

## O CIENTISTA DE DADOS ARQUIVÍSTICO:

o encontro da arquivística com as ciências de dados

Rogério Ferraz Garcia de Andrade<sup>1</sup>  
rogerio.ferraz@ironbit.com.br

### Resumo

Com o surgimento de uma nova geração de registros digitais, que ainda não encontram respaldo nas definições oficiais de documentos arquivísticos, mas fazem parte da realidade objetiva das instituições, abre-se uma lacuna para a emergência de um novo profissional arquivístico. Este, cotejando as *ciências de dados*, ou seja, compreendendo e aplicando os seus constructos conceitual e ferramental, é fortemente desafiado pela dinâmica imposta social e tecnologicamente. Uma das problemáticas a ser solucionada refere-se aos documentos produzidos em bases não estruturadas, que se caracterizam por estruturas fluídas, flexíveis e dinâmicas, pois alguns princípios arquivísticos, teorias e arquétipos jurídicos estão sendo disrompidos e deverão ser expandidos, repensados ou substituídos. Neste estudo, utilizamos como exemplo, as mídias sociais e os seus documentos digitais não estruturados, visando a evidenciar o novo campo de trabalho do cientista de dados arquivístico. Em um ambiente em evolução disruptiva, as mudanças de paradigma levarão à adequação de expectativas dos profissionais com formação arquivística (proveniente das ciências sociais) com aqueles formados nas ciências exatas ou que atuam no campo da tecnologia da informação. Para o futuro próximo, o melhor dos mundos seria a formação deste emergente profissional, o *cientista de dados arquivístico*, desde os estudos da graduação, incorporando os conhecimentos matemáticos e estatísticos e as variadas técnicas de desenvolvimento de soluções das ciências da informação e tecnologia.

**Palavras-Chave:** Mídias Sociais; Profissional de Arquivos; Ciência de Dados; Documento Digital não Estruturado; Cadeia de Custódia.

<sup>1</sup> Arquivo Público do Estado de São Paulo, São Paulo - SP. BRASIL

## 1 INTRODUÇÃO

A regulamentação da profissão de Arquivista e de Técnico de Arquivo foi regulamentada pela *Lei n° 6.546*, de 4 de julho de 1978, que estabelece os requisitos para o exercício das profissões e as suas principais atribuições. Segundo o seu artigo segundo, algumas atribuições dos profissionais arquivistas são (BRASIL, 1978):

*III - planejamento, orientação e direção das atividades de identificação das espécies documentais e participação no planejamento de novos documentos e controle de multicópias;*

*IV - planejamento, organização e direção de serviços ou centro de documentação e informação constituídos de acervos arquivísticos e mistos;*

*VI - orientação do planejamento da automação aplicada aos arquivos. (grifos nossos).*

Chama a atenção, que no longínquo ano de 1978, o legislador observou com bastante atenção temáticas que incorporam aspectos de inovação sistêmica à atividade arquivística. Como destacado no excerto acima, estão incorporadas no texto legal três importantes visões: **(a)** planejamento de novos documentos (inciso III), pois era observado a evolução dos microfimes, como também a introdução de documentos impressos eletronicamente; **(b)** planejamento e direção de estrutura de informação (inciso IV), antevendo-se que os arquivos deveriam ser fonte de informação tempestiva e relevante para os futuros sistemas de gestão documental; e **(c)** planejamento da automação aplicada aos arquivos (inciso V), que colocando-se em consonância com o movimento de informática que emergia na década de 70, observava a necessidade de automação das atividades dos arquivos (de gestão, informacional e de difusão) em virtude do crescente volume de documentos. Automação aqui, em um sentido *lato sensu*, poderia ser entendida também como a gestão computacional autônoma que atualmente se discute entre diversas disciplinas da tecnologia da informação.

Treze anos depois, a Lei n° 8.159/1991, que dispõe sobre a *Política Nacional de Arquivos*, tanto públicos como privados, legislou que *é dever do Poder Público a gestão*

*documental e a de proteção especial a documentos de arquivos*, postulados em no artigo primeiro. Com o advento da lei, houve também a introdução do conceito de *documento de arquivo*, conforme o artigo segundo (BRASIL, 1991):

*Consideram-se arquivos, para os fins desta lei, os conjuntos de documentos produzidos e recebidos por órgãos públicos, instituições de caráter público e entidades privadas, em decorrência do exercício de atividades específicas, bem como por pessoa física, qualquer que seja o suporte da informação ou a natureza dos documentos. (grifos nossos).*

Até 1991, a discussão de *qualquer suporte da informação ou a sua natureza* era mais simples do que observamos atualmente. O Brasil, por sua história ainda jovem, constituíra no papel (celulose), majoritariamente, os seus registros históricos.

Trazendo a discussão para a realidade atual, os suportes documentais aumentam sua complexidade em várias ordens de grandezas e em escalas exponenciais de volume. Vou tentar evidenciar esse fenômeno duplo com alguns exemplos de novos suportes: sites e portais; banco de dados não relacionais, imagens de todos os tipos (infravermelho, satélite, aerográfica, cartografia vetorizada, 3D *etc.*), animações, gráficos em 2D com linhas de evolução temporal, *podcasts*, *audiocasts*, entre outras.

