Versão

# PRESERVAÇÃO NA NUVEM: COMO SE PARECERÁ NO FUTURO UM SISTEMA DE PRESERVAÇÃO CONFIÁVEL?

**Luciana Duranti**

Luciana Duranti é docente de arquivística e diplomática no Master of Archival Studies na School of Information da University of British Columbia – Vancouver, Canadá. Suas pesquisas têm como foco a preservação da autenticidade de documentos e arquivos digitais em sistemas internos e ambientes em nuvem. Especialista em documentos diplomáticos e eletrônicos; desde 1998 exerce a função de diretora do projeto de pesquisa de documentos eletrônicos, InterPARES (Pesquisa Internacional sobre Documentos Arquivísticos Autênticos em Sistemas Eletrônicos).

## Introdução

William Lehr afirma que hoje a Internet é a "infraestrutura essencial" que assume o papel socioeconômico de "rede que originalmente dá o suporte para o acesso universal à telefonia", enquanto a "nuvem da Internet" possibilita "uma gama ampla de serviços dentro da rede, tais como: o acesso aos recursos computacionais, armazenamento on-line e outros serviços de alto nível, além dos serviços de transporte de dados"[1]. De maneira similar, Blanchette declara que a nuvem tornou-se "uma espécie de *meta-infraestrutura*" capaz de um crescimento sustentado sem precedentes"[2], em que a infraestrutura é "definida como os elementos de um ecossistema computacional que fornece ***serviços para*** *aplicativos*, em contraste com os aplicativos que oferecem *serviços para os usuários*"[3]. É por isso que os países estão começando a olhar para a nuvem como uma infraestrutura crítica, isto é, vital para o funcionamento de sua economia e sociedade. É, portanto, bastante lógico esperar que, no futuro, os sistemas de guarda e preservação documental estarão mais comumente, do que não, na nuvem. Se eles serão confiáveis ou até mesmo, de fato, "sistemas", ao invés de amálgamas constituídos variavelmente de serviços regulados por contratos padronizados, vai depender da habilidade dos profissionais de documentação[4] em desenvolver padrões internacionais para dados e documentos na nuvem e o impacto desses padrões na política governamental e na opinião pública.

1 Lehr, William. « Reliability and the Internet Cloud. » Em Yoo, Christopher S. e Blanchette, Jean-François, editores. *Regulating the Cloud. Policy for Computing Infrastructure.* Cambridge, Massachusetts e Londres, Inglaterra : The MIT Press, 2015), p. 105.
2 Blanchette, Jean-François. « Introduction ». Em Yoo, Christopher S. e Blanchette, Jean-François, editores. *Regulating the Cloud. Policy for Computing Infrastructure.* Cambridge, Massachusetts e Londres, Inglaterra: The MIT Press, 2015), p. 3.
3 Ibidem, p.5.
4 Neste contexto, "profissionais de documentação" (*record professionals*) é uma expressão utilizada genericamente para abranger gerentes de documentos, arquivistas, especialistas em governança da informação e todos aqueles profissionais cuja responsabilidade é gerenciar dados, documentos, registros ou arquivos, independentemente do contexto em que esses profissionais possam atuar.

## A Nuvem

Não há uma definição consensual de computação na nuvem. O que existe é apenas o reconhecimento de que se trata de um modelo de serviços que requer uma rede conectada e entregue em qualquer lugar para múltiplos usuários, não importa a localização destes ou os recursos dos provedores, oferecida sob demanda e paga proporcionalmente ao uso. Todavia, esse modelo pode ser modificado quando necessário, como por exemplo, ao oferecer o serviço apenas para um usuário específico, em uma única localização, fora da rede (off-line), por reserva ou por uma tarifa fixa. De fato, Weinman acredita que um enfoque híbrido é o melhor para a utilização dos serviços de rede[5]. Também o Instituto Nacional de Padrões e Tecnologia dos Estados Unidos (NIST), considera a computação em nuvem como um "paradigma em evolução", permitindo maneiras diferentes de combinar tecnologias novas e antigas[6].

Há muitas razões pelas quais profissionais de documentação têm estado relutantes em manter e preservar documentação em um ambiente de nuvem. Tais motivos dizem respeito à confiabilidade e à transparência do serviço, além de segurança, privacidade, controle e jurisdição[7]. Muito tem sido escrito sobre essas preocupações, assim como sobre os benefícios da nuvem, ligados principalmente a acesso, colaboração e vantagens econômicas[8]. Até o momento, tais benefícios não têm provado serem incentivos suficientes para a adoção geral dos serviços de nuvem, mas estão sendo desenvolvidos políticas, acordos contratuais, padrões de segurança e procedimentos de controle que podem servir de auxílio a uma mudança para a nuvem.

