

A gestão da preservação digital de dados de pesquisa proposta de um modelo processual

José Carlos Abbud Grácio¹
Miguel Ángel Márdero Arellano²

Resumo

O objetivo deste trabalho é apresentar os elementos relacionados à preservação digital para dados de pesquisa e um modelo de gestão a partir da conceptualização atual do seu ciclo de vida, onde as instituições de financiamento a pesquisas demandam um plano de gestão de dados dos pesquisadores que garanta que os dados das pesquisas financiados por essas instituições possam ser validados e reutilizados por outros pesquisadores. A usabilidade dos dados arquivados digitalmente nas instituições é apontada como o resultado da definição de uma política de preservação adotada pelas iniciativas de acesso aberto, de como o conteúdo arquivado deve ser descoberto e usado pelos pesquisadores, dentro de um modelo de gestão que justifica sua coleta e preservação, provendo novos significados para as partes interessadas ao longo do tempo. A identificação dos principais elementos que envolvem a gestão da preservação digital de dados de pesquisa focaliza o seu papel imperativo na gestão das necessidades imediatas das universidades.

Palavras chave: Preservação digital. Dados de pesquisa. Curadoria digital. Informação digital.

¹ Universidade Estadual Paulista

 <https://orcid.org/0000-0001-7620-1309> – E-mail: gracio@marilia.unesp.br
Marília – SP / Brasil

² Instituto Brasileiro de Informação em Ciência e Tecnologia
Brasília – DF / Brasil

 <https://orcid.org/0000-0001-5306-919X> – E-mail: miguel@ibict.br

Submetido em: 05 nov. 2019

Aceito em: 25 jan. 2020

Publicado em: 17 fev. 2020



The management of digital preservation of research data proposal for a procedural model

Abstract

The objective of this paper is to present the elements related to digital preservation for research data and a management model from the current conceptualization of its life cycle, where research funding institutions demand a research data management plan that guarantees research data funded by these institutions can be validated and reused by other researchers. The usability of digitally archived data in institutions is pointed to as the result of the definition of a preservation policy adopted by open access initiatives, of how archived content should be discovered and used by researchers, within a management model that justifies its collection and preservation, providing new meanings to stakeholders over time. Identifying the key elements involved in managing digital preservation of research data focuses on their imperative role in managing the immediate needs of universities.

Keywords: Digital preservation. Research data. Digital curation. Digital information.

1 Introdução

No contexto atual, a utilização dos dados em formato digital se faz cada vez mais presentes nas instituições e na vida das pessoas. Isso acontece também com os dados de pesquisa que, em sua maior parte, são coletados e armazenados em formato digital, o que possibilita uma análise mais rápida e eficiente.

Se a informação pode ser definida como um conjunto de dados com significado que tem como objetivo gerar novos conhecimentos nas pessoas, a partir de conhecimento prévio, a informação digital produzida a partir dos dados de pesquisa também tem esse objetivo e possuem desafios e características específicas.

Os principais desafios relacionados à informação digital produzida a partir dos dados de pesquisa são:

- A explosão da quantidade de informação digital produzida a partir das pesquisas realizadas;
- A obsolescência do hardware e do software que tratam essa informação;
- As mudanças no formato dos arquivos onde a informação é registrada;
- As mudanças nas mídias de armazenamento;
- As mudanças e avanços muito rápidos nas TICs;
- A busca e recuperação dessa informação digital;
- Os custos elevados das novas tecnologias para gerenciar essa informação;
- As necessidades de recursos humanos e tecnológicos para manter o acesso aos dados de pesquisa e aos seus resultados;
- A capacidade de assimilação das mudanças tecnológicas e de procedimentos dentro da organização, ou seja, a mudança na cultura organizacional.

Nas instituições de ensino superior (IES), a informação digital está em todas as áreas, tais como:

- Ensino: material didático
- Pesquisa: livros, revistas eletrônicas e dados de pesquisa
- Extensão: cursos de extensão e eventos
- Administração: documentos arquivísticos
- Comunicação: e-mail, Videoconferência, produções de Radio e TV, Notícias e fotografias

Todo esse acervo digital que faz parte das IES deve ser preservado com o propósito de registrar sua história e atender as legislações vigentes, como a Lei de acesso à informação. No caso dos dados de pesquisas, os mesmos são de grande importância para a continuidade das pesquisas, para validá-los e para que outros pesquisadores possam utilizar essas informações para novas pesquisas.

A National Science Board, em 2005, define os dados de pesquisa como: “...qualquer informação que possa ser armazenada em formato digital, incluindo texto, números, imagens, vídeos ou filmes, áudio, software, algoritmos, equações, animações, modelos, simulações, etc”.

Quando os dados não são arquivados de forma rápida e não há um planejamento inicial de guarda e proteção tornam-se um desafio sua preservação e acesso. Nos laboratórios e coordenações de projetos existem dados de centenas de estudos guardados em disquetes, fitas VHS, fitas cassetes, CDs e DVDs. O fato de alguns desses dados já estarem na Web não garante sua continuidade e integridade, pois é inegável que as URLs deixam de existir por vários motivos.