Há ainda um outro ponto importante que diferencia os novos suportes dos documentos, além da sua complexidade implícita e do volume em níveis astronômicos, com o passado recente: o paradigma dos dados não estruturados. Este conceito refere-se documentos digitais que são constituídos de dados que têm estrutura fluída, flexível e dinâmica, ou ainda, podem até não possuir estrutura nenhuma. São exemplos as mídias sociais, que possuem os mais variados suportes e seu conteúdo sendo ampliado a todo instante, em uma concepção que pode tornar-se infinita.

Impulsionados pela nova dinâmica social, profissionais de Arquivo e de Ciência de Dados estão se debruçando sobre as atividades relativas a boa técnica arquivística,

ou seja, em como identificar, organizar (classificar), avaliar, armazenar, preservar (ou descartar) e restaurar documentos produzidos em bases não estruturadas.

Olhando em retrospectiva, a legislação, em nível macro, anteviu várias inovações que surgiriam no mundo real das atividades humanas. No nível operacional, espera-se que a *Arquivística*, as *Ciências da Informação* e a *Tecnologia da Informação* respondam aos novos desafios ambicionados pela sociedade.

## 2. MUDANÇA DIGITAL

A mudança tecnológica do padrão celulose para o paradigma digital trouxe inéditos desafios à *Ciência Arquivística*. Antes a informação era somente registrada em apenas um suporte, o próprio papel. Com o paradigma digital, o suporte físico foi transformado em objeto conceitual, pois seus componentes (estrutura, conteúdo e contexto) não estão mais reunidos no próprio documento, mas dispersos em diversas formas de armazenagem e em diferentes lugares físicos (FONSECA, 2005).

O instrumento que incorpora essa transformação paradoxal, no campo arquivístico, é o documento digital arquivístico, cuja definição é *a informação registrada em dígitos binários, acessível e interpretável por meio de sistema computacional* (CTDE, 2009 apud CONARQ, 2013). Algumas de suas características são bastante singulares:

- i. **Participar de uma ação.**
- ii. **Forma documental fixa:** apresentação da mesma forma que tinha quando o documento foi armazenado.
- iii. **Conteúdo estável:** o documento digital tem que permanecer completo e inalterado.
- iv. **Relação orgânica:** relação explícita com outros documentos que participam da mesma ação, dentro ou fora do sistema.

Suas importantes especificidades são:

- i. **Documento digital não é virtual:** está fixado em um suporte (disco rígido, CD, DVD, entre outros).
- ii. **Conteúdo e suporte são entidades separadas:** o documento não se define pelo suporte.
- iii. **O documento digital é um objeto físico (suporte), lógico (software e formatos) e conceitual (conteúdo).**
- iv. **Fragilidade intrínseca do armazenamento digital,** pela degradação física do suporte.
- v. **Rápida obsolescência da tecnologia digital:** hardware, software e formatos;
- vi. **Instabilidade:** dificuldade em garantir a autenticidade dos documentos.

Entendendo-se a natureza do documento digital, o próximo passo passou a ser a sua gestão, de acordo com princípios arquivísticos. O *Conselho Nacional de Arquivos* (CONARQ) ciente dessa problemática, lançou em 2011 um manual de modelos de requisitos para o desenvolvimento de softwares. A obra, chamada de *e-ARQ Brasil: Modelos de Requisitos para Sistemas Informatizados de Gestão de Documentos* abarca os seguintes objetivos estratégicos (CONARQ, 2011):

- i. Especifica as condições a serem cumpridas pela organização produtora/recebedora de documentos e pelos próprios documentos, a fim de garantir sua confiabilidade, autenticidade e acesso ao longo do tempo.
- ii. Orientar a implantação da gestão arquivística de documentos.
- iii. Fornecer as especificações técnicas e funcionais, e os metadados, para orientar a aquisição e/ou a especificação e desenvolvimento de sistemas informatizados de gestão arquivística de documentos (SIGAD).

Os documentos gerados ou incorporados ao SIGAD garantem a implantação de programa de gestão arquivística de documentos, tanto para a modalidade papel quanto para os digitais, segundo os requisitos normativos do e-ARQ Brasil. Acreditando-se

nessa sistemática, os documentos produzidos dentro do ecossistema do SIGAD são confiáveis e unívocos, aptos para a gestão de documentos arquivísticos e seguros instrumentos de fonte de prova jurídica e de autenticidade (FLORES, ROCCO e SANTOS, 2016).

Como o mundo gira incessantemente e a habilidade humana para criar novas formas de comunicação parem ser infinitas, novas forma de registros digitais estão surgindo. São exemplos categóricos: *audiocast* de professores aos seus alunos via o *WhatsAPP*, um chat no *Tinder*, ou no *Telegram* ou no *Spotfi*; post no *Twitter* ou *Tic Toc* do presidente da empresa anunciando uma nova ação corporativa; comentários (positivos e negativos) de clientes no *Facebook* ou *Instagram* que geram contagem de *scores* de satisfação (NPS); contrato de compra de imóvel assinado no *Blockchain*; vídeo corporativo de boas-vindas a funcionários no *Youtube* ou no *Vimeo*; portal *web* da instituição que apresentam novas vagas de emprego; aulas no EaD corporativo, entre outras milhares de ações institucionais que são realizadas diariamente.