## Políticas

Muitos têm clamado por um arcabouço internacional coerente de políticas e estratégias governamentais que se dirija à jurisdição, segurança, privacidade e compartilhamento de riscos do ambiente de nuvem. Entre os envolvidos, a Comissão Europeia tem sido a mais ativa. Em sua Conferência de Segurança da Nuvem de 2015, foi acordado haver a necessidade tanto de abordagens políticas flexíveis que permitam o avanço tecnológico, como de um estreitamento de relações entre os setores público e privado, estabelecendo segurança em termos de redes, requisições de localização de dados, jurisdição estrangeira e acesso[9]. Em relação à privacidade, embora a Europa esteja desenvolvendo uma abordagem política unificada, será difícil harmonizá-la com a dos Estados Unidos, já que, na Europa, a privacidade é considerada um direito fundamental e um aspecto de dignidade (humana), enquanto, nos Estados Unidos, é parte da liberdade e uma commodity (mercadoria) alienável, de que se pode abrir mão para se obter um serviço customizado[10].

Virginia Greinman conduziu uma análise comparativa de estratégias nacionais de segurança de nuvem na Austrália, União Europeia, Japão, Singapura, Espanha, Reino Unido e Estados Unidos, fazendo uma série de

5 Weinman, Joe. «Cloud Strategy and Economics». Em Yoo, Christopher S. e Blanchette, Jean-François, editores. *Regulating the Cloud. Policy for Computing Infrastructure.* Cambridge, Massachusetts e Londres, Inglaterra: The MIT Press, 2015), pp. 25-28, 37-38.

6 NIST Grupo de Trabalho do Mapa de Padrões de Computação em Nuvem, "NIST Cloud Computing Standards Roadmap", NIST Special Publication, 500-291, versão 2, Departamento Norte-Americano de Comércio < Instituto Nacional de Padrões e Tecnologia, Julho de 2013.

7 Duranti, Luciana. "Preservation in the Cloud: Towards an International Framework for a Balance of Trust and Trustworthiness", em Katre, Ginesh e Giaretta, David (editores). APA/C-DAC International Conference on Digital Preservation and Development of Trusted Digital Repositories. 5-6 de Fevereiro de 2014, Nova Déli, Índia (Nova Déli: Excel India Publishers, 2014), pp 23-38.

8 Veja-se, por exemplo, Duranti, Luciana. «Archival Science in the Cloud Environment: Continuity or Transformation?» Atlanti, vol. 23 (2013) : 45-52 ; Duranti, Luciana e Rogers, Corinne. «Trust in digital records: An increasingly clouudy legal area». *Computer Law & Security Review 28.5 (Outubro de 2012): 522-531;* Duranti, Luciana e Rogers, Corinne. "Trust in online records and data". Em Lowry, James & Wamukoya, Justus (editores). *Integrity in Government through Records Management: Essays in Honour of Anne Thurston (*Farnham : Ashgate, 2014), pp, 203-216 ; Duranti, Luciana. «Digital Records and Archives in the Commercial Cloud». Em Yoo, Christopher S. e Blanchette, Jean-François, editores. *Regulating the Cloud. Policy for Computing Infrastructure. (*Cambridge, Massachusetts e Londres, Inglaterra: The MIT Press, 2015), pp. 197-214. Veja-se também *The Canadian Journal of Information and Library Science.* Edição especial sobre Dados, Documentos e Arquivos na Nuvem. Editora convidada: Luciana Duranti. Volume 39, Número 2, Junho de 2015.

9 Rede Europeia e Agência de Segurança da Informação (ENISA). *Security Framework for Government Clouds,* Fevereiro de 2015.

10 Renda, Andrea. "Cloud Privacy Law in the United States and the European Union. Em Yoo, Christopher S. e Blanchette, Jean-François, editores. *Regulating the Cloud. Policy for Computing Infrastructure. (*Cambridge, Massachusetts e Londres, Inglaterra: The MIT Press, 2015), pp. 135-164.

recomendações para o desenvolvimento de uma estratégia de nuvem unificada, que resulte em políticas nacionais consistentes. A primeira recomendação, suficientemente interessante para profissionais de documentação que continuam a se digladiar em seu próprio território terminológico, é desenvolver definições comuns para os termos usados com mais frequência, tais como "resiliência cibernética" (*cyber resilience*), que geralmente se refere à capacidade de manter a operação durante ataques, incidentes ou problemas técnicos. Essas definições deveriam ser acompanhadas por taxonomias e ontologias que apoiassem o desenvolvimento de uma lista uniforme ao longo de diferentes nuvens, países e continentes. A segunda recomendação é de desenvolver uma lista uniforme de ameaças e de agentes ameaçadores. A terceira é de estabelecer uma parceria entre países que partilhassem dos mesmos valores, como o da liberdade de expressão, acesso livre à informação e proteção da privacidade. A quarta é de uma identificação unificada dos agentes da nuvem (como, por exemplo, proprietários, provedores de serviço, vendedores, serviços de transporte, clientes, auditores e outras autoridades independentes de supervisão) em diferentes jurisdições, com suas responsabilidades e deveres. A quinta recomendação é a criação de um arcabouço de risco e desenvolver normas e princípios necessários para estabelecer e manter a ordem pública no ambiente de nuvem, a começar por padrões internacionais. A sexta e última recomendação é uma efetiva fiscalização da segurança da nuvem, baseada em transparência[11].