Existe também o caso onde os dados de pesquisa estão em formatos proprietários: SPSS, Excel, SAS, Word, Stata, ASCII, Access, Matlab, VHS, File Marker Pro, JPGs, etc., e em mídias de armazenamento usadas comumente pelos pesquisadores como os hard drives pessoais. O problema não é só perder os arquivos, mas perder aqueles dos quais os pesquisadores são institucionalmente responsáveis.

Para que esses dados não se percam, é necessário que os dados de pesquisa sejam preservados ao longo do tempo garantindo a autenticidade, integridade e acesso. As instituições financiadoras de pesquisas no Brasil também têm exigido que os pesquisadores e as instituições garantam, através de um plano de gestão de dados que deve ser apresentado, que os dados de pesquisa possam ser preservados e acessados por outros pesquisadores.

Dessa forma, as instituições necessitam desenvolver, dentro de suas políticas, uma política de preservação digital para dados de pesquisa que assegure ao pesquisador privacidade, segurança e a preservação ao longo do tempo para as informações produzidas em suas pesquisas e atenda às exigências das instituições financiadoras.

O objetivo deste trabalho é apresentar os elementos relacionados à preservação digital para dados de pesquisa, a partir do contexto atual onde as instituições de financiamento a pesquisas vêm exigindo um plano de gestão de dados dos pesquisadores que garanta que os dados das pesquisas financiados por essas instituições possam ser validados e reutilizados por outros pesquisadores.

2 O ciclo de vida dos dados de pesquisa

A disponibilidade dos dados de pesquisa permitem validar os resultados da pesquisa, analisar esses dados em outros contextos, comparar com dados atuais e permitir novas pesquisas a partir desses dados.

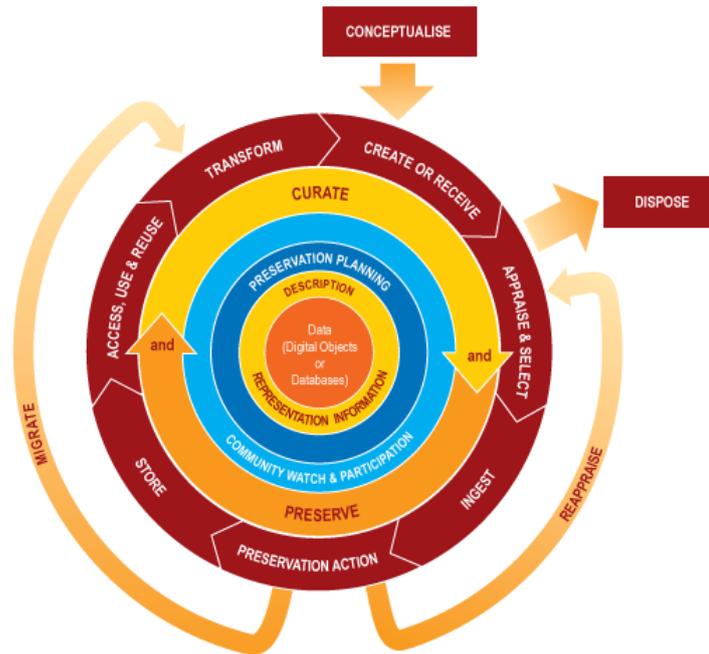
Ao agregar novos dados aos já existentes e novas pesquisas, esses dados passam a ter um “valor” maior e dessa forma agregar novas informações e gerar novos conhecimento.

Os dados de pesquisa precisam ser organizados e estruturados com o objetivo de permitir seu armazenamento nos formatos adequados, preservação e acesso, garantindo sua confiabilidade, autenticidade e integridade.

A Data Curation Center (DCC), um órgão britânico de curadoria e preservação de dados de pesquisa, define que a curadoria digital “envolve a manutenção, preservação e a adição de valor para os dados de pesquisa através de seu ciclo de vida” e afirma que “a curadoria digital e a preservação de dados são processos contínuos, exigindo uma reflexão considerável e o investimento de tempo e recursos adequados.” (DIGITAL, 2019)

O modelo do ciclo de vida da curadoria digital proposto pela DCC apresenta uma visão geral, desde a criação dos dados de pesquisa até sua utilização e reuso, contemplando as atividades de preservação digital. O modelo é apresentado abaixo:

Figura 1. Modelo do ciclo de vida da curadoria



Fonte: DCC, 2019.

O pesquisador necessita de um plano de gestão de dados que esteja alinhado com a política de preservação digital de dados de pesquisa da instituição. Do lado do pesquisador, o mesmo deve planejar a captura, a organização e a forma de armazenamento, enquanto que a instituição deve prover uma infraestrutura de repositórios para armazenar, indexar e prover as ferramentas de busca e recuperação desses dados. Dessa forma a preservação digital dos dados de pesquisa está inserida na curadoria digital.

Para a Comissão Europeia "um plano de gestão de dados é um documento que descreve como os dados de pesquisa recompilados ou gerado serão tratados durante um projeto de pesquisa e depois de sua conclusão, descreve quais serão mantidos seguindo metodologias e padrões específicos, e como esses dados serão compartilhados, se serão abertos e como ocorrerá a curadoria digital" (Silva, p.46).