Com a emergência de uma nova geração registros digitais, que ainda não encontram respaldo nas definições estritas de documentos arquivísticos oficiais, passaram a fazer parte da realidade objetiva (marketing, finanças, vendas, comunicação, regulamentações, até nuncios de demissões ocorreram no serviço público, entre outras) das instituições, não sendo mais apenas uma panaceia de adolescentes ou de *geeks*.

Do exposto, acredito que os SIGADs, e seu constructo conceitual aplicado, não comportam (ou para usar um termo caro à *Arquivística*, não interoperam) essas novas manifestações tecnológicas e, portanto, não conseguem processá-las na sua natureza original, apenas por meio de cópias autenticadas. É nesse momento especial que os profissionais arquivísticos devem olhar para um novo horizonte epistêmico: *a ciência de dados* (ANDRADE, 2021).

### 3. UM NOVO PROFISSIONAL ARQUIVÍSTICO

Independentemente de existir um movimento para o surgimento de um novo profissional que esteja cotejando as *ciências de dados*, como constructo conceitual e ferramental, em apoio às atividades arquivísticas, alguns autores questionam o atual paradigma da formação de arquivistas. Destacamos:

**O arquivista [...] começa a se desvincular do perfil tecnicista, para assumir o de produtor de conhecimentos novos.** O arquivista deve ser capaz de analisar criticamente a realidade que se coloca, à luz de conhecimentos teóricos e de atuar com competência de modo autônomo e consequente (RODRIGUES, 2008). (grifo nosso).

E,

**O arquivista hoje não pode esquecer que vive e atua profissionalmente na chamada 'era da informação', na qual, as tecnologias da informação e da comunicação têm presença marcante.** Os novos suportes documentais, com os quais terá de lidar, exigem conhecimento, competência, métodos e meios de produção, utilização e conservação física especiais.

São fatores novos os quais os arquivistas passam agora a serem instruídos e treinados, não obstante concorrem o risco de, em virtude de um vertiginoso crescimento das possibilidades da eletrônica nas áreas documentais, nunca conseguiram abarcar a plenitude destes conhecimentos tão mutantes e dependentes de equipamentos tão rapidamente tornados obsoletos (BELLOTTO, 2007). (grifo nosso)

Os profissionais arquivísticos têm diversos desafios pelo frente. Dentre estes se encontram os ligados a aquisição, manipulação, organização e visualização de dados digitais não estruturados.

Uma das respostas para o enfrentamento dessa realidade pode ser adquirida pelo estudo sistemático das bases matemáticas e informacionais, como por exemplo, estatística computacional, modelagem estatística, infraestrutura *cloud* e estruturas de dados e algoritmos, que atualmente são fundamentais para as estruturas dos documentos digitais, como também àqueles que não são reconhecidamente designados com documentos arquivísticos digitais, conforme discutiremos na próxima sessão.

Os profissionais de dados ou aqueles que seguirem esse caminho de estudos buscam, obrigatoriamente, em suas otimizações algorítmicas ou tecnológicas, muitas vezes no limite do conhecimento, o processamento em massa de dados e sua significação cognitiva para as pessoas (aqui entendidos como gestores, população em geral ou qualquer interessado na sua manipulação). Porém, é importante salientar que não necessariamente compreendem corretamente os princípios e os constructos arquivísticos.

Não obstante, observando-se a ciência arquivística, esse novo profissional emergente deverá se ater à rigorosa técnica da preservação do conteúdo e à autenticidade dos documentos durante todo o ciclo de vida dos documentos.

Na nossa opinião, é no exato encontro dessas duas visões de mundo é que nascerá uma nova geração de profissionais de arquivos, seguindo os passos de alguns bravos desbravadores. Se acreditamos que esse profissional híbrido, que ainda está sendo forjado na labuta do trabalho diário, será necessário se discutir as suas competências para os desafios que estão sendo demandados pela sociedade (em termos de informação tempestiva ou em questões de segurança jurídica e dos dados pessoais) e pelas inovações constante do universo digital em que vivemos (SMIT, 2017; ANDRADE. 2021).

#### **4. UM EXEMPLO DE APLICAÇÃO**

Nosso objeto de estudo será um posto da mídia social *Twitter* (<https://about.twitter.com/>). Um dos problemas de se tratar os documentos digitais de redes sociais é a sua estruturação como um documento único, pois se caracterizam por estruturas fluídas, flexíveis e dinâmicas, ou ainda, que podem não possuir estrutura nenhuma (ANDRADE, 2021).

Para a continuidade de nossa discussão, vamos a um exemplo, com um *post* no *Twitter* do Arquivo Público do Estado de São Paulo (APESP), para entendermos como se estrutura um documento típico, veja a figura 1.

**Figura 1:** *Post* do APESP no *Twitter* com a evidenciação de seus componentes do registro digital realizado.



**Fonte:** Twitter: <https://twitter.com/arquivoestadosp/status/1338502141749694465?s=20>,

visitado em 03 de janeiro de 2021.

A figura 1 apresenta a diplomática básica de um documento quando postado no *Twitter*: (1) o logo ou imagem da pessoa física ou jurídica detentora da conta; (2) o nome de quem é detentor da conta; (3) o apelido da conta no *Twitter*; (4) a data da postagem; (5) o campo para edição com até 280 caracteres, permitindo a inclusão de *links* como realizado pelo APESP; e (6) o campo para inserção de imagens ou de vídeo.