Ainda que as recomendações de Greiman sejam valiosas e um alvo a ser alcançado, a criação de políticas nacionais coerentes e um arcabouço regulatório internacional apoiado em padrões pode ajudar, ao assegurar que os agentes da nuvem entrem em acordos contratuais consistentes e completos, ou seja, por meio de ações completas.

## Acordos Contratuais

Um estudo conduzido pelo Projeto InterPARES Trust[12] comparou contratos de serviço de nuvem vigentes com requisitos de manutenção e preservação de documentos para determinar se tais requisitos são contemplados pelos contratos padronizados. Não foi surpreendente para os pesquisadores concluir que não há uma terminologia padronizada para se referir ao conteúdo do usuário ao longo de diferentes contratos de serviço de nuvem. Decidiram, então, adotar o termo "dados" como a menor unidade de informação para abranger qualquer tipo de conteúdo. Com base em uma revisão da literatura, concluíram também que os contratos de serviço de nuvem consistem em vários tipos de documentos legais: um documento geral especificando os serviços (como, por exemplo, os termos de uso); um documento para serviços específicos (como, por exemplo, o acordo de nível de serviço); vários tipos de documentos cobrindo áreas como privacidade e uso aceitável. O gerenciamento documental e os requisitos arquivísticos foram identificados com base nos padrões relevantes ISO e ARMA,[13] bem como em padrões europeus. Além disso, os pesquisadores examinaram as Diretrizes do Acordo de Padronização do Nível de Serviço da Nuvem (NT em inglês) da Comissão Europeia. Os requisitos identificados diziam respeito a: controle de acesso; proteção da

11 Greiman, Virginia. «National Strategies for Cloud Innovation and Security». Em Endicott-Popovski, Barbara (editora), *Proceedings of the 3rd International Conference on Cloud Security Management. Universidade de Washington, Tacoma, Estados Unidos, 22-23 de outubro de 2015* (Reading, UK: ACPI, 2015), pp. 46-57.

12 InterPARES Trust (ITrust 2013-2018 – www.interparestrust.org) é um projeto de pesquisa fundado pelo Conselho de Pesquisa de Ciências Sociais e Humanidades do Fundo Canadense de Parcerias. ITrust explora assuntos relativos a documentos digitais assegurados pela Internet, especificamente aqueles dentro de um sistema de armazenamento *multi-tenancy*, multi-jurisdicional, baseado em terceiros da Internet, ou seja, da Nuvem. O objetivo do projeto é gerar paradigmas teóricos e metodológicos para desenvolver políticas locais, nacionais e internacionais, procedimentos, normas, padrões e legislação, para assegurar confiança pública fundamentada em evidências de boa governança, uma economia digital forte e uma memória digital duradoura. A parceria de pesquisa abrange cerca de 50 universidades e organizações, nacionais e multinacionais, públicas e privadas, nos Estados Unidos, América Latina, Europa, África, Australásia (NT: região que inclui a Austrália, Nova Zelândia, Nova Guiné e algumas ilhas menores da parte oriental da Indonésia) e Ásia. Seus pesquisadores são especialistas em arquivologia, gerenciamento documental, diplomática, lei, tecnologia da informação, comunicação e mídia, jornalismo, ***e-commerce***, informática da saúde, segurança cibernética, governança e garantia da informação, forense digital, engenharia da computação e política da informação. ITrust é a 4ª fase da Pesquisa Internacional de Documentos Permanentes Autênticos em Sistemas Eletrônicos (InterPARES – 1998-2018, www.interpares.org), que desenvolveu o conhecimento essencial para a preservação de longo prazo de documentos autênticos criados e/ou mantidos em formato digital e forneceu a base para padrões, políticas, estratégias e planos de ação capazes de assegurar longevidade de tais materiais e a capacidade de seus usuários de confiar em sua autenticidade.

13 **NT**: *ARMA International* é uma instituição privada que tem o compromisso de fornecer à sua comunidade de membros e profissionais de gerenciamento de informações as melhores práticas no segmento. Ver mais em www.arma.org

privacidade; confiabilidade demonstrável e contínua; transparência do gerenciamento de conta, localização do servidor, destruição e recuperação de dados[14].

Com base em tudo isso, os contratos foram examinados em relação a fatores-chave: "propriedade dos dados", disponibilidade, recuperação e uso; retenção e disposição de dados; armazenamento e preservação de dados; segurança; localização e transferência de dados; fim de serviço ou término do contrato[15].

## Propriedade dos dados

No que diz respeito à propriedade dos dados, o problema identificado foi o seguinte: quando o usuário[16] confia seus dados a um provedor e usa sua plataforma e aplicativo para gerar dados adicionais, o provedor criará dados relativos a essas ações para processamento, gerenciamento de dados etc. Enquanto o conteúdo gerado e/ou armazenado na nuvem pelo usuário é de sua propriedade, os metadados criados pelo provedor não o são, e como o usuário deles necessita para provar a integridade dos dados, é essencial que os acordos contratuais determinem se e como o usuário tem o direito de acessar e utilizar os metadados do provedor.

