

REGISTROS DIGITAIS ARQUIVÍSTICOS DO SIPAC PROTOCOLO: um olhar para os metadados de preservação digital

SIPAC protocol archival digital records: a look at digital preservation metadata



 **Pedro Felipy Cunha da Silva**

Universidade Federal da Paraíba
E-mail: pedrofelipy@hotmail.com
João Pessoa, PB – Brasil

 **Wagner Junqueira de Araújo**

Universidade Federal da Paraíba
E-mail: wagnerjunqueira.araujo@gmail.com
João Pessoa, PB – Brasil



Resumo

Introdução: Um avançado processo de transformação digital já estava em curso, quando recebeu o impacto da ampliação do trabalho remoto durante e após a pandemia de COVID-19. **Objetivo:** Este estudo buscou compreender como estão sendo gerenciados os metadados de preservação relacionados aos registros digitais arquivísticos das instituições federais de ensino superior brasileiras. **Metodologia:** Parte de um estudo de caso do Centro de Ciências Jurídicas da Universidade Federal da Paraíba e da aplicação de pesquisa junto às Instituições Federais de Ensino usuárias do software SIPAC Protocolo, por meio de um questionário enviado para dezesseis instituições. As respostas foram tabuladas utilizando o software Microsoft Excel, e posteriormente analisadas com o auxílio do software Rapid Miner. **Resultados:** Obteve-se uma descrição dos metadados armazenados, com a diferenciação daqueles gerados de forma automática dos inseridos manualmente pelo usuário. **Conclusão:** Observou-se que esses metadados, quando associam as transações e requisitos, validam o registro produzido, permitindo que os dados nele constantes possam se tornar um documento manifestado e acessível ao usuário no decorrer do tempo. Esse processo nem sempre cria um registro digital arquivístico; apenas quando evidencia e serve de prova das atividades de uma pessoa ou instituição. Foi possível notar ainda que o caráter dinâmico dos recursos digitais requer a reunião de informações com nível de segurança razoável, garantindo que o ambiente computacional no qual a informação foi criada poderá ser reproduzido com o passar do tempo. Desta forma, os registros permanecerão acessíveis, independentemente da existência do formato e das condições técnicas em que foram concebidos.

Palavras-chave: Registro digital arquivístico. Metadado. Preservação digital. Arquivologia. SIPAC Protocolo.

Abstract

Introduction: An advanced digital transformation process was already underway, when it was impacted by the expansion of remote work during and after the COVID-19 pandemic. **Objective:** This paper aimed to understand how the preservation metadata related to archival digital records of Brazilian universities are being managed. **Methodology:** To do so, it was based both on the case study of the Federal University of Paraíba Law School and the application of research with the Brazilian Federal Universities that use the SIPAC Protocol software. A questionnaire was sent to sixteen institutions. Their responses were tabulated using Microsoft Excel software and analyzed using the Rapid Miner software. **Results:** A description of the stored metadata was obtained, differentiating those generated automatically from those entered manually by the user. **Conclusion:** We noted that these metadata, when associating their specific transactions and requirements, validate the register in production, allowing the data contained to become a manifested and accessible document over time. However, not always the described process creates an archival digital record, only to make evident and to prove the activities of a person or an institution. We also noted that the dynamic of digital resources requires the gathering of information with a reasonable level of security, allowing the computational environment in which the information was created to be reproduced over time, so the records remain accessible regardless of format and technical conditions in which they were initially created.

Keywords: Archival digital record. Metadata. Digital preservation. Archival science. SIPAC Protocol.

LICENÇA DE USO

Os autores cedem à [Revista Brasileira de Preservação Digital](#) os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a Licença Creative Commons Attribution (CC BY) 4.0 International. Esta licença permite que terceiros remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os autores têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHERS

Universidade Estadual de Campinas – Sistema de Bibliotecas / Instituto Brasileiro de Informação em Ciência e Tecnologia – Rede Brasileira de Serviços de Preservação Digital – Cariniana. As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Gildenir Carolino Santos, Miguel Angel Márdero Arellano.

CREDIT

- **Reconhecimentos:** Não aplicável.
- **Financiamento:** Não aplicável.
- **Conflitos de interesse:** Os autores certificam que não têm interesse comercial ou associativo que represente um conflito de interesses em relação ao manuscrito.
- **Aprovação ética:** Não aplicável.
- **Disponibilidade de dados e material:** Não aplicável.
- **Contribuições dos autores:** Conceitualização, Curadoria de dados, Análise formal, Investigação, Metodologia, Administração do projeto, Visualização e Escrita – rascunho original*: SILVA, P. F. C. da. Escrita – revisão & edição*: ARAÚJO, W. J.

ODS: 4. Educação de Qualidade

Submetido em: 04/05/2023 – Aceito em: 01/11/2023 – Publicado em: 04/03/2024

1 Introdução

Um avançado processo de transformação digital já estava em curso, quando recebeu o impacto da ampliação do trabalho remoto durante e após a pandemia de COVID-19. Diante desse contexto, ações passaram a ser registradas e documentadas de forma eletrônica com maior intensidade e em formatos cada vez mais diversificados.

Tal cenário trouxe para a Ciência da Informação e, sobretudo para a Arquivologia, a necessidade de reflexões sobre a preservação dos documentos e registros criados digitalmente. Para isso, buscou-se identificar, a partir dos metadados, elementos que contribuem para a preservação dos registros digitais arquivísticos. Para fins deste estudo, foram analisados em um primeiro momento os documentos e registros produzidos em formato nato-digital pelo Centro de Ciências Jurídicas da Universidade Federal da Paraíba (CCJ/UFPB), e, numa segunda análise, por um conjunto de instituições federais de ensino superior brasileiras que utilizam o módulo Protocolo do Sistema Integrado de Patrimônio, Administração e Contratos (SIPAC).

O SIPAC é um dos *softwares* que compõem o conjunto de sistemas de gestão SIG. É formado por 29 módulos e 4 portais. O objeto do trabalho é o SIPAC Módulo Protocolo, sistema de produção, tramitação e arquivamento de registros digitais, documentos avulsos e processos administrativos, utilizado por todas as unidades administrativas da UFPB, em atendimento a demanda de transformação digital, imposta pelo Decreto Federal 8.539/2015.

2 Revisão da literatura

O termo **registro digital arquivístico** possui como alicerce teórico a combinação da definição de registro digital (tradução nossa), proposto pela pesquisadora canadense Corinne Rogers (2015); com os conceitos de arquivo, documento, documento digital e gestão de documentos, dispostos no Dicionário Brasileiro de Terminologia Arquivística, publicado pelo Arquivo Nacional (2005).

No dicionário citado, **documento** é definido como “unidade de registro de informações, qualquer que seja o suporte ou formato” (ARQUIVO NACIONAL, 2005, p. 73). Enquanto **documento digital**, é aquele “codificado em dígitos binários, acessível por meio de sistema computacional” (ARQUIVO NACIONAL, 2005, p. 75). A **gestão de documentos** é conceituada como o “conjunto de procedimentos e operações técnicas referentes à produção, tramitação, uso, avaliação e arquivamento de documentos em fase corrente e intermediária, visando sua eliminação ou recolhimento” (ARQUIVO NACIONAL, 2005, p. 100). E, para além das definições quanto à instituição, instalações e móvel de guarda, o termo **arquivo** é definido como o “conjunto de documentos produzidos e acumulados por uma entidade coletiva, pública ou privada, pessoa ou família, no desempenho de suas atividades, independentemente da natureza ou suporte” (ARQUIVO NACIONAL, 2005, p. 27).

Registros digitais:

[...] consistem em dados gerados pelo usuário (conteúdo), metadados gerados pelo sistema de identificação de origem e localização, metadados gerados pelo aplicativo que gerencia aparência e desempenho do registro (exemplo: formato de arquivo nativo), metadados gerados pelo aplicativo de descrição dos dados (exemplo: metadados do sistema operacional do sistema de arquivos) e metadados gerados pelo usuário que descrevem os dados (ROGERS, 2015, p. 16, p. 18, tradução nossa).

Consoante tais conceitos, a adoção do termo registro digital, pode ser entendida como uma possibilidade terminológica que inclui, porém amplia, o conceito de documento digital, forma pela qual o termo *digital record* também pode ser traduzido.

A condição arquivística dos registros digitais se dará quando sua produção e acumulação forem realizadas como prova do desempenho das atividades das pessoas, físicas ou jurídicas, por eles responsáveis. E sua gestão, de maneira análoga à gestão de documentos, aos procedimentos relativos ao seu ciclo de vida.

Ao se combinar as definições do Arquivo Nacional (2005) de “documento” (ARQUIVO NACIONAL, 2005, p. 73), de “documento digital” (ARQUIVO NACIONAL, 2005, p. 75) e a de “gestão de documentos”, com o conceito de Rogers (2015), o termo registro digital arquivístico é entendido, para fins deste trabalho, como uma unidade de registro de informações e como um tipo de documento digital. A aproximação com a definição de documento digital se dá por sua composição em dígitos binários e a necessidade de sistema computacional para sua leitura. Contudo, esse tipo de documento não se limita à forma clássica, uma vez que também considera outros elementos como parte da sua constituição, conforme define Rogers (2015) ao se referir aos dados e metadados.

Embora o Dicionário Brasileiro de Terminologia Arquivística não traga um conceito para registros digitais arquivísticos, dado o seu caráter contemporâneo, o Conselho Nacional de Arquivos, por meio da Câmara Técnica de Documentos Eletrônicos (CTDE) apresenta, dentre outras coisas, a necessidade de gestão dos metadados, conforme prevê a Resolução 43/2015, quando descreve as diretrizes para a implementação de repositórios digitais arquivísticos confiáveis.

Percebe-se que a utilização de ambientes digitais e a consequente produção de registros digitais arquivísticos torna mais complexa a tarefa de gerenciar tais elementos (Rogers, 2015), dada a multiplicidade de possibilidades de interação e consequente significação pelo usuário (Henntonem, 2017).