Esta estrutura é fixa para qualquer *post* no *Twitter*. Uma vez postado será permanente o seu conteúdo, pois os campos 5 e 6 não podem mais ser editados, porém poder ser cancelados (deletados). Entretanto, a fluidez do documento digital reside nas interações e na utilização das ferramentas de engajamento que a rede social permite, que é o principal objetivo das mídias sociais. Observe a figura 2.

A figura 2 apresenta os dados fluidos do *post* no *Twitter*, que aumentam a cada interação dos seus usuários. São evidenciados os serviços que a plataforma proporciona gratuitamente, ou seja, engajamento, exposição sentimentos (positivos, neutros ou negativos), interação e criação de reputação. Os dados dinâmicos são: (7) mensagens dos membros da rede social. Este recurso é excelente para a análise de sentimentos com relação ao *post*; (8) retuite do *post*; (9) aprovação social do *post*; (10) oferta serviços adicionais aos membros da rede social.

**Figura 2:** *Post* do APESP no *Twitter* com a evidenciação da extensão infinita do registro documental nas mensagens e funcionalidades oferecidas pela rede social.



**Fonte:** Twitter: <https://twitter.com/arquivoestadosp/status/1338502141749694465?s=20>,

visitado em 03 de janeiro de 2021 e editado por Andrade (2021).

Fazendo a análise do *post*-exemplo, após 27 dias, obtivemos as seguintes informações: nenhuma mensagens dos usuários e um retuite, do Arquivo Nacional, em 15/10/2020.

Imagine, agora, que o analista responsável pela comunicação fosse realizar o seu relatório de mídia. Provavelmente, com os dados acima já contados, tiraria um *print* da tela do computador do *post* original, colaria no *Word* e concluiria seu relatório. Ato contínuo, abriria o seu SIGAD, faria o *upload* do relatório e a sua obrigatória autenticação (pois é um documento exógeno) e o tramitaria para quem de direito. **O relatório se tornou um documento arquivístico, mas os originais, não!**

Extrapolando para o mês de dezembro de 2020, a produção documental no *Twitter* foi de 74 tuites: 13 mensagens (8 positivas e 5 neutras), 84 retuites e 436 likes. Com esses números, o trabalho do analista já seria custoso. Poderíamos recorrer à API (interface de programação de aplicação, em português) do próprio *Twitter* para fazer esse relatório, o que seria realizado em segundos. Entretanto, interações, análises de sentimentos e o engajamento continuariam sendo realizados manualmente, se necessário à sua análise.

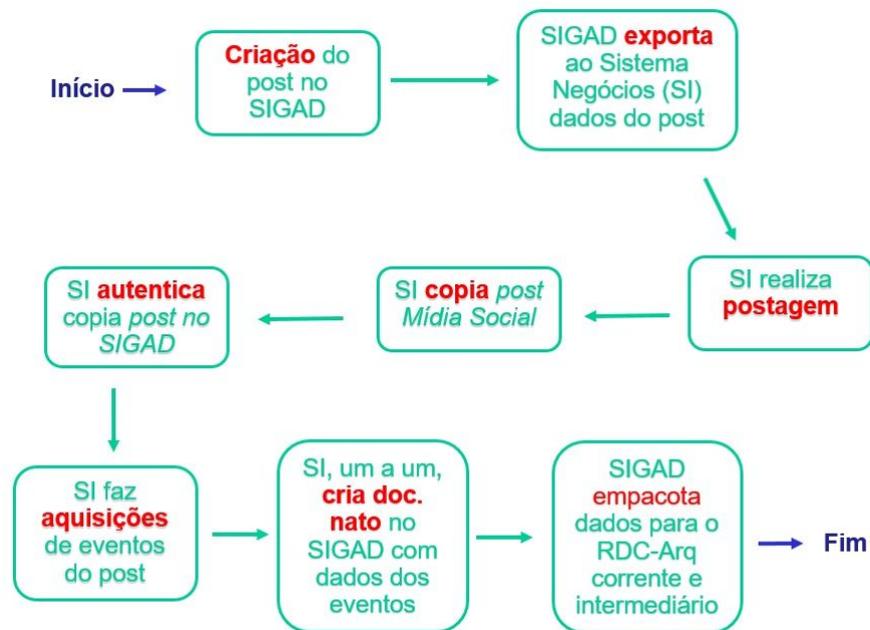
Agora, vamos aumentar a complexidade e volume de análise. E se tivéssemos que verificar todas as outras postagens do APESP e suas manifestações, inclusive vídeos e áudios, em outras redes sociais? E se fossem verificados também os *posts* da Secretaria a qual o APESP está subordinado? Ou da Administração Direta? Certamente, não haveria um instrumento de controle e de informação, como o relatório citado. Teríamos apenas o planejamento de mídia tradicional e os *posts* seriam desconsiderados como documentos de arquivo, como até a presente data são desconsiderados, apesar de conterem manifestações administrativas e comunicações oficiais.

Há um limite físico para o trabalho que o ser humano pode realizar. No mundo dos documentos digitais das redes sociais, com o volume de informação em *terabytes*, torna-se impossível a manipulação pelas pessoas. Muito menos a sua compreensão cognitiva e a aquisição de informação relevante em tempo real. É o tipo de situação que só pode realizado pelo processamento computacional intenso.