## Disponibilidade, recuperação e uso

Em relação à disponibilidade e ao acesso, todo contrato deveria distinguir um do outro, já que são conceitos legalmente distintos: enquanto disponibilidade é um fato, acesso é um direito. Todavia, este último não pode ser obtido sem o primeiro. A legislação na América do Norte e Europa garante o direito à informação detida por organismos públicos e, por vezes, também por organizações privadas, e essa informação deve ser prestada dentro de um período específico. Quando os dados são armazenados em um ambiente de nuvem, "a disponibilidade dos dados armazenados implica também a disponibilidade da infraestrutura, *hardware* e *softwares*, os quais facilitam a recuperação e a legibilidade dos dados", já que dificuldades técnicas podem tornar o processo mais lento, levando o proprietário dos dados, obrigado a dar acesso às informações, a sofrer sanções[17]. Assim, os acordos contratuais devem especificar o grau de confiabilidade do provedor.

Enquanto "disponibilidade" é "a quantidade de tempo esperado que um sistema esteja em funcionamento" o que pode ser expresso estatística ou percentualmente, "confiabilidade" é a característica de se comportar consistentemente de acordo com as expectativas[18]. Por isso, um contrato deve considerar não apenas disponibilidade, mas também "consistência e precisão de acesso". Isto significa que cópias dos dados precisam ser distribuídas para diversos *data centers* – garantindo redundância, além de que as cópias precisam continuar consistentes durante o período em que os usuários acessam os mesmos dados ao mesmo tempo. Isto ainda não é atualmente possível, pois "os provedores não dispõem de acordos explícitos entre si que ajudem a assegurar a confiabilidade da Internet como um todo." Isto vai "requerer colaboração entre múltiplas autoridades regulatórias... assim como agentes-chave", incluindo provedores de serviço, usuários, "comunidades de segurança/proteção pública e comunidades internacionais de comércio e de padronização"[19]. Enquanto isso, usuários potenciais deveriam questionar se o provedor tem arquitetura e processos de serviços que assegurem confiabilidade e estratégias de respostas críveis se um problema acontece, e se é auditado por alguma autoridade.

## Retenção e disponibilidade de dados

O problema da retenção e da disponibilidade de dados é complexo, pois independentemente do que está incluído no acordo contratual, seu cumprimento é difícil de verificar. A razão é que a disponibilidade pode necessitar da transferência de um sistema para outro para retenção, o que pode envolver perda de

14 Bushey, Jessica ; Demoulin, Marie e McLelland, Robert. «Cloud Service Contracts : An Issue of Trust» *The Canadian Journal of Information and Library Science,* 39, no 2 (Junho de 2015) : 128-153.
15 Ibidem, 135.
16 No contexto desse capítulo, o termo usuário refere-se a qualquer cliente do serviço de nuvem.
17 Bushey, Jessica ; Demoulin, Marie e McLelland, Robert, op. cit., pp. 137-8.
18 Lehr, op. cit., p. 95.
19 Ibidem, pp. 100-101.D

autenticidade (ou seja, identidade e integridade) ou destruição, podendo gerar violações de confidencialidade ou privacidade, persistência de algumas cópias e respectivos metadados, além da persistência dos metadados gerados pelo provedor a partir dos dados do usuário. Contratos padronizados normalmente não contêm cláusulas relativas à retenção e à disponibilidade sistemáticas. No máximo, asseguram que os dados apagados do usuário se tornarão permanentemente inacessíveis dentro de 6 meses, o que não satisfaz os requisitos de manutenção dos documentos. Por isso, qualquer acordo contratual deve conter termos de uso que contemplem esses assuntos.

## Armazenamento e preservação de dados

O armazenamento e a preservação de dados impactam a qualidade dos dados e sua possibilidade de servir como fontes em geral ou como evidência legal, em particular, especialmente em jurisdições nas quais a autenticidade dos dados é uma inferência feita a partir da integridade do sistema onde os dados estão alocados. Acordos contratuais geralmente não especificam como os dados são mantidos ao longo da mudança de tecnologias e de formatos de dados. Tais acordos afirmam, em geral, que os usuários são responsáveis por fazer cópias de seus dados. Todos os procedimentos de preservação, incluindo armazenamento apropriado, cuidados, custódia e controle de dados, são referidos pelos provedores como "procedimentos de cópia ou ***backup***"[20].

## Segurança

Segurança, sob a perspectiva da manutenção dos documentos e da preservação de dados, tem a ver com a proteção dos dados contra acessos, usos, alterações ou destruições não autorizadas. O provedor deve ser capaz de produzir trilhas auditáveis, *logs* de acesso e de captura, mantendo e deixando à disposição metadados associados com o acesso, recuperação, uso e gerenciamento dos dados, além daqueles vinculados aos próprios dados. Todavia, contratos padronizados ligam medidas de segurança com os tipos de serviços oferecidos e as tarifas pagas pelos usuários, o que pode se tornar um problema para aqueles que escolhem a nuvem por razões econômicas. Apesar do fato de que a transferência de dados para um provedor não elimina a responsabilização do proprietário dos dados, o Direito, em geral, considera importante que a segurança seja incluída em termos técnicos, físicos e gerenciais nos acordos contratuais[21]. Por causa desse requisito, a Aliança pela Segurança da Nuvem (CSA em inglês) fornece, atualmente, orientações sobre segurança enquanto serviço, direcionada principalmente para sistemas amplos. Logicamente, tais sistemas superam a "complexidade e inconsistência" de múltiplos sistemas pequenos, já que "a escala possibilita economias em tudo, desde o monitoramento de irregularidades, até o recrutamento e treinamento de pessoal chave" e "melhora a funcionalidade – filtros de *spam*, detecção de intrusos e outras atividades, baseadas em análise *big data*." Todavia, é possível argumentar que "a vulnerabilidade cresce com a agregação de conteúdo e atividade" e que as preocupações com privacidade se tornam maiores com os grandes provedores[22].