A diversificação da natureza dos registros digitais é uma realidade e o aumento do seu volume tem sido exponencial nas últimas décadas (Katu, 2016), de tal maneira que é necessário um esforço intelectual para a compreensão das novas demandas de gestão arquivística, necessárias para atuar com essa nova possibilidade de acervo.

Desta forma, este trabalho buscou compreender como estão sendo geridos os metadados de preservação digital relacionados aos registros digitais arquivísticos, no âmbito das instituições federais de ensino superior brasileiras, a partir do estudo de caso do CCJ/UFPB e da aplicação de pesquisa junto as IFES usuárias do SIPAC Módulo Protocolo.

3 Materiais e métodos

Trata-se de um estudo de caso que, de acordo com Gil (2010, p. 37), “é uma modalidade de pesquisa amplamente utilizada nas ciências sociais. Consiste no estudo profundo e exaustivo de um ou poucos objetos, de maneira que permita seu amplo e detalhado conhecimento”.

Para a identificação dos metadados de preservação digital, optou-se como amostra o conjunto de universidades federais que utilizam o mesmo sistema de produção de registros digitais, o SIPAC Módulo Protocolo.

Com o objetivo de identificar os metadados necessários para a preservação digital dos registros digitais arquivísticos, com base em Rogers (2015) e Sayão (2010),

este trabalho buscou conhecer quais metadados dessa finalidade estão sendo utilizados inicialmente pela UFPB e, de forma complementar, por outras instituições congêneres.

Para isso, foram enviados questionários para todas as universidades federais brasileiras que utilizam o mesmo sistema de produção de registros digitais que a UFPB, a saber, os Sistemas Integrados de Gestão (SIG), plataforma desenvolvida e distribuída pela Universidade Federal do Rio Grande do Norte (UFRN).

Importa esclarecer que tal sistema é constituído de diversos subsistemas, a exemplo do SIGAA, SIPAC, SIGRH, SIGPP, SIGED, SIGELEIÇÃO, entre outros. Contudo, considerando a amplitude de registros gerados pelos diferentes módulos e a necessidade de metadados específicos para cada um, este trabalho se limitou aos metadados gerados para/por um único módulo: o SIPAC Protocolo.

Esta escolha se deu com base nos dados obtidos no diagnóstico documental inicial realizado no CCJ/UFPB (SILVA, 2018). Optou-se pela investigação acerca do Módulo Protocolo do SIPAC – Sistema Integrado de Patrimônio, Administração e Contratos –, que é responsável, dentre outros aspectos, pela produção e tramitação de documentos e processos administrativos, e figura como o *software* com maior número de registros digitais produzidos no âmbito deste sistema.

Destaca-se o fato de o SIPAC já ser o sistema de produção de documentos mais utilizado pelas Instituições Federais de Ensino Superior (IFES), além de estar presente ainda em outros órgãos do poder executivo federal, em órgãos estaduais e em instituições das cinco regiões do país.

Com o propósito de conhecer como as IFES realizam a preservação digital por meio dos metadados, foi elaborado um questionário contendo seis perguntas, e esse instrumento de coleta de dados foi enviado para dezesseis instituições por meio da Plataforma Integrada de Ouvidoria e Acesso à Informação (FalaBR), site criado pelo Governo Federal no ano de 2020, em substituição ao Sistema Eletrônico de Informação ao Cidadão (eSIC), como ferramenta de aplicação da Lei de Acesso à Informação.

O envio dos questionários foi realizado em 08/09/2020 para as seguintes universidades:

Quadro 1 – Universidades Federais que utilizam o SIPAC

| |
|--|
| Fundação Universidade Federal da Grande Dourados – UFGD |
| Fundação Universidade Federal de São João Del Rei – UFSJ |
| Fundação Universidade Federal de Sergipe – UFS |
| Fundação Universidade Federal do ABC – UFABC |
| Fundação Universidade Federal do Maranhão – UFMA |
| Fundação Universidade Federal do Vale do São Francisco – UNIVASF |
| Universidade Federal da Integração Latino-Americana – UNILA |
| Universidade Federal da Paraíba – UFPB |
| Universidade Federal de Itajubá – UNIFEI |
| Universidade Federal de Lavras – UFLA |
| Universidade Federal do Oeste da Bahia – UFOB |
| Universidade Federal do Recôncavo da Bahia – UFRB |
| Universidade Federal do Rio Grande do Norte – UFRN |
| Universidade Federal Rural da Amazônia – UFRA |

| |
|--|
| Universidade Federal Rural do Rio de Janeiro – UFRRJ |
|--|

| |
|--|
| Universidade Federal Rural do Semiárido – UFERSA |
|--|

Fonte: dados da pesquisa, 2020.

Dentre as instituições pesquisadas, a UFRA ignorou a manifestação realizada para o seu Sistema de Acesso à Informação e não respondeu o pedido de informação enviado pela plataforma FalaBR. Já a UNIFEI respondeu que não utiliza os sistemas SIG. As outras treze instituições enviaram respostas consideradas satisfatórias para os objetivos desta pesquisa, cujos resultados serão descritos no item a seguir.

Convém elucidar que, após a realização dessa fase do trabalho, obteve-se a informação no portal de cooperação da UFRN que são 29 universidades federais que utilizam a plataforma SIG e não 16, como apresentado no ENARQUIFES (2019). Entretanto, considerando as dificuldades encontradas em obter as respostas das instituições dentro dos prazos legais, em virtude da adoção do regime de trabalho remoto por ocasião da pandemia de coronavírus, bem como as limitações de cronograma constantes no planejamento deste trabalho, optou-se por realizar a análise com as 16 instituições já pesquisadas, com a compreensão de que essa amostra já compreende mais do que 50% das universidades federais que utilizam o SIPAC.

As respostas obtidas das IFES quanto aos metadados de preservação digital foram tabuladas utilizando o *software Microsoft Excel* para a construção de uma base de dados, que foi analisada com o auxílio do *software* de análise e visualização de dados *Rapid Miner*.

4 Discussão e Resultados

A combinação do conceito de Rogers (2015) sobre os registros digitais arquivísticos (que, além de relacioná-los aos dados de conteúdo gerados pelo usuário, os explicam como constituídos de quatro conjuntos diferentes de metadados) à reflexão de Sayão (2010) acerca do OAIS, considerado o mais importante modelo de requisitos de preservação de objetos digitais da atualidade, nos levou à reflexão constante no quadro 2.

Quadro 2 – Metadados para preservação digital de registros digitais arquivísticos

| | |
|--|---|
| (1) metadados de identificação de origem e localização; | infraestrutura conceitual que descreve o ambiente [...] |
| (2) metadados de aparência e desempenho; | [...] as interfaces externas [...] |
| (3) metadados de descrição automática pelo sistema operacional e/ou de sistema de arquivos e | [...] os componentes funcionais e os objetos de informação, associados com um sistema responsável pela preservação de longo prazo de materiais digitais |
| (4) metadados de descrição manual gerados pelo usuário. | |
| (Rogers, 2015, p. 16,18). | (Sayão, 2010, p. 13). |

Fonte: elaborado pelos autores (2020) a partir de Rogers (2015) e Sayão (2010).

O quadro acima reflete de forma conceitual os metadados necessários para a preservação digital dos registros digitais arquivísticos e, com base nele, foram

elaboradas as questões enviadas às universidades federais que utilizam o SIPAC, conforme questionário abaixo:

Quadro 3 – Questionário sobre metadados de preservação digital para as universidades federais que utilizam o SIPAC

METADADOS DE PRESERVAÇÃO DIGITAL DE FORMA GERAL

Quais metadados são criados e armazenados pela sua instituição para preservação dos registros e documentos produzidos nos sistemas SIPAC e SIGAA (ou nos sistemas de produção de documentos acadêmicos e administrativos)?

METADADOS GERADOS DE FORMA AUTOMÁTICA

A sua instituição armazena metadados que identificam a origem e a localização dos usuários que produzem os registros e documentos? Caso sim, quais metadados são armazenados?

A sua instituição armazena metadados que diferenciem os registros e documentos gerados em ambiente digital (nato digitais) daqueles convertidos para o meio eletrônico (digitalizados)? Caso sim, quais metadados são armazenados em cada caso?

A sua instituição armazena metadados que descrevam o dispositivo e o sistema operacional no qual os registros e documentos foram produzidos? Caso sim, quais metadados são armazenados a respeito do dispositivo e do sistema operacional?

A sua instituição armazena metadados que descrevam os dados constantes nos registros e documentos criados/inseridos nos sistemas SIGAA/SIPAC (ou nos sistemas de produção de documentos acadêmicos e administrativos)? Caso sim, quais são os metadados armazenados para descrição dos dados?