Seguindo com o exemplo do *Twitter*, vamos percorrer o trabalho que um cientista de dados arquivístico poderia realizar para a proposição de um modelo de transformação de um documento digital do *Twitter* em um documento arquivístico

e processar essa magnitude de dados, em computação constante. A primeira atividade que este profissional deveria realizar é o fluxo do documento digital de um *post*. Observe-o à figura 3.

**Figura 3:** Fluxo teórico de um *post* para criação de plataforma de documentos digitais arquivísticos de mídias sociais.



Fonte: Andrade (2021).

O fluxo de documento digital é realizado por meio de oito processos distintos, conforme apresentado na figura 3:

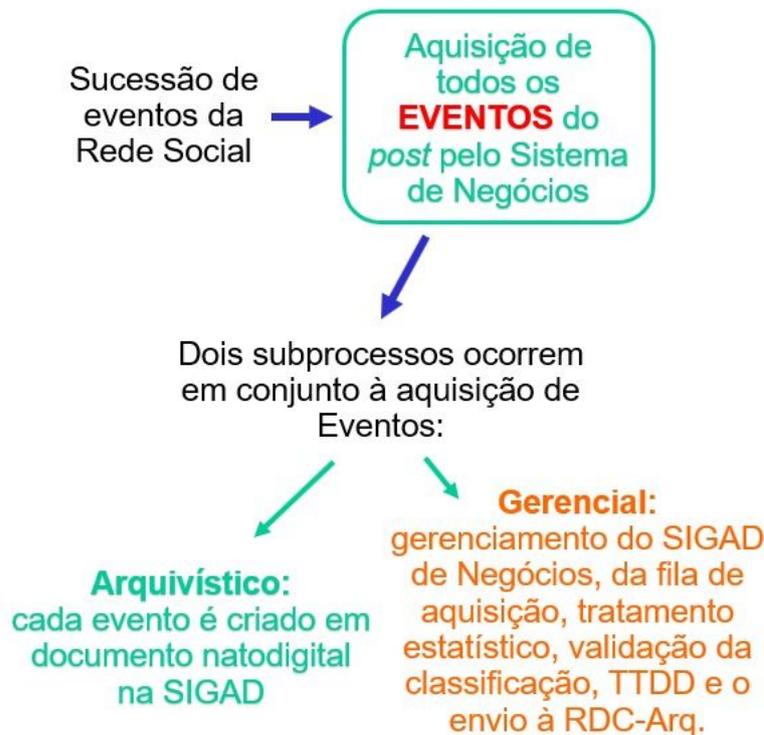
- i. **Criação de SIGAD dedicado:** para a gestão das redes sociais deve ser criado SIGAD específico, que a literatura denomina de SIGAD de Negócio, visando ao controle arquivístico de *posts* nas redes sociais. Para que o conteúdo do *post* (texto e mídia) nasça natodigital, o criador os criará dentro do SIGAD. Quanto à classificação documental, haverá a sua inclusão de acordo com a TTDD da instituição. Por exemplo, segundo *Plano de Classificação* do APESP (APESP, 2019, pag. 193), o

*post* da figura 1 poderia ser classificado com o código 002.05.02.001, de guarda permanente, o que definiria o futuro destino do *post* na origem.

- ii. **O SIGAD de Negócio exporta ao Sistema de Negócio (SN) dados para a postagem:** os itens 5 e 6 a figura 1 são enviados ao SN, respectivamente, dados do conteúdo em texto com imagens ou vídeo, devidamente classificados também.
- iii. **O Sistema de Negócio realiza a postagem na rede social,** automaticamente.
- iv. **O Sistema de Negócio autentica cópia do *post* no SIGAD de Negócio:** de posse da imagem do *post* na rede social o SN a autentica no SIGAD.  
**Um fato importante:** observe a figura 1 e depois, a 2. O conteúdo na figura 2 foi alterado. Para o caso de ocorrer mal uso das informações de postagens, a diplomática forense teria os meios para contradizer arguições de falsidade, com base nos registros arquivísticos realizados, mesmo que o *post* original tenha sido apagado da rede social.
- v. **Sistema de Negócios autentica cópia do *post* no SIGAD:** o SN, de posse da imagem do *post*, faz a autenticação no SIGAD.
- vi. **Sistema de Negócios faz aquisições dos dados dos eventos do *post*:** esse processo não é realizado apenas uma vez, é um *looping* que pode ser repetido milhares de vezes, a cada nova interação com o *post* criado na rede social. Neste processo, a aplicação receberá cada evento e montará a sequência lógico-temporal das interação com o objetivo de recriar a história original do *post* e suas interações.
- vii. **Sistema de Negócios cria documentos natodigitais no SIGAD de Negócios :** o SN, de acordo com a fila de espera, cria documentos, evento a evento, de cada *post* individualmente, no SIGAD.
- viii. **SIGAD de Negócios empacota os documentos para remessa ao RDC-Arq de corrente e intermediário:** em tempos regulares o SIGAD empacota os documentos e os remete ao arquivo RDC-Arq de corrente e intermediário, garantindo a sua preservação arquivística e da cadeia de custódia.