## Localização e transferência de dados

O problema da segurança liga-se diretamente com o da localização dos dados e fluxo internacional de dados. Isto é uma preocupação tanto em termos de leis de proteção de dados quanto de leis estrangeiras que permitem que agências de investigação acessem dados guardados por provedores registrados em sua jurisdição ou que nela façam negócios regularmente. A localização dos dados pode ser também um critério ao se determinar a lei que se aplica em caso de litígio, embora os provedores normalmente escolham a jurisdição compatível com seu próprio sistema legal[23]. O fato de que a nuvem é, em si, independente de localização, fomentou um debate sobre o quanto seria desejável um movimento de limitação de dados

20 Bushey, Jessica ; Demoulin, Marie e McLelland, Robert, op. cit., p. 140.
21 Ibidem, p. 141.
22 Blumenthal, Marjorie. «Finding Security in the Cloud». Em Yoo, Christopher S. e Blanchette, Jean-François, editores. *Regulating the Cloud. Policy for Computing Infrastructure*. (Cambridge, Massachusetts e Londres, Inglaterra : The MIT Press, 2015), p. 64.
23 Goh, Elaine. «Clear skies or cloudy forecast ? Legal challenges in the management and acquisition of audiovisual materials in the cloud». *Records Management Journal* 24, 1 (2014) : 59.

além das fronteiras nacionais, mas a estratégia internacional está deixando de requisitar que os dados permaneçam na jurisdição de criação, subestimando, assim, a importância de acordos multilaterais entre países para colaboração em segurança. Blumenthal especula que os países deveriam olhar para a nuvem pública como uma "infraestrutura crítica". O questionamento da autora baseia-se nas premissas de que "o mercado para sistemas públicos de nuvem vão continuar a se desenvolver... problemas de incentivo... provavelmente continuarão a existir... e... progressos quanto à segurança continuarão lentos, dados os desafios persistentes em segurança cibernética e a inclusão de desafios novos associados à nuvem"[24]. Certamente, considerando que a infraestrutura crítica é constituída pelo que é mais importante para o funcionamento de um país, confiar a manutenção e preservação de documentação pública à nuvem pública, tanto apoiaria tal determinação dos governos, como facilitaria a escolha da nuvem pública para documentos e arquivos de negócios e organizações privadas. No entanto, as infraestruturas críticas dependem de outras infraestruturas e alguns serviços de nuvem dependem não somente de infraestruturas elétricas e de comunicação, mas também de outros serviços de nuvem. Existe potencial para que "nuvens federativas ajudem-se umas às outras ao compartilhar recursos por ocasião de uma crise" e "não é surpresa que uma nova linha de ofertas de serviços de recuperação contra desastres haja emergido", como escreve Blumenthal[25]. Fato é que a nuvem é a plataforma de escolha para aplicativos móveis e para os dados gerados por seu uso, assim como aqueles criados por aparelhos inteligentes em casa e no trabalho, e os produtores de documentos geram um percentual crescente de dados na nuvem pública. Por isso, é apenas questão de tempo para que os serviços de nuvem pública sejam considerados como críticos e para que tenhamos que confiar neles para a manutenção e a preservação dos dados gerados nas plataformas públicas de nuvem.

## Término de serviço, rescisão de contrato

A perspectiva de considerar a nuvem pública como uma infraestrutura crítica aliviaria receios ligados ao término de serviço e à rescisão de contrato. Atualmente, é possível que, se o provedor da nuvem deixar de existir ou encerrar um ou mais serviços (encerramento ou suspensão de serviços podem acontecer por violação dos termos, inatividade ou conveniência), os dados deixados para trás sejam deletados ou fiquem inacessíveis. Se contratos por serviços pagos estabelecem sua duração, serviços gratuitos não têm uma duração predefinida e podem fechar as contas de forma unilateral. Contratos padronizados normalmente solicitam que o usuário apague o *software* e os aplicativos, e podem impedir o usuário de acessar os dados deixados no provedor. Mesmo quando os dados são devolvidos ao usuário, não é assegurado que eles estejam em um formato utilizável e interoperável. Se o contrato é rescindido pelo usuário, a restituição dos dados pode ser cara e eles podem não estar em formatos acessíveis. Além disso, o usuário pode não ter o direito de acessar os metadados gerados pelo sistema para a sua manutenção de documentos ou para propósitos legais e pode não ter a garantia de que o provedor destruirá cada cópia dos dados alocados nos *data centers*. Portanto, a rescisão de contrato precisa ser abordada de maneira explícita e detalhada no acordo feito entre provedor e usuário.