METADADOS INSERIDOS MANUALMENTE PELO USUÁRIO

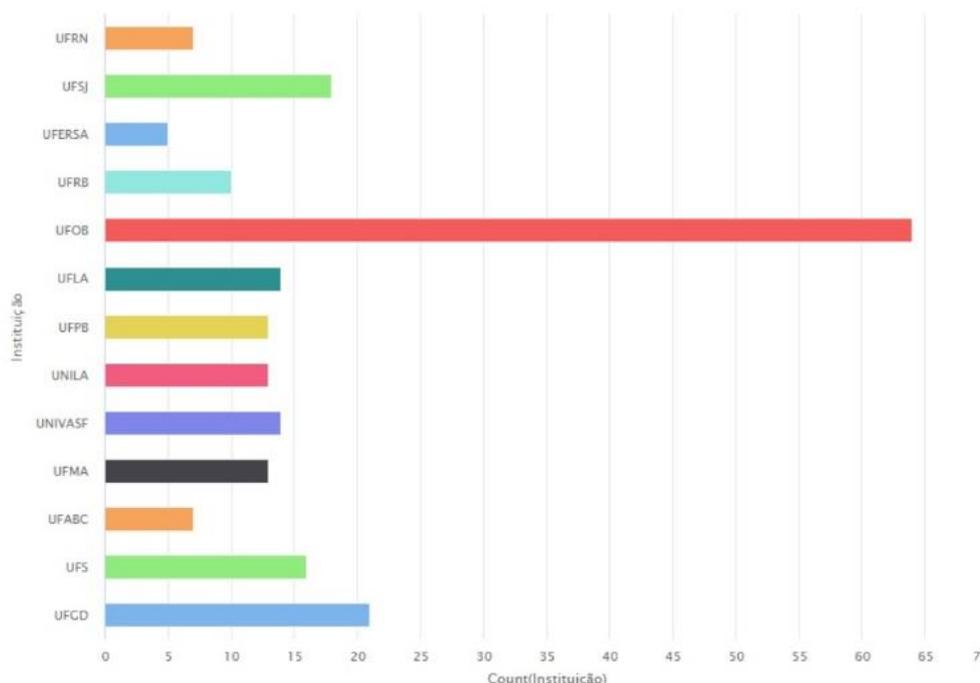
Existem metadados que descrevam os dados constantes nos registros e documentos criados/inseridos nos sistemas SIGAA/SIPAC (ou nos sistemas de produção de documentos acadêmicos e administrativos) que são incluídos pelo usuário de forma manual? Caso sim, quais são os metadados armazenados para descrição dos dados inseridos manualmente pelo usuário?

Fonte: elaborado pelos autores (2020).

Quanto aos resultados das respostas ao questionamento do nome do setor responsável pelas respostas, verifica-se que 23% das respostas foram fornecidas por setores de Arquivo e Protocolo; 77%, por setores de Tecnologia e Informática, com destaque para a nomenclatura “Superintendência de Tecnologia da Informação”, comum em duas instituições: UFMA e UFPB. Tal cenário sugere que a atuação dos arquivistas na gestão dos registros em ambientes digitais já é uma realidade, entretanto, ainda são os profissionais de tecnologia da informação que ocupam a maior parte dos espaços institucionais sobre o tema.

O gráfico 1 apresenta total de metadados diferentes informados como resposta ao quesito: 1 - Quais metadados são criados e armazenados pela sua instituição para preservação dos registros e documentos produzidos nos sistemas SIPAC e SIGAA?

Gráfico 1 – Metadados de preservação digital nas IFES que utilizam o SIPAC



Fonte: dados da pesquisa, 2020

O gráfico acima demonstra a amplitude da quantidade de metadados informados pelas universidades para a preservação dos seus registros digitais. Com destaques para o maior e menor resultado, respectivamente, a Universidade Federal do Oeste da Bahia (UFOB), com 64 (sessenta e quatro) tipos de metadados, e a Universidade Federal do Semiárido (UFERSA), com 5 (cinco) tipos de metadados.

Ressalta-se que a quantidade de campos de metadados indicados como resposta ao primeiro questionamento desta pesquisa não é um indicador da qualidade dos metadados de preservação digital das IFES participantes. A análise qualitativa a seguir oferecerá outros elementos para a compreensão da aderência dessas instituições aos preceitos de preservação digital por meio de metadados.

Dessa forma, para além da análise quantitativa, buscou-se compreender quais são os metadados informados pelas IFES para a preservação dos seus registros digitais. Os tipos de metadados com incidência maior do que um foram destacados. Os destaques podem ser visualizados no quadro abaixo:

Quadro 4 – Metadados de preservação digital de registros digitais arquivísticos das IFES que utilizam o SIPAC/Módulo Protocolo

| Grupo 1 - Três ocorrências | Grupo 2 - Duas ocorrências |
|----------------------------|------------------------------------|
| Assunto | Assunto detalhado |
| Data de criação | Data de cadastro |
| Identificador | Data do documento |
| Número de páginas | Data do <i>upload</i> |
| Unidade de origem | Nome do arquivo |
| | Para documento – Assunto detalhado |
| | Para documento – Observação |
| | Tamanho |
| | Tamanho do arquivo |
| | Tipo de documento |

Fonte: dados da pesquisa, 2020.

Inicialmente, analisaremos os campos de metadados do grupo um, que foram aqueles com três ocorrências, destacando os desafios relacionados à preservação digital para cada um deles. O primeiro termo, “assunto”, é utilizado por algumas instituições como o campo por meio do qual o usuário escolhe o código de classificação de documentos estabelecido pelo Conselho Nacional de Arquivos.

A opção de escolher o código do assunto exige do usuário um conhecimento abrangente dos instrumentos de gestão de documentos das atividades meio e fim da administração pública. Tal código é o fundamento legal para o estabelecimento da temporalidade do documento no sistema, ou seja, a definição do seu ciclo de vida. Do mesmo modo, essa escolha permite a definição da destinação do documento, isto é, a decisão a respeito da sua guarda temporária ou permanente. São vários os problemas associados a esse metadado, caso ele não seja adequadamente preenchido, seja pela falta de treinamento do usuário ou de adequação do sistema.

Primeiro, a escolha do código só irá garantir o cumprimento da temporalidade e destinação dos documentos se tal função estiver programada no sistema, de forma que ter a informação do código é importante, porém, não é suficiente para que a gestão do documento ocorra.

Segundo, permitir ao usuário escolher dentre um rol de centenas de códigos, em alguns casos com descrições com pouca ou nenhuma aproximação com a finalidade do documento, coloca em risco a finalidade de gerir tais documentos, pois aumenta a chance de erros de classificação, podendo um documento de guarda permanente ser definido como de guarda temporária e vice-versa.

Seria necessária uma ferramenta capaz de realizar a classificação de forma automatizada, de maneira que o usuário não precise escolher o código de classificação, mas ao escolher o tipo de documento, com base em formulário que use um dicionário de termos e vocabulário controlado previamente estabelecido na programação do sistema. De forma que cada tipo documental já esteja vinculado ao código de classificação do assunto correspondente, aumentando a eficiência já na fase da produção, tramitação e tratamento dos documentos.

Além das dificuldades já mencionadas, nem toda universidade usa esse campo “assunto” para o código de classificação do CONARQ. Existem ainda os casos em que esse campo é utilizado de forma indiscriminada, ampliando ainda mais os riscos à preservação desses registros.

Em se tratando do item “data de criação”, fundamental para a gestão do ciclo de vida de qualquer objeto digital, é possível perceber que, em algumas IFES, ele não é utilizado de modo a favorecer a preservação digital, haja vista que não é feita diferenciação entre o metadado gerado para os documentos nato digitais daqueles gerados para os documentos digitalizados. Isso quer dizer que, se um documento digitalizado for recebido pela instituição em uma determinada data, a ausência de campo específico para a separação da data de recebimento da data de criação do documento original pode facilitar a inserção de informação incorreta.

A ausência de elementos que diferenciem os metadados de documentos nato digitais daqueles que são digitalizados fragiliza o ambiente de preservação digital, uma vez que trata todos os registros como se fossem originais, quando os digitalizados deveriam ser tratados como cópias ou representantes digitais, com requisitos de autenticidade e validade jurídica absolutamente diferenciados.

Dentre as instituições pesquisadas, verificou-se que já existem aquelas que diferenciam os documentos entre nato digitais e digitalizados por meio de booleanos que marcam como “virtual” os documentos gerados em ambiente externo ao sistema, e como “escrito” aqueles produzidos diretamente no SIPAC.

O terceiro tipo de metadado indicado com maior frequência foi o campo “identificador”. Verifica-se que nenhuma das instituições apresentou uma descrição para esse campo, embora outros campos tenham sido descritos. Isto é, as respostas apontam ser consenso entre os responsáveis pela criação dos metadados a existência de um campo com esta nomenclatura. Infere-se, portanto, que o identificador se caracteriza por uma numeração com a função de identificar cada registro criado de forma individual dentro da estrutura do sistema. Contudo, tem-se que fazer o destaque de que as instituições se utilizam de termos diferentes para representar a mesma informação no que se refere ao nome dos metadados. Por exemplo: enquanto a UFRN utiliza o termo “identificador persistente”, UFGD, UFMA e UNIVASF utilizam o termo “identificador”.

Essa situação se repete com inúmeros outros termos, e a falta de padronização na utilização de termos para finalidades semelhantes representa uma dificultador para a interoperabilidade dos sistemas e, conseqüentemente para a preservação digital dos registros. Constitui-se, portanto, como um desafio para a área de Arquivologia oferecer contribuições no sentido de diminuir a assimetria terminológica dos tipos de metadados de preservação digital do SIPAC.

O quarto metadado com a maior incidência, “número de páginas”, apresenta diferentes problemas. Primeiro, o fato de os documentos escritos diretamente no sistema serem exibidos de forma contínua, isto é, sem a paginação, o que torna tal campo desnecessário para os documentos nato digitais. Em segundo lugar, verificou-se na amostra que algumas instituições adotam a digitalização dos versos das páginas dos documentos mesmo quando eles estão em branco, enquanto outras só digitalizam as páginas com conteúdo, de modo que, nos casos de processos administrativos digitalizados, dados de menor volume contidos em partes não significativas das páginas capturadas podem ser excluídos de forma errônea. Por isso, conforme ocorre com o campo “data de criação”, a utilização de termos comuns para registros nato digitais daqueles que são digitalizados pode acarretar prejuízos para a preservação dos registros em ambiente digital.

O quinto elemento de metadado com maior incidência entre as IFES foi o campo “unidade de origem”, cuja função é indicar o setor de lotação do usuário responsável pela criação ou recebimento do documento que dá origem ao registro digital. Esse campo é passível de reflexões sobre a vinculação do setor de lotação do servidor no momento da criação do registro. Ou seja, se um servidor criou um registro num determinado ano em que estava lotado no setor X, mas após alguns anos já não se encontra mais lotado naquele setor, para fins de manutenção da organicidade e do contexto correto de produção daquele documento, é importante que esse campo de metadado fixe a unidade de lotação do servidor para aquela data específica, de maneira que a preservação dessa informação não seja alterada ou comprometida, seja pela movimentação do servidor ou pela alteração no nome do setor em possíveis reformas administrativas.

