têm muito a aprender e a oferecer. Cite-se, como exemplo, os já consolidados critérios de avaliação e seleção de informação, baseados no ciclo de vida e no *continuum* dos documentos/informação, que vão particularmente ao encontro dos já sabidos e divulgados critérios de seleção definidos pela curadoria digital. Evidentemente, e sem querer simplificar demasiado a análise das partes envolvidas nesta equação, tanto os/as os/as

arquivistas como os/as cientistas de dados ganham largamente ao cruzar os seus conhecimentos e as suas perspetivas nesses processos.

As consequências que se invocam, por detrás dessas movimentações, têm como ponto de partida e como ponto de chegada o “*open government*”. Nessa arena, também os/as arquivistas têm, claramente, uma palavra a dizer e outra a receber. As oportunidades e os desafios trazidos pela ciência aberta e pelo movimento dos dados abertos, enquanto fenómenos suportados pela rápida evolução das tecnologias digitais, encontram, entre os/as profissionais de arquivo, algumas oportunidades evidentes, que se podem traduzir no incremento da sua visibilidade académico-científica e técnico-profissional.

As análises apresentadas na segunda secção – “Resolvendo a equação” – que procuraram responder à questão relacionada com os exemplos de iniciativas institucionais e supra-institucionais de dados abertos, realizadas com o envolvimento de profissionais e de investigadores em temas arquivísticos, reconhecidos ao nível internacional, evidenciam que a integração de profissionais, instituições e conhecimentos técnico-práticos e académico-científicos relativos à Arquivística, na vertente da gestão documental, são uma mais-valia para assegurar e sustentar a continuidade e a legitimidade destas iniciativas. Mas o benefício também se verifica na outra direção, dado que passa a haver um maior reconhecimento do papel e das potencialidades das instituições, um aumento da experiência e da competência dos/as profissionais, e um maior *apport* de conhecimento para a investigação e reflexão que permitem avançar a Ciência nesta área.

Finalmente, quanto às limitações e aos futuros estudos, esperamos investigar iniciativas experienciadas em contextos científicos mais próximos (*e.g.*, Portugal e Brasil), alargando assim o espectro a outras realidades arquivísticas.

Referências

- Akmon, D., Zimmerman, A., Daniels, M., & Hedstrom. (2011). The application of archival concepts to a data-intensive environment: working with scientists to understand data management and preservation needs. *Archival Science*, 11(3/4), 329–348.
- Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), 399–418.
- Borgman, C. L., Wallis, J. C., Enyedy, N. (2007). Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. *International Journal Digital Libraries*, 7(1–2):17–30.
- Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating Data Literacy into Information Literacy programs: core competencies and contents. *Libri*, 63(2). <https://doi.org/10.1515/libri-2013-0010>.
- Casellas i Serra, L. E. (2013). *The mapping, selecting and opening of data: the records Management contribution to the Open Data project in Girona City Council*. Apresentado na ICA Annual Conference, Bruxelas. Obtido de International Transparency España. (2012). *Índice de transparencia de los ayuntamientos*

- (*Transparency index of municipalities*). Obtido de <https://transparencia.org.es/ita-2012/>
- Dobreva, M., & Duff, W. M. (2015). The ever changing face of digital curation: introduction to the special issue on digital curation. *Archival Science*, 15(2), 97–100. <https://doi.org/10.1007/s10502-015-9243-7>.
- Harris, V. (1998). Postmodernism and archival appraisal: Seven theses. *South African Archives Journal*, 40, 48.
- Hey, T., Tansley, S., & Tolle, K. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research: Washington.
- Kassen, M. (2014). Globalization of e-government: open government as a global agenda; benefits, limitations and ways forward. *Information Development*, 30(1), 51–58.
- Lee, C. A., & Tibbo, H. (2011). Where's the Archivist in Digital Curation? Exploring the possibilities through a matrix of knowledge and skills. *Archivaria*, 72, 123–168.
- McDonald, J., & Léveillé, V. (2014). Whither the retention schedule in the era of big data and open data? *Records Management Journal*, 24(2), 99–121. <https://doi.org/10.1108/RMJ-01-2014-0010>
- McLeod, J. (2012). Thoughts on the opportunities for records professionals of the open access, open data agenda. *Records Management Journal*, 22(2), 92–97. <https://doi.org/10.1108/09565691211268711>.
- Nesmith, T. (2015). Toward the archival stage in the history of knowledge. *Archivaria*, 80, 119–145.
- Noonan, D., & Chute, T. (2014). Data Curation and the University Archives. *American Archivist*, 77(1), 201–240.
- Piwowar, H., Chapman, W. (2009). Public sharing of research datasets: a pilot study of associations. *Journal of Informetrics*, 4(2):148–156.
- Ridener, J. (2009). *From polders to Postmodernism: a concise history of archival theory*. Litwin Books: Sacramento.
- Schwab, A., & Zhang, Z. (2018). A new methodological frontier in entrepreneurship research: Big Data studies. *Entrepreneurship Theory and Practice*, 43(5), 843–854. <https://doi.org/10.1177/1042258718760841>.
- Vancauwenberghe, G., & Crompvoets, J. (2018). Governance of Open Data initiatives. In: B. van Loenen, G. Vancauwenberghe, & J. Crompvoets (Eds.), *Open Data Exposed* (Vol. 30, pp. 79–100). https://doi.org/10.1007/978-94-6265-261-3_5
- Walford, A. (2018). If everything is information. In: H. Knox, & D. Nafus (Eds.), *Ethnography for a data-saturated world*. Manchester University Press: Manchester. <https://doi.org/10.7765/9781526127600.00013>.
- Walters, T., & Skinner, K. (2011). *New roles for new times: Digital curation for preservation*. Association of Research Libraries: Washington. Obtido de http://www.arl.org/storage/documents/publications/nrnt_digital_curation17mar11.pdf.
- Warnow-Blewett, J., Genuth, J., & Weart S. R. (2001). *AIP study of multi-institutional collaborations: final report: highlights and project recommendations*. Ameri-

can Institute of Physics, College Park, MD. Obtido de <https://history.aip.org/history/pubs/collabs/highlights.html>

Zimmerman, A. (2008). New knowledge from old data: the role of standards in the sharing and reuse of ecological data. *Science Technologies & Human Values*, 33(5), 631–652.

Pre-Print