Para encerramos a questão do fluxo, gostaria de pormenorizar o item “VI”, sobre aquisições dos evento de interações advindos da rede social. Por gentileza, observe a figura 4. Podemos observar dois subprocessos de tratamento dos dados do documento digital, que é uma estrutura viva, em construção, e que está em constante transformação pelas próprias características intrínsecas do ambiente das redes sociais.

**Figura 4:** Detalhamento do Processo de Aquisição de Eventos.



Fonte: ANDRADE, 2021.

Os eventos são enviados automaticamente por uma API das redes sociais, e consumidos por uma aplicação intermediária de extração e gerenciamento de filas na nuvem do Sistema de Negócios. As transações são realizadas em via única (*one-way*) e não há trocas de informação com a rede social.

O Sistema de Negócios gerenciará os dois subprocessos da aquisição de eventos do *post*:

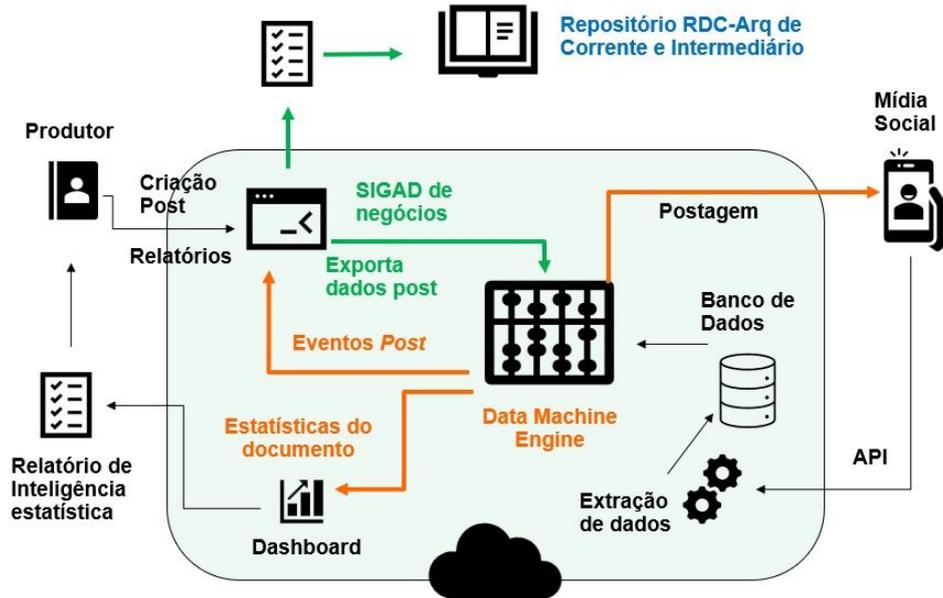
- i. O primeiro, de montagem da estrutura da história temporal do *post* no SIGAD de Negócios, com a criação de documentos natodigitais de cada evento.
- ii. O segundo processo é a inteligência necessária para o que primeiro processo ocorra sem ferir a cadeia de custódia, ou seja, mantendo a autenticidade lógico-computacional. Há também serviços computacionais adicionais visando ao gerenciamento da fila de aquisição, ao tratamento estatístico, à validação da classificação arquivística, ao gerenciamento da TTDD e ao envio de pacotes ao RDC-Arq de corrente e intermediário.

Nesses dois processos reside o trabalho diferenciado do cientista de dados arquivístico, unindo a Arquivística com a Ciência de Dados e o seu tripé referente à aplicação estatístico-matemático, à estruturação da *cloud computing* e à programação e estruturação de dados não estruturados, pois seria impossível o gerenciamento de massa de dados proveniente mídias sociais por meio do esforço humano. Na figura 5, esquematizei a arquitetura de serviços do modelo proposto, a título de visualização geral.

Dos todos elementos apresentados na figura 5, quatro merecem maiores detalhamentos:

- i. **API (de *application programming interface*):** chamada em português de *Interface de Programação de Aplicação*, designa um conjunto de rotinas e padrões estabelecidos por uma aplicação para a utilização de suas funcionalidades. A API do *Twitter* está disponível em <https://developer.twitter.com/en/docs/twitter-api> e apresenta as funcionalidades disponíveis que alimentarão os eventos para criação do documento integrado: tuites, localização, mensagens diretas, likes *etc.*

**Figura 5** - Esboço da arquitetura de serviços realizados pelo modelo para o documento digital proveniente de redes sociais.