A partir do estudo dos acordos contratuais e os temas que abordam ou deveriam abordar[26], fica claro que o aspecto da manutenção de dados na nuvem que é mais subestimado pelos provedores é a *preservação*. Preservar dados na nuvem pode ser um processo de caixa preta, no qual os profissionais de documentação podem saber o que eles colocam para ser preservado e o que querem acessar e recuperar – provavelmente as mesmas coisas que eles colocam – mas frequentemente não sabem qual tecnologia é usada pelos provedores para gerenciar, armazenar ou processar seus dados. Os servidores de nuvem pública podem nem saber onde os dados estão, podendo subcontratar outros provedores para alguns de seus serviços e efetivamente o fazendo e mantendo, potencialmente, servidores ou sendo registrados como servidores de países diferentes. Mesmo sabendo da tecnologia utilizada pelos provedores para preservação, esses profissionais de documentação responsável pelos dados não podem contar que os mesmos *hardware* e *software* permanecerão em serviço pelo tempo em que os dados necessitam ser preservados, ou que as tecnologias que os substituam serão compatíveis com as anteriores. Além disso, é improvável que eles venham a ter

24 Blumenthal, Marjorie. «Finding Security in the Cloud». Op. cit., p.65.
25 Ibidem, p. 68.
26 Esse estudo produziu uma lista de checagem de assuntos a serem abordados em contratos entre provedores de nuvem e usuários. Tal lista mostrou-se mais útil aos profissionais de documentação do que um modelo de contrato, dados os contextos organizacionais divergentes, capacidades e vulnerabilidades. Essa lista pode ser encontrada no *website* do Itrust, em https://interparestrust.org/trust/research_dissemination, dentro dos Documentos da Pesquisa InterPARES Trust, NA14.

a *expertise* necessária em todas ou mesmo na maior parte das tecnologias necessárias ou usadas para a preservação digital atualmente e no futuro, já que ninguém sabe como as tecnologias de informação e de comunicação vão evoluir ao longo do tempo. Preocupados com essas questões, os pesquisadores do InterPARES Trust concluíram que tanto provedores quanto usuários se beneficiariam do desenvolvimento de um padrão internacional para a Preservação como um Serviço Confiável (PaaST, em inglês).

## Preservação como um Serviço Confiável (PaaST)

O propósito do estudo chamado Preservação como um Serviço Confiável (PaaST) é o de determinar os requisitos necessários para que os dados sejam preservados confiavelmente na nuvem. Ou seja, tornar confiáveis os dados destinados a uma preservação de longo prazo em provedores de nuvem pública, com a expectativa de recuperar dados idênticos aos transferidos em todas as dimensões essenciais, ou dados cujas diferenças em relação aos inicialmente transferidos fossem conhecidas com suficiente clareza e precisão, para melhor avaliar sua adequação para qualquer intenção de uso.

Para enriquecer a formulação dos requisitos para a preservação de longo prazo, por meio do aumento de *expertise técnica que contribua para o seu desenvolvimento e para adequá-los à implementação e à articulação e aceitação como padrão, a* InterPARES Trust (ITrust) propôs ao Grupo de Gerenciamento de Objeto (OMG),[27] um consórcio internacional, aberto a membros, com padrões de tecnologia sem fins lucrativos, para usar os requisitos do PaaST como base para um software de preservação com padrão OMG. Tais padrões são ditados por vendedores, usuários finais, instituições acadêmicas e agências governamentais. Entre esses padrões, dois são especialmente relevantes para a iniciativa PaaST: a Linguagem Modelada Unificada – UML[28] e a especificação dos Serviços de Gerenciamento de Documentos – RMS[29] que define uma abordagem tecnologicamente agnóstica para o gerenciamento de arquivos digitais que são mantidos dentro de sistemas de negócios ou outros aplicativos. Um padrão de preservação seria complementar à especificação OMG RMS – embora generalizável para todo tipo de dados, sendo articulada em UML.

O OMG aceitou a proposta do ITrust e estabeleceu um grupo de trabalho, sob a égide da Força Tarefa do Compartilhamento Governamental de Informação e Domínio de Serviços, para apoiar o desenvolvimento do padrão de preservação. O grupo de trabalho OMG e os pesquisadores do ITrust podem se sobrepor como membros e têm trabalhado em paralelo. "A especificação do OMG tem sido elaborada como um Modelo de Plataforma Independente, dando liberdade aos desenvolvedores para escolher a abordagem e as tecnologias para desenvolver implementações", e é desenhada para manter controle inquebrantável sobre os documentos correntes de um produtor quando gerenciados em um sistema que implementa o padrão OMG RMS. O projeto PaaST se baseia em produtos anteriores desenvolvidos no curso dos três projetos InterPARES precedentes e no padrão do Sistema de Informação de Arquivo Aberto ISO (OAIS)[30]. Ele diverge do padrão OAIS, pois este é um modelo de referência que define as funções e informações necessárias para a preservação, mas não aborda como podem ser implementadas. Embora o PaaST mantenha-se neutro em relação aos métodos e tecnologias usadas para a implementação, ele tem sido desenvolvido para facilitar a produção de um *software* de implementação. O escopo do PaaST é mais estreito que o do padrão OAIS, especificamente excluindo funções cuja realização não é automatizada, tais como solicitar e negociar acordos de submissão na Entidade Funcional e Administrativa OAIS e a produção de recomendações e planos na Entidade Funcional de Planejamento de Preservação OAIS[31].