No segundo grupo de campos de metadados, no qual aparecem aqueles com duas ocorrências, iremos realizar a análise de forma coletiva para alguns elementos. O primeiro item, “assunto detalhado”, se configura como um campo de livre escrita para o usuário descrever o conteúdo ou a finalidade do documento. Entretanto, embora esse item seja importante para facilitar a recuperação das informações, ele não é considerado um item obrigatório pelo sistema, para a criação de um registro. Ou seja, sem uma descrição obrigatória e com a falta de classificação automática já mencionada na análise do campo “assunto”, as informações de contexto do documento podem não ser adequadamente preservadas.

São mencionados ainda os campos “assunto detalhado – para documento” e “observação – para documento”. O primeiro cumpre exatamente a mesma função do campo “assunto detalhado”, porém com a vinculação aos documentos avulsos. O segundo é destinado a apresentar informações adicionais ao assunto detalhado, também no caso de documentos avulsos, sendo, na avaliação desta pesquisa, um item de metadado redundante em relação aos que já foram apresentados.

Percebe-se, de acordo com as respostas recebidas das IFES, que em algumas instituições os metadados associados a um processo administrativo são tratados de forma diferenciada dos metadados associados a documentos avulsos, por isso a expressão “para documento” após o termo que nomeia o campo de metadado. Nas instituições que adotam essa abordagem, também existe um campo “para processo” e, em algumas, outros campos “para memorando” e “para ofício”.

Os dados demonstram ser uma prática comum entre as IFES participantes desta pesquisa: separar os processos administrativos dos documentos avulsos, de modo que sejam criados e mantidos metadados diferentes para cada espécie. E ainda, a criação de um metadado específico para memorandos, que após alteração no manual de redação oficial da presidência da república, em alguns órgãos – inclusive a UFPB –, passou a ser substituído pelo ofício.

Os estudos realizados por esta pesquisa apontam para a necessidade de metadados que preservem o contexto e a descrição dos documentos. Entretanto, a criação de campos específicos para diferentes espécies não parece ser o melhor caminho para essa demanda.

Embora não seja intenção deste trabalho propor um modelo de metadados descritivos para a preservação digital, a literatura indica que um único campo intitulado “assunto” poderia ser utilizado para diferentes espécies documentais sem prejuízo de finalidade. Observa-se, no caso das IFES analisadas, que o estabelecimento de metadados com diferentes termos para processos administrativos, documentos avulsos e ofício, além de reunir numa mesma seara todos os documentos que não sejam processos administrativos ou ofícios, ocasionam duas dificuldades: limitam a três as possibilidades de descrição, e multiplicam por três a necessidade de metadados para a mesma finalidade, aumentando o retrabalho, diminuindo as possibilidades de interoperabilidade e, conseqüentemente, de eficiência.

Uma alternativa possível para a diminuição dos problemas de ordem terminológica resultantes da diferente aplicação de termos pelas instituições usuárias do SIPAC seria a criação de um dicionário unificado no sistema, porém, tal ação requer um esforço interinstitucional e multiprofissional de padronização.

Os itens “data de cadastro”, “data do documento” e “data do *upload*” reforçam a falta de padronização existente para a utilização de campos de metadados nas IFES. Tal situação se amplia quando se percebe que algumas instituições não realizam a separação entre os metadados aplicados aos documentos digitalizados daqueles produzidos no ambiente digital, e outras fazem uma separação apenas por espécie documental como já mencionado.

Por exemplo, enquanto a UFGD possui um campo “data de criação” para documentos nato digitais e outro “data de inserção” para documentos digitalizados – cenário bastante coeso com a proposta da literatura –, a UFS possui um campo “data de autuação” para a espécie documental processo e outro “data de criação” para outras espécies documentais, porém, sem a distinção nato digital e digitalizado – cenário inadequado de acordo com a análise dos autores utilizados por esta pesquisa.

Colocar em um mesmo campo de metadado a data que um documento foi criado no sistema, independente da espécie, e a data em que ele foi recebido de forma

digitalizada ou ainda a data em que ele foi inserido no sistema, configura-se como um equívoco para a preservação digital, pois torna ineficiente a gestão da temporalidade e a destinação desses documentos.

Situação semelhante ocorre com os campos “tamanho” e “tamanho do arquivo”. Embora esse campo tenha a finalidade de descrever o tamanho do registro em bits no sistema, a ausência de uniformidade pode prejudicar a necessidade de interoperabilidade futura entre os sistemas.

O campo “nome do arquivo” se apresenta como responsável pela preservação do nome original do arquivo antes da sua inserção, no caso de arquivos anexados. Destacamos a redação dada pela UFGD, que o configurou como “nome do arquivo com extensão”, de maneira a preservar também o formato do documento digital inserido, ou seja, nessa instituição, se um documento produzido em PDF com nome “arquivo” for incluído, o metadado armazenado será “arquivo.pdf” e não apenas “arquivo”, de modo que, conforme preconiza a literatura, outro elemento de preservação do contexto técnico será mantido.

O último campo, “tipo do documento”, de acordo com as respostas, tem como função descrever qual a espécie do documento, se é ata, relatório, certidão, etc. Porém, de acordo com os conceitos arquivísticos, para de fato ser caracterizado como tipo documental, precisaria unir a espécie documental com a atividade, por exemplo, ata da reunião do conselho universitário, relatório de atividade de extensão, certidão de aprovação de afastamento docente para pós-doutorado, etc.

As IFES adotam diferentes possibilidades para esse campo, a começar pela forma como o nomeiam. Se consideradas todas as formas constantes nas respostas, o número de ocorrências deixaria de ser dois e passaria a sete, ocupando, portanto, o topo da lista, tendo sido informado por UFGD, UFS, UFMA, UNIVASF, UNILA, UFPB e UFRB, com destaque para a descrição informada pela UFGD, que escolheu o termo “tipo de conteúdo” para indicar a mesma finalidade.

Essa discussão é permeada de detalhes simples, mas importantes para definir se a preservação digital acontecerá de forma integral ou parcial. Nesse sentido, considera-se importante mencionar que, no caso dos marcadores booleanos utilizados para definir um registro como “virtual” ou “escrito”, é preciso lembrar que existem registros no SIPAC que não se configuram nem como documentos natos digitais, nem como documentos digitalizados. Tratam-se, portanto, dos processos e documentos que foram criados de forma analógica ou física, cuja tramitação foi registrada no SIPAC, isto é, o sistema contém apenas as informações de distribuição e movimentação dos documentos. A UFPB possui tais registros. Sendo assim, esse se torna outro elemento necessário para reflexão: podem os registros de tramitação ser tratados da mesma forma que os registros que contém conteúdo de documentos, sejam eles natos digitais ou digitalizados? Como deve ser a preservação dos processos que iniciaram físicos e tornaram-se digitais? Esses documentos, conhecidos como híbridos, carecem de metadados específicos para a sua preservação, que contemplem, dentre outros fatores, a existência de normas institucionais que detalhem a necessidade ou não de digitalização do conteúdo em suporte analógico, bem como as informações quanto ao local de arquivamento dos documentos produzidos no momento anterior à migração para o suporte eletrônico.

Em relação a análise das respostas ao quesito 2 – A sua instituição armazena metadados que identificam a origem e a localização dos usuários que produzem os registros e documentos? Caso sim, quais metadados são armazenados?

Das treze instituições participantes, apenas a UFERSA informou não armazenar metadados de identificação da origem e localização dos usuários que

produzem registros e documentos em ambiente digital. As outras doze IFES informaram um total de 43 tipos de metadados utilizados para essa finalidade, com destaque, para o campo “endereço IP”, com quatro ocorrências (UFGD, UFABC, UNILA e UFRB), seguido do campo “IP da máquina do usuário”, com duas ocorrências (UFLA e UFRN); juntos, representam 14% das respostas sobre essa questão.

Considerando que a UFPB é a base deste estudo, destaca-se que a resposta fornecida, embora não contenha o campo IP, apresenta como opções de metadados os itens “login do usuário”, “nome do usuário”, “matrícula SIAPE do usuário”, “unidade de lotação do usuário” e “data de criação”.

Convém esclarecer que os metadados relacionados ao IP dizem respeito à identificação do dispositivo utilizado pelo usuário quando do acesso ao sistema para o cadastro de um documento, sendo possível inferir que os dois termos utilizados pelas universidades possuem significação equivalente e a existência de dois termos se dá pela ausência de padronização terminológica da área.

A importância de se registrar o IP para a finalidade digital é inquestionável, entretanto, tal metadado também se configura como um importante elemento arquivístico de preservação dos registros digitais, haja vista que pode oferecer o contexto testemunhal que assegure ou não a autenticidade de um registro. A título de exemplo, a captura do IP de um documento que só possa ser produzido pelo reitor e tenha sido criado a partir de um dispositivo que ele não utiliza eventualmente pode contribuir na verificação da autenticidade desse documento.

Os únicos dois outros campos de metadados que apresentaram duas ocorrências, foram justamente “login do usuário” e “unidade de origem – para processo”, elementos presentes na resposta da UFPB.

Verifica-se que foram informados pelas IFES uma média 4 campos de metadados para identificação de origem e localização dos registros digitais. Embora a UFPB não tenha indicado o endereço IP como resposta a essa questão, é preciso destacar que a universidade informou que faz o armazenamento desse metadado para outra finalidade: a de descrição do dispositivo.

Destacamos que a UFOB respondeu que armazena o metadado “Assinante do documento”, entretanto, não faz diferença entre os metadados dos documentos nato digitais e dos digitalizados. No SIPAC, o assinante do documento nato digital é o servidor com a competência institucional para a assinatura daquele documento, portanto, autor ou participante da ação documentada. De forma oposta, no caso dos documentos digitalizados, o SIPAC identifica como assinante o servidor responsável pelo recebimento daquele documento para a inclusão no sistema, sendo, portanto, uma assinatura de autenticação, mas não de autoria.