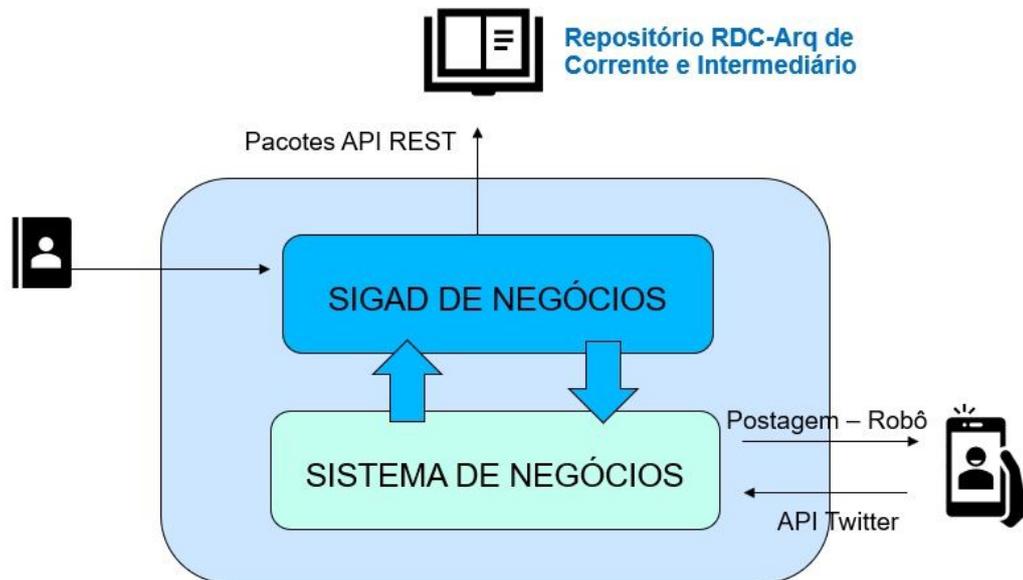
Fonte: Andrade (2021).

- ii. **SIGAD de Negócios:** é desenvolvido segundo as orientações do e-ARQ Brasil (CONARQ, 2011). Sua finalidade é suportar um processo de negócio específico, não possuindo nenhuma dependência de banco de dados, obrigatoriamente (FLORES, 2018).
- iii. **Repositório RDC-Arq de corrente e intermediário:** o repositório segue as diretrizes para implementação de repertórios arquivísticos, porém não é de guarda permanente (CONARQ, 2015), o que nos permite atualizar os documentos de cada *post* em tempos regulares, não ferindo o disposto no Art. 25 da lei 8159, que proíbe a apagar documentos nos arquivos permanentes sem o devido processo legal. Por outro lado, o tratamento dessas atualizações regulares deverá ter regra específica de governança (para apagar registros, por exemplo) a ser implementada pelo Sistema de Negócios.

- iv. **Data Machine Engine:** é o motor de processamento da aplicação, ou seja, onde se processa a inteligência do modelo proposto. Na modelagem que estou desenvolvendo será uma aplicação de *Machine Learning* (Aprendizado por Máquinas), que coordenará o processo iniciando-se a cada nova postagem, em qualquer rede social, quando responsável (produtor) criar o conteúdo do *post* no SIGAD. Como comentado, controlará o SIGAD de Negócios e o Repositório de corrente e intermediário.

Por fim, acreditamos a cadeia de custódia não é quebrada, entretanto foi estendida. O modelo da arquitetura foi uma das preconizadas pela *Resolução 43 do CONARQ* (CONARQ, 2015), conforme a figura 6.

**Figura 6:** protótipo do modelo sugerido.



**Fonte:** realizado por Andrade (2021) e baseado em CONARQ (2015).

## 5 CONCLUSÃO

A *ciência arquivística* está sendo estimulada a repensar os seus constructos epistemológicos. De um lado, para responder às demandas da sociedade, principalmente

com relação à disponibilização de informações ao público, ou seja, ampliar a sua atuação em termos sociais na difusão destes, deixando de enfatizar sua atuação para dentro das instituições. Há um universo de dados e de informações que estão escondidos nas organizações e muitas vezes a sociedade não compreende racionalmente os investimentos realizados nos arquivos.

Em outra frente, na expansão dos princípios arquivísticos visando a atender as inovações tecnológicas que estão disrompendo negócios, teorias e arquétipos jurídicos. Nesse sentido, um exemplo real foi tratado nesse trabalho, utilizando-se as mídias sociais e seus documentos digitais não estruturados, que estão entre nós há mais de 10 anos, inclusive garantido direitos e obrigações às pessoas.

É um documento relevante que deve ser repensado arquivisticamente. Porém, o exemplo é apenas uma “ponta do iceberg” de uma onda de novos documentos digitais não estruturados de inúmeras inovações e suas adaptações que estão aparecendo todos os dias. Citamos o *blockchain* e os microvídeos do *Tic Tok*, mas a inventabilidade brasileira também chegou ao novíssimo PIX, que foi criado para ser um instrumento financeiro de transferência de valores monetários, mas virou um novo meio de mensagens (inclusive para paqueras e para cobranças), ao custo de R\$ 0,01 a cada 100 caracteres a quem os envia (aos bancos emissores há outro custo adicional).

Obviamente, toda essa revolução tem os seus custos. No caso do emergente cientista de dados arquivístico é concernente à sua formação (teórica e prática) e à sua inclusão em posições de liderança em projetos nos Arquivos. Quando me refiro a essa inclusão, quero enfatizar a sua atuação como membro ou líder atuante das decisões estratégicas e não apenas como mão de obra contratada para projetos itinerantes, atuando como um recurso de TI e não absorvendo a teoria arquivística, como costumeiramente acontece.

Poderá existir um ponto de ruptura em que as ciências da informação se sobreponham à arquivística, devido ao seu ferramental matemático-analítico,

principalmente, se esta ficar enclausurada em seus princípios históricos-tecnicistas, como comentam Rodrigues (2008), Bellotto (2007), Smit (2017), Flores (2019) e Andrade (2021) supracitados e tantos outros autores e acadêmicos.