O projeto PaaST começou por identificar os blocos construtivos da preservação, como segue:

- o conhecimento do que vai ser preservado e de suas propriedades significativas;

27 **NT**: em inglês, *Object Management Group*, é uma organização internacional que aprova padrões abertos para aplicações orientadas a objetos; ver mais em: https://www.omg.org/
28 **NT**: em inglês, *Unified Modeling Language* é uma organização internacional que aprova padrões abertos para aplicações orientadas a objetos; ver mais em: https://www.omg.org/
29 **NT**: em inglês, *Records Management Services*.
30 Organização de Padrões Internacionais. Space data and information transfer systems – Open Archival Information System (OAIS) – Reference model. ISO 14721 : 2012. http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284
31 Duranti, Luciana; Jansen, Adam; Michetti, Giovanni; Mumma, Courtney; Prescott, Daryll; Rogers, Corinne; e Thibodeau, Kenneth. «Preservation as a Service for Trust (PaaST)». Em Vacca, John R. (editor). *Security in the Private Cloud* (CRC Press - uma publicação de Taylor & Francis Group, LLC), no prelo.

- o conhecimento de como os dados são codificados;
- o conhecimento de como os dados são armazenados;
- a habilidade de aplicar requisitos específicos de preservação, especialmente aqueles relativos a propriedades significativas, em quaisquer mudanças, sejam de dados, de *hardware* ou de *software* dos quais dependam, ou ambos, para a manutenção da autenticidade;
- o conhecimento de que quaisquer mudanças no armazenamento, codificação ou das tecnologias usadas no processamento dos dados não tenham nem corrompido nem impedido de visualizar os dados apropriadamente;
- a manutenção de ligações corretas e completas entre dados relacionados, e entre eles a informação contextual sobre os mesmos;
- o conhecimento de como produzir cópias autênticas dos dados preservados.

Estes blocos construtores são a base dos requisitos funcionais do PaaST, que define as capacidades, os metadados e todas as outras informações necessárias para preservar os dados e produzir cópias autênticas deles. Os requisitos têm a intenção de serem aplicados em uma gama de situações, "permitindo tanto a alocação de diferentes tarefas de preservação a agentes diferentes como a execução dessas tarefas por um ou mais agentes, usando diferentes métodos e tecnologias. Portanto, os requisitos de preservação estão articulados como serviços, ou seja, conjuntos de capacidades relacionadas que podem ser executadas usando tecnologias potencialmente diferentes e não relacionadas, sob controles operacionais ou administrativos separados e independentes"[32]. Como os requisitos PaaST não assumem ou requerem que as atividades ou controles de preservação sejam implementadas em um sistema integrado, o contexto geral de processos de preservação é chamado de ***ambiente de preservação***, em vez de ***sistema de preservação*** e é definido como a totalidade das infraestruturas tecnológicas e as ferramentas utilizadas na preservação digital. No entanto, isto não exclui a possibilidade de ter um sistema de preservação abrangente e integrado. Além disso, as responsabilidades de preservação podem ser distribuídas de tal modo que algumas atividades podem ser realizadas internamente (*in house*), enquanto outras são executadas por um ou mais provedores de nuvem (como, por exemplo, um provedor poderia ofertar armazenamento e gerenciamento de dados e outros serviços especializados, como a migração de mídias ou conversão de dados, em uma base necessária).

Os requisitos de preservação estão agrupados em conjuntos de recursos relacionados e são chamados de *serviços*. Cada serviço – além dos recursos específicos dentro deste – poderia ser executado usando tecnologias diferentes e potencialmente não relacionadas, sob um controle operacional separado e independente. Os serviços de preservação são constituídos por:

- *Requisição de entrada/submissão,* que absorve dados em um Ambiente de Preservação;
- *Caracterização*, que explora as propriedades técnicas, arquivísticas e representacionais dos dados;
- *Autenticidade*, que captura e relata informações sobre a identidade e a integridade dos dados, além de aplicar métodos de autenticação;
- *Armazenamento para Preservação,* que controla o armazenamento dos dados para manter sua identidade, previne que se corrompam e satisfaz outros requisitos de preservação;
- *Mudanças na Preservação,* que gerencia mudanças tecnológicas, como o formato de migração ou substituição do *software,* para assegurar sobrevivência e usabilidade;
- *Acesso,* que possibilita entregar cópias dos dados.