A falta de metadados de identificação do assinante que diferenciem o autor do documento nato digital, daquele que é responsável pela autenticação de um representante digital, coloca em risco a preservação digital desses registros, uma vez que não se pode confundir o autor de um documento com alguém que autentica uma cópia de documento produzido por outrem, de tal maneira que a utilização de um único metadado para as duas situações configura-se como um elemento passível de interpretação equivocada quando da necessidade de acesso futuro a essa informação.

Ainda sobre esse quesito, destacamos a existência de metadados informados exclusivamente por duas instituições. A UFRN foi a única universidade a indicar o armazenamento de metadado de geolocalização. Do mesmo modo, a UFGD foi a única a informar o armazenamento das telas e *URL*'s acessadas pelo usuário, ou seja,

a IFES realiza a captura e guarda de cada um dos endereços eletrônicos acessados pelos usuários do seu sistema.

O armazenamento da geolocalização e da informação sobre as URL's acessadas podem colaborar com a gestão e preservação ao oferecer a possibilidade de auditoria e de verificação da autoria dos registros.

Partindo para a questão 3 - A sua instituição armazena metadados que diferenciem os registros e documentos gerados em ambiente digital (nato digitais) daqueles convertidos para o meio eletrônico (digitalizados)?

Três instituições disseram não realizar armazenamento desse tipo de metadados (UFERSA, UFOB e UNIVASF). A UFABC informou que, por não ser a desenvolvedora do *software*, preferia que as respostas fossem solicitadas à UFRN, responsável pela criação do sistema. Sendo assim, as respostas a essa questão foram fornecidas por UFGD, UFS, UFMA, UNILA, UFPB, UFLA, UFRB, UFSJ e UFRN.

Destaque-se que a UFOB, que contém o maior número de campos de metadados de preservação digital informado dentre as IFES respondentes, não diferencia metadados de digitalizados dos metadados dos registros nato digitais, o que corrobora o entendimento de que a quantidade de campos não reflete a qualidade dos metadados de acordo com os requisitos propostos pela literatura.

Das nove instituições respondentes a esse quesito, foram obtidos 62 (sessenta e dois) campos de metadados dedicados à diferenciação entre registros nato digitais e registros provenientes de digitalização. Os dados obtidos revelam que o metadado mais utilizado pelas instituições para a distinção entre os registros gerados em ambiente digital daqueles digitalizados é justamente um campo que responde a essa questão, nomeado de diferentes formas, mas com a mesma finalidade. O item corresponde a 23% das respostas e suas variações podem ser identificadas no quadro 5.

Quadro 5 – Metadados de diferenciação entre nato digital e digitalizado nas IFES que utilizam o SIPAC

| Nome do campo de metadado | IFES |
|---|-------|
| Sinalizador marcando que se trata de um nato digital | UFGD |
| Sinalizador marcando que se trata de um digitalizado | |
| Forma do documento – nato digital | UFS |
| Forma do documento – digitalizado | |
| Sinalizador se o documento é anexado | UFMA |
| Sinalizador se o documento é nato digital | |
| Booleano que indica que o documento é virtual e que foi escrito pelo próprio usuário dentro do sistema. Caso contrário, é um documento virtual cujo conteúdo está em um arquivo anexado ou é um documento não virtual (representação de um documento físico); | UNILA |
| Booleano que indica se o documento é eletrônico ou físico; | |
| Para os nato digitais – marcação como escrito | UFPB |
| Para os digitalizados – marcação como anexado | |
| Booleano para virtual | UFRB |
| Booleano para escrito | |
| Para documentos – forma do documento (nato digital ou anexado) | UFSJ |
| Para processos – processo eletrônico (sim ou não) | |

Fonte: dados da pesquisa, 2020.

A UFLA, embora não tenha informado um campo específico para indicar se o registro é nato digital ou digitalizado, apresentou metadados diferentes para cada condição dos registros, indicando que a instituição faz a distinção entre os diferentes tipos de registros, embora não possua um campo específico para determinar essa finalidade.

Em sequência, o outro campo de metadado relacionado a esse quesito é o item correspondente ao tipo de conferência realizada quando o registro se trata de um documento digitalizado. O item corresponde a 6% das respostas, conforme o quadro abaixo.

Quadro 6 – Metadados acerca do tipo de conferência para os documentos digitalizados nas IFES que utilizam o SIPAC

| Nome do campo de metadado | IFES |
|---|------|
| Tipo de conferência do digitalizado | UFGD |
| Tipo de conferência do digitalizado | UFS |
| Tipo de conferência para digitalizados | UFMA |
| Para documentos digitalizados – tipo de conferência | UFSJ |

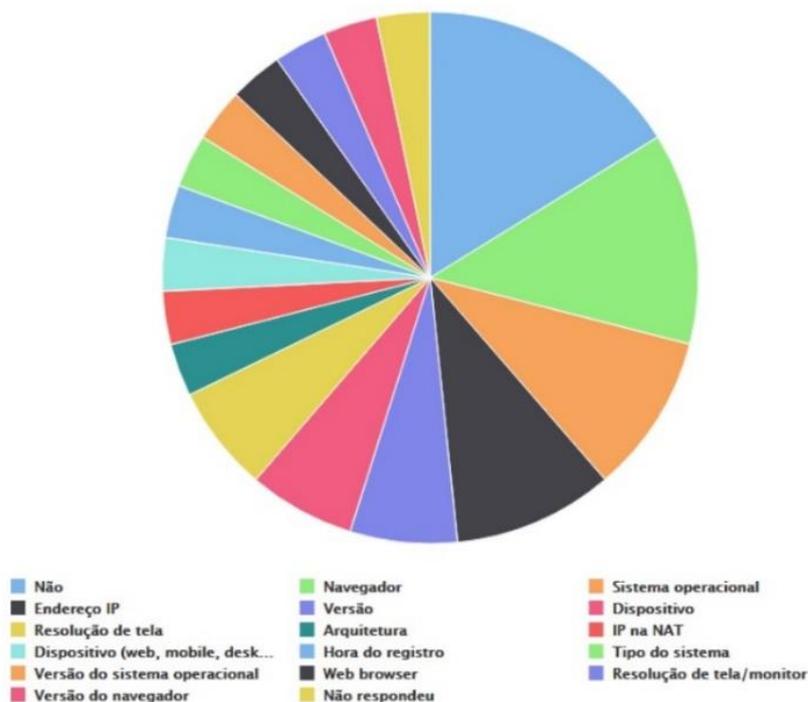
Fonte: dados da pesquisa, 2020.

A UFPB não informou a existência de metadado especificando o tipo de conferência para digitalizados na resposta a esse quesito, porém, indicou a existência do campo “Para digitalizados – Tipo de conferência (original, cópia autenticada em cartório, cópia autenticada administrativamente ou cópia simples)” na resposta ao quesito 6, tendo sido, inclusive, a única instituição a descrever qual tipo de conferência era realizada.

Verifica-se que, dentre as instituições que responderam a essa questão, a média de campos de metadados dedicados à finalidade de diferenciar nato digitais de digitalizados é de 20 itens. Na resposta desse item, a UFPB informou apenas 5, contudo, conforme apontado anteriormente, as respostas aos quesitos se complementam e o número de campos indicado não significa maior ou menor grau de aderência à categoria em questão. A análise dos aspectos positivos e negativos do conjunto de metadados utilizados pela UFPB será discutida até o final deste artigo.

Em continuidade, as respostas ao quesito 4 - A sua instituição armazena metadados que descrevam o dispositivo e o sistema operacional no qual os registros e documentos foram produzidos? Caso sim, quais metadados são armazenados a respeito do dispositivo e do sistema operacional?

Gráfico 2 – Metadados de descrição do dispositivo e sistema operacional armazenados pelas IFES que utilizam o SIPAC



Fonte: dados da pesquisa, 2020.

Das treze instituições respondentes, 5 cinco IFES responderam que não armazenam metadados de descrição do dispositivo e sistema operacional. São elas: UFRN, UFERSA, UFRB, UFOB e UFMA. Similarmente, 2 instituições – UFS e UFSJ – deixaram esse quesito sem resposta, diminuindo para 6 o número de universidades que informaram quais metadados são armazenados para essa finalidade: UFGD, UFABC, UNIVASF, UNILA, UFB e UFLA.

Dentre as instituições que indicaram uma resposta, foram apresentados 25 campos de metadados, ou seja, as IFES armazenam uma média de 4 campos de metadados para identificação do dispositivo ou sistema operacional. Dentre eles, aquele com maior incidência é o navegador ou *web browser*, com cinco ocorrências. Em seguida, está o endereço IP, com três ocorrências. Por fim, com duas ocorrências aparecem os campos “dispositivo”, “resolução de tela” e “versão”.

Destaque-se que o item “dispositivo” tem como finalidade a identificação de qual tipo de equipamento está sendo utilizado para a produção do registro, por exemplo, computador, tablet ou celular. Uma das instituições apresentou como possibilidades para esse campo as opções *web*, *mobile* e *desktop*. No campo “resolução de tela”, que também foi apresentado com a nomenclatura “resolução de monitor”, a intenção é obter a informação de qual era a disposição do “documento manifestado” para o usuário que o produziu ou inseriu. Já o item “versão”, não explicado de maneira clara pelos órgãos respondentes, indicaria a versão do sistema utilizado quando da criação do registro, de modo a preservar o ambiente técnico necessário para a preservação do documento ao longo do tempo de acordo com as características do *software*.

Chama a atenção o fato de a UFGD ser a única instituição a possuir um metadado destinado a preservar a arquitetura do registro. Entende-se que a adoção de um campo com a intenção de armazenar informações para visualização do documento, de acordo com as características computacionais necessárias para sua visualização da forma como foi produzido, atende de forma *sui generis* às recomendações da literatura para a preservação digital.