Por fim, como toda mudança de paradigma, haverá a adequação de expectativas dos profissionais de arquivos com profissionais de dados provenientes das escolas de ciências exatas. Quanto maior a empatia de ambos os lados, melhor será a solução dos desafios em curso. Entretanto, o melhor dos mundos seria a formação do *cientista de dados arquivístico* iniciar-se nos estudos da graduação. Acredito que este aspecto será um tema importante de discussão na *Academia*. Agradeço a atenção e aguardo as suas considerações. Obrigado!

## REFERÊNCIAS

ANDRADE, Rogério Ferraz Garcia. **Arquivística, Redes Sociais e Cadeia de Custódia: o campo o campo de trabalho do Cientista de Dados Arquivístico**. São Paulo. Revista do Arquivo Público do Estado de São Paulo, Ed. 12. 2021.

ARQUIVO PÚBLICO DO ESTADO DE SÃO PAULO. **Plano de Classificação e Tabela de Temporalidade de Documentos da Administração Pública do Estado de São Paulo: atividade-meio**. 2ª edição. 2018. Disponível em [http://www.arquivoestado.sp.gov.br/site/assets/publicacao/anexo/2019\\_PUBLICACAO\\_PC\\_E\\_TTDM\\_EIO\\_PARA%20DIVULGACAO.pdf](http://www.arquivoestado.sp.gov.br/site/assets/publicacao/anexo/2019_PUBLICACAO_PC_E_TTDM_EIO_PARA%20DIVULGACAO.pdf), visitado em 10 de janeiro de 2021.

BRASIL. **LEI nº 6547, Regulamentação das Profissões de Arquivista e Técnico de Arquivo**. 1978. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/LEIS/1970-1979/L6546.htm](http://www.planalto.gov.br/ccivil_03/LEIS/1970-1979/L6546.htm). Visitada em: 10 de janeiro de 2021.

BRASIL. **LEI 8.159, Política Nacional de Arquivos Públicos**. 1991. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/LEIS/L8159.htm](https://www.planalto.gov.br/ccivil_03/LEIS/L8159.htm). Visitado em: 26 de dezembro de 2020.

CONSELHO NACIONAL DE ARQUIVOS. **e-ARQ Brasil: Modelos de Requisitos para Sistemas Informatizados de Gestão de Documento**. 2011. Disponível em: <http://siga.arquivonacional.gov.br/images/publicacoes/e-arq.pdf>, visitado em 27 de dezembro de 2020

(\_\_\_\_.); **Oficina e-ARQ Brasil: Modelos de Requisitos para Sistemas Informatizados de Gestão de Documento**. Técnica de Documentos Eletrônicos. Rio de Janeiro, 2013. Disponível em: [https://www.gov.br/conarq/pt-br/assuntos/camaras-tecnicas-setoriais-inativas/camara-tecnica-de-documentos-eletronicos-ctde/earq\\_oficina\\_2013.pdf](https://www.gov.br/conarq/pt-br/assuntos/camaras-tecnicas-setoriais-inativas/camara-tecnica-de-documentos-eletronicos-ctde/earq_oficina_2013.pdf), visitado em: 28 de dezembro de 2020.

(\_\_\_\_.); **RESOLUÇÃO N° 39, DE 29 DE ABRIL DE 2014**. 2014. Disponível em: <https://dhg1h5j42swfq.cloudfront.net/2016/02/20153108/Diretrizes-para-Implementa%C3%A7%C3%A3o-de-Reposit%C3%B3rios-Digitais-Confi%C3%A1veis.pdf>, visitado em 6 de janeiro de 2021.

(\_\_\_\_.); **Diretrizes para a implantação de repositórios arquivísticos digitais confiáveis - Resolução 43**. 2015. Disponível em: [https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/conarq\\_diretrizes\\_rdc\\_arq\\_resolucao\\_43.pdf](https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/conarq_diretrizes_rdc_arq_resolucao_43.pdf), visitado em 4 de janeiro de 2021.

FLORES, Daniel; ROCCO, B. C. B; SANTOS, H. M. **Cadeia de custódia para documentos arquivísticos digitais**. Acervo, Rio de Janeiro, v. 29, n. 2, p. 117-132, jul./dez. 2016 – p. 117. 2016. Disponível em: <http://revista.arquivonacional.gov.br/index.php/revistaacervo/article/view/717/732>, visitado em 4 de janeiro de 2021. <https://www.youtube.com/watch?v=E0AgfsvmkGs>. Visitado em 07 de janeiro de 2021.

(\_\_\_\_.). **Transformando um GED ou sistema de negócio em um SIGAD e interoperando-o via RESP API ao RDC-Arq: cadeia de custódia digital arquivística de documentos digitais autênticos com preservação digital sistêmica**. Palestra ao SINFO-UFRN, em 19 de julho de 2018. 2018. Disponível em:

FONSECA, M. O. K. **Arquivologia e Ciência da Informação**. Rio de Janeiro: Editora FGV, 2005.

RODRIGUES, Ana Célia. **Diplomática contemporânea como fundamento metodológico da identificação de tipologia documental em arquivos**. Tese (Doutorado) - Universidade de São Paulo, São Paulo, 2008.

SMIT, Johanna W. **Arquivologia, Ciência da Informação e as duas faces de Jano**. Brazilian Journal of Information Science. Reserch Trends. 11:4. 2017.