Os serviços não são, necessariamente, independentes. Como exemplo, a *Caracterização* seria chamada, normalmente, de Requisição de entrada/submissão, para decidir aceitar ou não um conjunto de dados submetidos à preservação. Analogamente, se requisitado por um cliente de acesso, a *Autenticação* pode ser necessária para permitir a avaliação das cópias entregues ao usuário.

"O PaaST não inclui serviços que não sejam devotados especificamente à preservação digital, ainda que possam estar relacionados de perto, ou então serviços gerais possíveis de se esperar de um provedor de serviço, não importa se os objetos informacionais foram pensados para uma preservação de longo prazo. Os Serviços Gerais poderiam abranger telecomunicações, gerenciamento de dados, *login* de sistema, segurança, ferramentas de busca genérica, subsistemas de armazenamento, ferramentas de transferência de arquivos etc. Os Serviços de Preservação farão uso frequente de tais serviços gerais, mas sua presunção

32 Ibidem.

de disponibilidade significa que não precisam estar articulados com os requisitos do PaaST. Um exemplo comum de um serviço relacionado que não está incluído no PaaST é o sistema de gestão de acervos, muito usado por instituições como arquivos e bibliotecas para gerir os instrumentos de pesquisa de materiais pelos quais são responsáveis"[33].

Aos agentes que prestam os serviços são atribuídos quatro *papéis* principais na preservação digital: os três primeiros estão voltados para suprir o serviço (preservação) e o quarto, para acessá-lo: o *Detentor inicial* é a parte que, pelo menos no início das atividades de preservação, detém, possui ou controla os dados passíveis de preservação; o *Diretor de Preservação* é a parte que tem a responsabilidade de preservar os dados; o *Provedor de Serviço de Preservação* é a parte que fornece recursos tecnológicos e serviços para conduzir a preservação digital; e o *Cliente de Acesso* é a parte que deseja obter, ou efetivamente obtém acesso aos dados preservados. Há um agente secundário, o *Remetente* (*Submitter)*, parte que realmente envia os dados para preservação ao Provedor de Serviço de Preservação e é o interlocutor padrão para questões ou problemas que o Provedor de Serviço tenha com o envio.

Os requisitos funcionais para cada serviço são numerados de forma exclusiva. Algumas precondições precisam estar presentes para que o serviço possa funcionar. Um fluxograma principal detalha as operações sequenciais do serviço e fluxogramas alternativos abordam erros e condições excepcionais que possam surgir. O PaaST é respaldado por termos e condições específicas para a preservação de qualquer agrupamento de dados, classificados debaixo de dois títulos: um *Acordo de Preservação* e um *Contrato de Serviço de Preservação.* O diagrama UML é acompanhado por casos de utilização e outras documentações. As operações e os atributos seguem convenções terminológicas padronizadas que dão apoio ao acesso e à reutilização em outras operações e serviços.

O objetivo do projeto ITrust é gerar os paradigmas teóricos e metodológicos necessários para desenvolver políticas locais, nacionais e internacionais, bem como procedimentos, normas, padrões e legislação capaz de assegurar credibilidade pública aos documentos digitais. A pesquisa, até o momento, levou a resultados que permitem vislumbrar um sistema de preservação que seja capaz de existir em uma ou múltiplas nuvens, bem como em um ambiente híbrido. Esta visão está refletida no desenho de um ambiente ideal de preservação confiável, por meio da organização de requisitos, recursos, agentes, ações, entidades e relacionamentos em um modelo UML que possa acomodar uma larga variedade de contextos. A pesquisa de Preservação como um Serviço Confiável (PaaST) está completando o desenvolvimento de um grupo de Serviços de Preservação que assegure a autenticidade contínua de qualquer tipo de informação carregada na nuvem.

## Conclusão

Como se parecerão os sistemas de preservação confiáveis, no futuro? Provavelmente, não se parecerão com sistemas no sentido tecnológico do termo. Em vez disso, serão parecidos com agrupamentos de partes conectadas, formando um todo complexo que, espera-se, seja governado por um conjunto comum de princípios, regras e procedimentos. Serão híbridos, compreendendo serviços na nuvem e internos. Se vão ser coerentes, integrados, interdependentes e interoperáveis, resilientes, disponíveis e confiáveis, isto vai depender dos desenvolvimentos tecnológicos futuros e das vantagens econômicas.

A confiabilidade da manutenção e da preservação documental nesses sistemas estará diretamente ligada à confiabilidade dos provedores de serviços e à segurança da infraestrutura da arquitetura de nuvem e suas operações. Se tais sistemas serão efetivos e por si só considerados eficazes na manutenção e preservação de documentos, isto vai depender da capacidade dos provedores de nuvem de se interconectarem tanto verticalmente (com provedores de serviço especializado apoiados em provedores maiores) como horizontalmente, como uma federação dando suporte não apenas à disponibilidade por meio da redundância, mas também ao acesso universal a todos os tipos de intercâmbios. Quanto à autenticidade do material que será confiado a esses sistemas, só podemos desejar que uma pesquisa internacional e interdisciplinar contínua seja capaz de assegurar que todos esses mecanismos de controle e segurança de dados continuem sendo uma preocupação central.

33 Ibidem.