A UFPB apresentou 7 campos de metadados destinados à descrição do dispositivo e sistema operacional na resposta a esse questionamento.

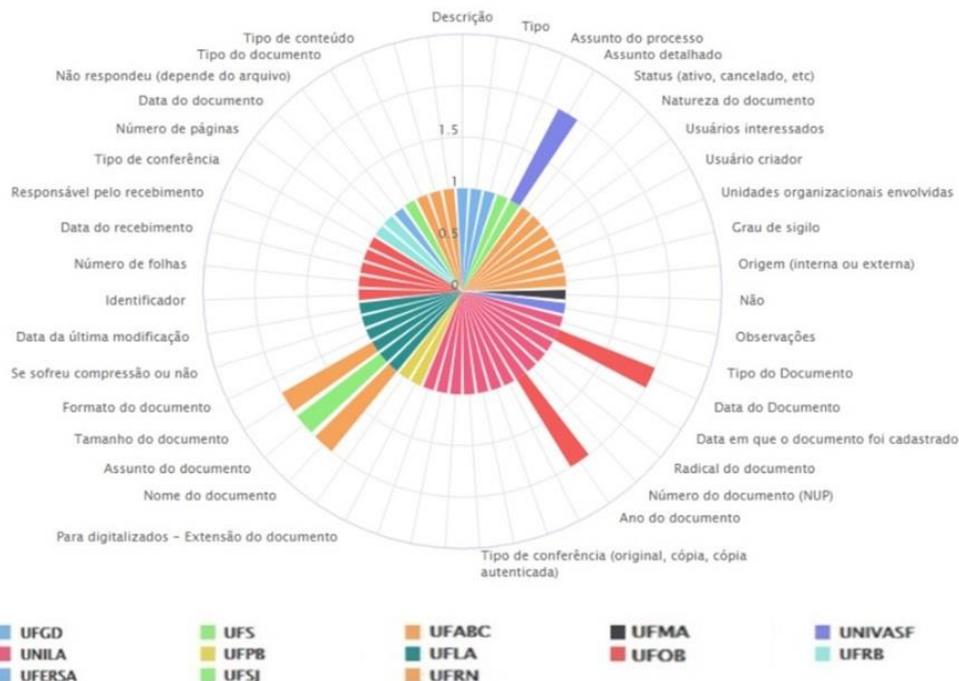
Outro fato que chama a atenção é o caso da UFRN, instituição proprietária do SIPAC, ter respondido que não armazena metadados destinados à descrição do dispositivo e sistema operacional nessa questão. Cabe destacar que, na resposta ao quesito 2, que trata dos metadados que identificam a origem e a localização dos usuários, a instituição já havia indicado o armazenamento dos seguintes campos: “informações do sistema operacional” e “informações do dispositivo”. Ou seja, a UFRN armazena metadados para a finalidade dessa questão, embora não tenha apresentado respostas no quesito 4.

Analisemos o quesito 5 – A sua instituição armazena metadados que descrevam os dados constantes nos registros e documentos criados/inseridos nos sistemas SIGAA/SIPAC (ou nos sistemas de produção de documentos acadêmicos e administrativos)? Caso sim, quais são os metadados armazenados para descrição dos dados?

Das treze IFES respondentes, apenas UNIVASF e UFERSA não apresentaram respostas a esse quesito. A primeira respondeu que não armazena esse tipo de metadado e a segunda, embora tenha dito que armazena, informou que “esses dados dependem do arquivo que está sendo inserido no sistema, ficando impossível de dizer exatamente quais são eles visto a variedade de tipos de documentos criados/inseridos nos SIGs”.

O gráfico 3 contém a ilustração das respostas obtidas pelas instituições participantes da pesquisa sobre os metadados que descrevem os dados constantes nos registros digitais.

Gráfico 3 – Metadados descritivos de dados armazenados pelas universidades federais que utilizam o SIPAC



Fonte: dados da pesquisa, 2020.

Identificam-se diversos campos de metadados nas respostas, entretanto, conforme o gráfico demonstra, apenas 6 campos obtiveram destaque quanto ao número de ocorrências, são eles: “ano do documento”, “assunto detalhado”, “assunto do documento”, “data do documento”, “nome do documento” e “tamanho do documento”. Todos esses campos já foram objeto de análise desta pesquisa, com exceção do “ano do documento”, cuja função é registrar o ano em que o registro foi criado.

Considerando que as respostas dessa questão apresentaram campos de metadados bem diferentes quando comparados com as respostas das questões anteriores, e informados de maneira exclusiva por algumas instituições, destacamos os itens constantes no quadro abaixo:

Quadro 7 – Metadados de descrição de dados de universidades federais que utilizam o SIPAC

| Descrição do campo de metadado | Universidade |
|---|--------------|
| Tempo de vida | UFRN |
| Se sofreu compressão ou não | UFLA |
| Código de verificação para averiguar a autenticidade do documento | UNILA |
| Numeração original do documento (no caso de documento externo) | |
| Unidades organizacionais envolvidas | UFABC |

Fonte: dados da pesquisa, 2020.

Verifica-se, pela UFRN, a aplicação do conceito de tempo de vida. Sabe-se que “ciclo de vida” é uma abstração utilizada tanto pela Arquivologia para a definição das idades documentais, como por outras ciências, para a gestão de objetos digitais de diferentes naturezas. Embora não tenha ficado evidente se tal metadado está relacionado à temporalidade arquivística ou à definição do intervalo entre a data de criação e a data da consulta do registro, o armazenamento de tal informação – seja uma ou outra – constitui-se como importante elemento para a preservação digital.

A informação sobre “se sofreu compressão ou não”, constante na resposta da UFLA, demonstra uma importante preocupação com a preservação do contexto técnico do registro, fator preponderante na definição de uma política de preservação, uma vez que a não preservação desse tipo de dado pode comprometer, por exemplo, a visualização de imagens JPEG.

Do mesmo modo, os três outros campos de metadados informados pela UNILA e UFABC, “código de verificação de autenticidade”, “numeração original de documento externo” e “unidades organizacionais envolvidas”, fornecem elementos de contexto também considerados relevantes pela literatura de preservação digital de registros arquivísticos.

Dando sequência, considerando as respostas de todas as instituições, foram informados 51 campos de metadados destinados à preservação dos dados contidos nos registros digitais pelas 11 universidades que responderam satisfatoriamente a essa questão, de modo que a média de itens armazenados para essa finalidade é de 5 campos.

Constata-se que a instituição com o maior número de metadados dedicados a essa finalidade é a UNILA, que informou armazenar 13 campos de metadados para descrição de dados dos seus registros. Em seguida, estão a UFABC e a UFOB, com 7 campos de metadados cada. Logo após a UFLA, com 6; seguida da UFRN, com 5 campos.

A verificação da aderência desses metadados ao modelo baseado em Rogers (2015) e Sayão (2010) aponta que nenhuma instituição apresenta um conjunto de metadados que possa ser considerado referência na descrição das informações de *hardware*, *software*, dependência entre os componentes digitais e demais itens necessários à preservação digital do aspecto computacional. Ao contrário, as respostas obtidas demonstram um cenário bastante preocupante quanto ao armazenamento dessas informações.

A UFPB informou armazenar apenas dois campos de metadados para essa finalidade, a saber, “para digitalizados – extensão do documento” e “para nato digitais – conteúdo do documento em forma HTML”. Ou seja, a UFPB, embora quantitativamente abaixo da média das demais universidades em relação ao número de campos de metadados de descrição dos dados contidos em seus registros digitais, é a única a armazenar conteúdo em forma HTML de acordo com as respostas recebidas. É preciso destacar, ainda, que outros campos de metadados que também se configuram como descritores de dados já estão sendo armazenados pela UFPB, conforme respostas a questões anteriores.

Destaca-se que a simples diferenciação entre nato digitais e digitalizados ainda deixa de fora alguns elementos arquivísticos relevantes do contexto de produção dos registros digitais. Por exemplo, se uma certidão foi criada diretamente no sistema como um registro nato digital, logicamente ela é um documento produzido pela UFPB. Porém, uma certidão incluída como documento digitalizado, tanto pode ser sido criada pela UFPB, como pode ter sido criada por um outro órgão, e não existe metadado sendo armazenado que estabeleça essa diferença. Isto é, esses registros estão sendo

tratados de forma igual, quando são absolutamente diferentes do ponto de vista da responsabilidade institucional sobre eles.

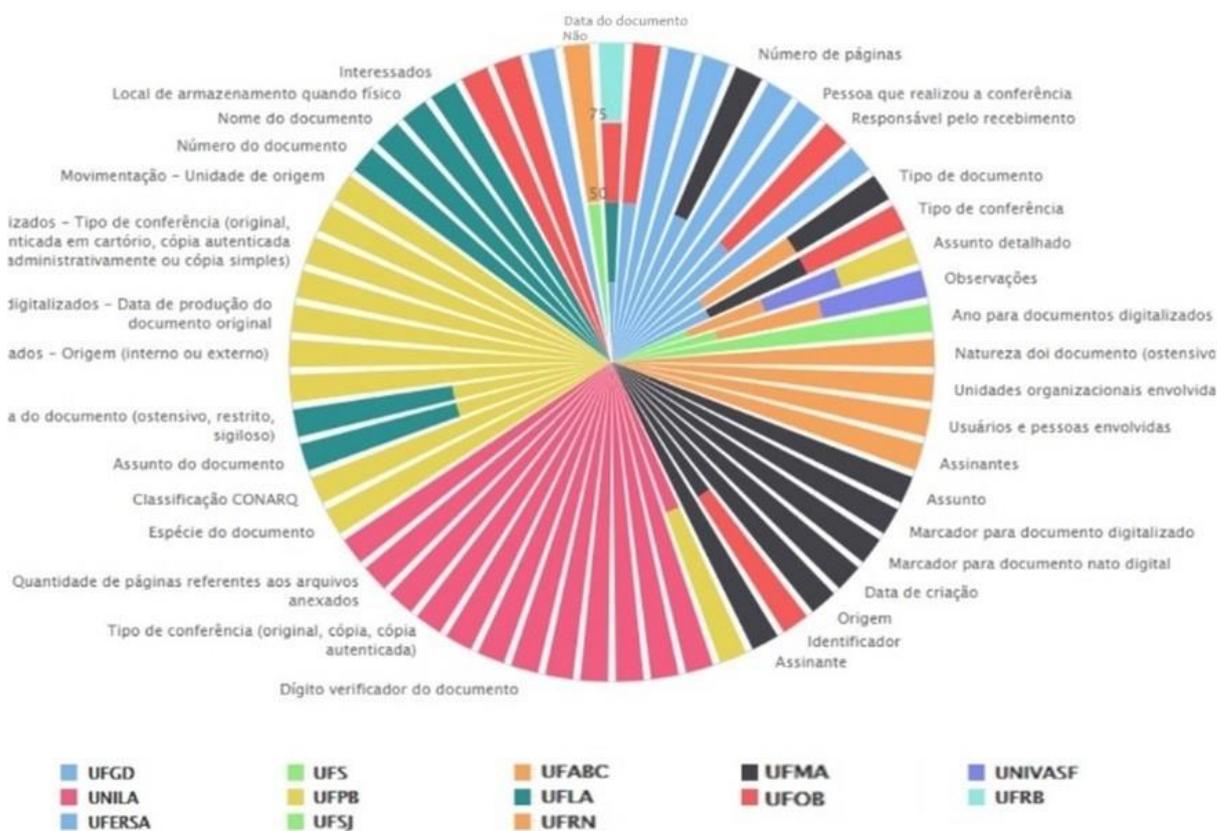
Os digitalizados provenientes de documentos originais produzidos pela UFPB precisam ter informações quanto à localização física dos seus originais a eles associados, enquanto os digitalizados provenientes de documentos recebidos pela UFPB precisam estar atrelados a metadados que traduzam de forma fidedigna o seu contexto de produção, e não apenas o tipo de conferência realizada para a digitalização.

Passaremos agora à análise do quesito 6 - Existem metadados que descrevam os dados constantes nos registros e documentos criados/inseridos nos sistemas SIGAA/SIPAC (ou nos sistemas de produção de documentos acadêmicos e administrativos) que são incluídos pelo usuário de forma manual? Caso sim, quais são os metadados armazenados para descrição dos dados inseridos manualmente pelo usuário?

As instituições UFRN e UFSJ responderam que não armazenam esse tipo de metadado. Já a UFRSA, informou que armazena, contudo, que “esses dados dependem do arquivo que está sendo inserido no sistema, ficando impossível de dizer exatamente quais são eles visto a variedade de tipos de documentos criados/inseridos nos SIGs”.

As respostas das outras universidades federais que foram consideradas satisfatórias para as finalidades deste trabalho podem ser visualizadas no gráfico 4.

Gráfico 4 – Metadados descritivos incluídos de forma manual pelo usuário nas universidades federais que utilizam o SIPAC



Fonte: dados da pesquisa, 2020.

É possível perceber, com base no gráfico acima, que UNILA e UFPB se destacam como as IFES com o maior número de metadados descritivos de dados inseridos de forma manual pelo usuário. As duas instituições informaram armazenar 13 campos de metadados cada para esta finalidade.

Chama a atenção o fato de a UNILA ter sido a única instituição a mencionar o armazenamento do NUP – Número Único de Protocolo –, demanda de padronização da tramitação de documentos eletrônicos regulamentada, de utilização obrigatória para todos os órgãos públicos federais.

Os campos de metadados informados pela UFPB como resposta ao quesito 6 foram: “tipo do documento”, “espécie do documento”, “classificação CONARQ”, “assunto do documento”, “assunto detalhado”, “natureza do documento”, “hipótese legal da restrição de acesso – para documentos restritos”, “origem – para documentos digitalizados”, “data de produção do documento original – para documentos digitalizados”, “data de recebimento do documento original – para documentos digitalizados”, “tipo de conferência – para documentos digitalizados”, “movimentação – unidade de destino” e “movimentação – unidade de origem”.

Salienta-se que, no campo “natureza do documento”, existem as opções de “ostensivo”, “restrito” ou “sigiloso”. No campo “origem – para documentos digitalizados”, as opções são “interno” ou “externo”. Já no campo “tipo de conferência – para documentos digitalizados”, conforme mencionado em análise anterior, as opções são “original”, “cópia autenticada em cartório”, “cópia autenticada administrativamente” ou “cópia simples”.

Ao considerar o conjunto das respostas de todas as IFES, os destaques que se apresentam foram os campos “assunto detalhado” e “data do documento”, com quatro incidências cada. Em seguida, com três ocorrências cada, “observações”, “tipo de conferência” e “tipo de documento”, todos já descritos em análises das questões anteriores.

Mediante o exposto, verifica-se que 100% das universidades federais brasileiras constantes nesta análise armazenam algum tipo de metadado para a preservação digital dos registros digitais arquivísticos criados no SIPAC. Contudo, constata-se a falta de padronização na elaboração dos nomes utilizados para os campos, aspecto considerado um problema para a interoperabilidade interinstitucional de sistemas no futuro.

Além disso, também é possível perceber que algumas IFES possuem problemas de ordem terminológica na composição dos seus metadados. Tal constatação se baseia na ausência de conexão entre o conceito de termos arquivísticos utilizados e sua aplicação prática no SIPAC.

No caso da UFPB, são armazenados metadados específicos para a preservação digital dos registros digitais criados no SIPAC, relacionados a todas as categorias analisadas por esta pesquisa. No entanto, as respostas da UFPB, assim como das outras IFES participantes, limitaram-se aos descritores automático e manuais do SIPAC Módulo Protocolo, sem mencionar os metadados existentes para os outros módulos, SIGAA e demais sistemas de criação de registros digitais, igualmente importantes para a preservação digital dos registros documentais da instituição em ambientes digitais.

A complexidade dos registros digitais arquivísticos envolve, embora não se limite a ela, a compreensão dos elementos arquivísticos necessários para a sua preservação digital. Contudo, um dos principais desafios para essa demanda está em transpor requisitos pensados tradicionalmente sob a perspectiva dos documentos

analógicos, para requisitos digitais com a finalidade de preservação do contexto e dos requisitos arquivísticos dos registros digitais.

De acordo com o modelo conceitual de requisitos para metadados de preservação digital, baseado no trabalho de Rogers (2015) e Sayão (2010), identificaram-se categorias sob as quais esses metadados precisam ser criados.

A primeira, relacionada à origem, localização e descrição da infraestrutura conceitual que descreve o ambiente, precisa incluir, para além das informações sobre setor de lotação do usuário que criou o registro, outros metadados que tragam luz ao caráter orgânico e contextual dos documentos arquivísticos, tais como informações sobre a vinculação daquele registro a alguma comissão, núcleo, projeto, grupo de trabalho ou qualquer outro tipo de unidade que possa não estar sendo identificada pelo sistema de produção do registro, como se percebe, no caso da UFPB, a falta de identificação do campus quando da produção de registros no SIPAC.

A segunda categoria, aparência e desempenho, nos metadados encontrados por esta pesquisa, faz menção ao formato do arquivo digital e outras informações de cunho tecnológico. Mas é preciso lembrar que, do ponto de vista arquivístico, existem elementos de aparência que traduzem aspectos diplomáticos e contextuais predominantemente arquivísticos e que não estão sendo capturados. Por exemplo, os registros audiovisuais criados pela plataforma de conferência *web* da Rede Nacional de Pesquisa possuem atributos visuais que os diferenciam daqueles criados pela plataforma *Google Meet* – ambos utilizados para gravação de atividades acadêmicas e administrativas da UFPB.

Em vista disso, preocupa-nos que, no futuro, sem essa informação, tais registros possam ter a sua integridade ou outros aspectos de sua autenticidade comprometidos. Isto é, não identificar a plataforma em que o registro foi criado originalmente antes de ser inserido no SIPAC é apenas um exemplo da necessidade de se criar um padrão para metadados de preservação arquivística de registros digitais.

Passando para a terceira categoria, que se refere à descrição do dispositivo e sistemas utilizados, entendemos que, embora eminentemente tecnológica, pode contribuir de forma efetiva com a manutenção dos requisitos arquivísticos, ao incluir informações sobre documentos inseridos provenientes de outros sistemas. Dessa forma, quando um documento que foi criado em outro *software* de governo precisar ser inserido no SIPAC, deveria ser possível identificar qual o sistema utilizado para a produção daquele registro, como por exemplo, o Sistema de Concessão de Diárias e Passagens (SCDP), Sistema Integrado de Administração Financeira (SIAFI), Sistema de Gestão de Pessoas (SIGEPE), entre outros. Embora já seja possível inserir essas informações em campo destinado a observações, a existência de um campo específico favoreceria a preservação e a recuperação de informações desses contextos.

Por fim, na categoria que trata dos metadados de descrição automática e manual dos dados, percebe-se que também existem possibilidades de ampliação que envolvam aspectos arquivísticos. Os eventos de gestão de processos administrativos, como desmembramentos, desentranhamentos, juntadas, são algumas das informações que não estão sendo objeto de metadados e que possuem relevante função arquivística. Do mesmo modo, embora a UFPB já possibilite a classificação dos registros de acordo com o Plano de Classificação de Documentos de Arquivo do CONARQ, existem alguns problemas do ponto de vista arquivístico que merecem uma reflexão.

Para cada código de classificação utilizado, existe uma previsão legal do tempo pelo qual aquele registro digital precisa ser mantido no ambiente de gestão, se há a necessidade de guarda precaucional e, caso haja, por qual período, e qual deve ser a sua destinação final, podendo ser a eliminação ou a guarda permanente. O que chama a atenção é o fato de, na UFPB, a informação quanto ao código de classificação ainda não estar sendo utilizada para a efetiva gestão do ciclo de vida. Ou seja, embora existam registros no SIPAC classificados, não há ainda nenhuma ação efetiva de transferi-los do ambiente de gestão para o ambiente de preservação, pois não estão sendo vinculados os metadados de temporalidade.

Tal situação nos leva para outra reflexão, que é a necessidade de interoperabilidade entre os sistemas para que a gestão dos documentos aconteça de forma plena. Sob esse aspecto, o CCJ possui um relevante volume de documentos que já poderiam ser destinados, porém necessitam de informações constantes de outros sistemas da universidade para que a destinação aconteça e, como os sistemas não interoperam e a UFPB não armazena metadados que subsidiem essa gestão, a destinação fica condicionada à captura manual das informações de vinculação do registro ao evento que determina a sua temporalidade, e a um comando posterior que possibilite a sua eliminação, transferência ou recolhimento ao repositório arquivístico digital confiável, cuja existência e funcionamento ainda estão em processo de implantação.

Destaca-se, no caso específico da UFPB, que a instituição possui grupos de metadados para identificação de origem e localização, de distinção entre nato digitais e digitalizados, para descrição dos sistemas e dispositivos utilizados pelo usuário, para descrição automática de dados e para descrição manual de dados pelo usuário, mas não há a diferenciação precisa de metadados de preservação para os níveis lógico, físico e conceitual, três diferentes níveis necessários para a preservação no ambiente digital.

Deveria existir um conjunto de metadados específicos para a preservação do nível lógico (entendido como a cadeia de bits que precisa de informações contextuais em linguagem de máquina para ser compreendido), um conjunto de metadados relacionados à preservação do objeto digital no nível físico (entendido como o artefato ou o *hardware* utilizado e necessário para armazenamento e visualização do documento e as intervenções de migração e alterações de suporte realizadas ou planejadas), e um conjunto de metadados destinados à preservação do objeto digital no nível conceitual (ou seja, descritores do documento digital arquivístico manifestado que contribuam na organização e recuperação da informação pelo usuário).

Constata-se que os metadados podem e devem servir de parâmetro para a gestão dos registros digitais arquivísticos. Contudo, para isso, os sistemas precisam trabalhar de forma integrada, acompanhar as mudanças na legislação, mas também reunir informações contextuais de *hardware*, *software* e elementos descritivos que garantam a preservação dos requisitos arquivísticos dos registros digitais.

6 Conclusão

Dado o caráter dinâmico dos recursos digitais e a forte obsolescência tecnológica decorrente desse universo, é preciso reunir informações que garantam, ou ao menos busquem um nível de segurança razoável, de que o ambiente computacional no qual a informação foi criada poderá ser reproduzido com o passar

do tempo, de modo que os registros permaneçam acessíveis independentemente da existência do formato e das condições técnicas em que foram concebidos.

É justamente o conjunto desses atributos, aqui denominados de metadados, que permitem que os dados constantes de um documento em ambiente digital possam se tornar um documento manifestado e acessível ao usuário no decorrer do tempo, associado ao conteúdo que o fundamenta, e as transações e requisitos que o tornam válido, que constitui um registro digital, que, quando criado para evidenciar e servir de prova das atividades de uma pessoa ou instituição, é um registro digital arquivístico.

O registro digital arquivístico agrega dados e metadados que descrevem autoria, conteúdo e classificação, mais acrescenta elementos como competência, localização do dispositivo utilizado, vínculos institucionais internos e externos, e componentes digitais que o constituem como objeto único ou associado a outros.

Compreendemos que o conceito de registro digital arquivístico ainda precisa de outras reflexões que o esmiúcem e ampliem, porém, o estudo realizado até aqui nos leva à compreensão de que o registro digital arquivístico se caracteriza como uma proposta conceitual mais ampla do que o documento arquivístico digital (ACKER, 2019; RIES, 2018; ROGERS, 2015), pois ressignifica os limites do que é considerando relevante para a preservação dos requisitos arquivísticos dos documentos em ambiente digital, com destaque para os metadados.

A comparação das propostas constantes na literatura com as respostas das universidades federais pesquisadas, incluída a UFPB, sobre os campos de metadados existentes no SIPAC Módulo Protocolo, permitiu visualizar que as IFES brasileiras armazenam metadados para a preservação digital, porém as informações quanto ao aspecto material dos registros digitais são limitadas, não há padronização dos campos de metadados entre as instituições, existem incoerências terminológicas relacionadas aos conceitos arquivísticos, predomínio da atuação de profissionais de tecnologia em detrimento dos profissionais de arquivologia, lacunas de informação sobre as diferenças entre documentos nato digitais, digitalizados e híbridos, além de uma série de questões relacionadas com a ausência de uma política nacional ou no âmbito dos órgãos vinculados ao Ministério da Educação acerca dos critérios para a preservação digital dos registros digitais arquivísticos.

Destacamos que a UFPB reúne elementos de preservação digital para todas as áreas indicadas pela literatura, estando, portanto, alinhada ao modelo conceitual de preservação digital de registros digitais arquivísticos, no que se refere ao conjunto de metadados utilizado para o SIPAC Módulo Protocolo. Entretanto, verificou-se que em todas as categorias de análise existem oportunidades de criação de metadados que ampliem, sob o ponto de vista arquivístico, as possibilidades de preservação. Isto é, não é suficiente ter metadados de preservação digital: é preciso que haja metadados de preservação digital específicos para registros digitais arquivísticos a fim de preservar o contexto arquivístico dos registros digitais.

Embora a maioria dos metadados armazenados pela UFPB sejam relacionados ao contexto descritivo e conceitual dos documentos – aspecto relevante para a preservação digital, ainda é necessário o desenvolvimento de campos que consolidem esse nível da preservação e incluam metadados vinculados aos elementos de *hardware* e *software* necessários para a preservação de longo prazo.

Propõe-se, como sugestão para investigação futura, a verificação da adequação da UFPB no que se refere à preservação digital das demais plataformas utilizadas para a criação de registros digitais arquivísticos (sítio, redes sociais, aplicativos institucionais e os outros módulos do SIG utilizados pelo órgão).

Constata-se a necessidade de reflexão sobre o profissional de arquivo frente a essas demandas. O Arquivista deve tornar-se um agente da curadoria digital, ou ainda, um Arquivista Digital, isto é, um planejador ativo das estratégias de preservação de longo prazo dos registros digitais arquivísticos sob sua responsabilidade. Para isso, o conhecimento sobre metadados (agora parte do acervo) deve fazer parte das suas competências profissionais para que ele possa contribuir com a criação de informações: de conteúdo, que interpretem a cadeia de bits para as máquinas; e de descrição, que facilitem a recuperação para os usuários. Além disso, precisará estar preparado para criar e manter um repositório que contemple, dentre outros elementos, os desafios da privacidade e da propriedade intelectual, como também da gestão do uso e gestão do acesso.

A sociedade do futuro só conhecerá a sociedade do presente se os documentos arquivísticos digitais estiverem acessíveis, e isso só será possível a partir da compreensão dos registros digitais arquivísticos. Entendemos que esta pesquisa inicia a discussão, sem a intenção de esgotá-la ou limitá-la.

Referências

ACKER, A. Quando é um registro? Estrutura de pesquisa para localização de registros eletrônicos na infraestrutura. GILLILAND, A. J.; MCKEMISH, S.; LAU, A. J. (org.). **Pesquisa no multiverso arquivístico**. Tradução de Ana C. Rodrigues. Salvador: 9Bravos, 2019.

ARQUIVO NACIONAL (Brasil). **Dicionário Brasileiro de Terminologia Arquivística**. Rio de Janeiro: Arquivo Nacional, 2005.

CONSELHO NACIONAL DE ARQUIVOS (BRASIL). Câmara Técnica de Documentos Eletrônicos. **Resolução nº. 43**, de 4 de setembro de 2015. **Estabelece diretrizes para a implementação de repositórios arquivísticos digitais confiáveis para o arquivamento e manutenção de documentos arquivísticos digitais em suas fases corrente, intermediária e permanente, dos órgãos e entidades integrantes do Sistema Nacional de Arquivos - SINAR**. Disponível em: <https://bit.ly/41OfYBf>. Acesso em: 21 abr. 2023.

ENCONTRO NACIONAL DE ARQUIVISTAS DAS INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR, 6., 2019, Goiânia. **Anais**. Goiânia: ENARQUIFES, 2019.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.

HENTTONEN, P. Privacy as an archival problem and a solution. **Archival Science**, v. 17, n. 3, p. 285–303, 2017. ISSN: 13890166, DOI: <http://dx.doi.org/10.1007/s10502-017-9277-0>.

KATUU, S. Managing digital records in a global environment. **The Electronic Library**, [S.l.], v. 34, n. 5, p. 869–894, 2016. ISSN: 02640473. DOI: 10.1108/EL-04-2015-0064.

RIES, T. The rationale of the born-digital dossier genetique: Digital forensics and the writing process: With examples from the Thomas Kling Archive. **DIGITAL**

SCHOLARSHIP IN THE HUMANITIES, ENGLAND, v. 33, n. 2, p. 391–424, 2018. Disponível em: <https://doi.org/10.1093/llc/fqx049>. Acesso em: 07 set. 2020.

ROGERS, C. Diplomatics of born digital documents - considering documentary form in a digital environment. **Records Management Journal**, [S.l.], v. 25, nº 1, p. 6–20, 2015. ISSN: 09565698, DOI: 10.1108/RMJ-03-2014-0021. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/RMJ-03-2014-0021/full/pdf>. Acesso em: 15 out. 2023

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v.15, n.30, p. 1–31, 2010. DOI: <https://doi.org/10.5007/1518-2924.2010v15n30p1>.

SILVA, P. F. C. da. **Diagnóstico documental do Centro de Ciências Jurídicas da Universidade Federal da Paraíba**. João Pessoa, 2018. Disponível em: <https://bit.ly/3LmQq7h>. Acesso em: 12 abr. 2020.