A elaboração de um plano de gestão de dados de pesquisa exige uma mudança de cultura do pesquisador e da instituição na qual ele está inserido. Para Silva (2016 p. 47):

É evidente que muito poucos pesquisadores se preocupam com os registros de dados de pesquisa. Eles geralmente os mantêm até não precisarem mais deles, mas graças aos planos de gerenciamento de dados de pesquisa, eles podem garantir que seus objetivos sejam compatíveis com a preservação e o acesso, enquanto que com os dados da pesquisa são uma produção cada tempo mais importante, mas cara no processo de pesquisa acadêmica, em todas as disciplinas.

Os responsáveis pela gestão dos dados de pesquisa precisam comunicar à comunidade científica sobre as ações que devem ser tomadas desde o início do ciclo

de vida dos dados. A participação deles na avaliação dos materiais se podem ou não ser publicados, se são úteis e se as expensas na sua preservação a longo prazo são justificáveis.

Para a FAPESP, um Plano de Gestão de Dados deve responder a duas perguntas básicas:

1. Quais dados serão gerados pelo projeto?
2. Como serão preservados e disponibilizados, considerando questões éticas, legais, de confidencialidade e outras?

O plano submetido à FAPESP deve seguir as seguintes diretrizes (FUNDAÇÃO, 2019):

- Descrição dos dados e metadados produzidos pelo projeto;
- Quando aplicável, restrições legais ou éticas para compartilhamento de tais dados, políticas para garantir a privacidade, confidencialidade, segurança, propriedade intelectual e outros;
- Política de preservação e compartilhamento;
- Período de carência e período durante o qual os dados serão preservados e disponibilizados;
- Descrição de mecanismos, formatos e padrões.

Nesse contexto, torna-se necessário, além de um plano de ação, que as instituições tenham uma Política de Preservação Digital para dados de pesquisa dentro da política de gestão de dados do repositório, já que esta não garante a disponibilidade de uma infraestrutura mínima para o repositório de pesquisa, propondo também que sejam aplicadas medidas de segurança e de proteção da privacidade. Dessa forma a política objetiva, além de atender as agências de financiamento, preservar e dar acesso aos dados de seus pesquisadores.

3 Aspectos da gestão da preservação digital de dados de pesquisa

A preservação digital pode ser definida como “os processos de gestão envolvidos na administração das atividades necessárias para garantir que um objeto digital possa ser acessado e utilizado no futuro, a partir das TIC existentes na época e com garantias de sua autenticidade e integridade”.

Nesse sentido, ela deve fazer parte de todo o ciclo de vida de dados de pesquisa, desde o planejamento até o seu acesso e reuso.

Uma Política de Preservação Digital de Dados de Pesquisa e os processos decorrentes para sua implantação devem englobar os elementos e os aspectos da Preservação Digital, os avanços nas Tecnologias de Informação e Comunicação (TIC) e as mudanças que afetam a Cultura Organizacional.

A gestão da preservação digital de dados de pesquisa refere-se ao armazenamento, acesso e preservação de dados produzidos a partir de uma determinada pesquisa. Toda organização de pesquisa precisa ter uma política de dados de pesquisa, que inclua um esquema de gestão de dados.

O elemento da preservação digital na política de gestão envolve as condições de continuidade do conteúdo, a regularidade de realização de cópias de segurança, retenção de bitstreams, remoção de itens, autorizações de atualizações, atribuição de checksum, arquivamento e transferências de bancos de dados.

As primeiras iniciativas que adotaram uma política de preservação de dados provieram das financiadoras das pesquisas que estipularam as obrigações que um pesquisador deve cumprir como condição do financiamento.

Um plano de gerenciamento de dados deve considerar os muitos aspectos do gerenciamento de dados, entre eles o da sua preservação desde antes do início do projeto, garantindo que os dados sejam bem gerenciados no presente e preparados para preservação no futuro.

Um dos aspectos que envolvem a preservação dos dados de pesquisa é incluir na política de gestão que existam mecanismos para garantir que quando o pesquisador sair da universidade por aposentadoria ou mudança, os dados com valor de uso a longo prazo estejam depositados num repositório de dados institucional para seu arquivamento.

As práticas de gerenciamento de dados cobrem todo o ciclo de vida dos dados, desde o planejamento da investigação até a realização, bem como de fazer backup de dados, uma vez que eles são utilizados para a preservação de dados a longo prazo entregues após a conclusão da investigação.

A simulação de dados em grande escala movimenta a ciência computacional atual, reunindo um conjunto de instruções e de dados que determinarão parte dos resultados.

A integridade na pesquisa é garantida pela quantidade de informação necessária para o conhecimento pessoal. Um dos componentes dessa garantia de integridade são os metadados de preservação, os quais provêm informação para dar suporte ao processo permanência dos registros.

Entre os primeiros serviços de preservação de dados estão:

- ICPSR da University of Michigan em funcionamento desde 1960 e que é considerado o maior arquivo de dados das ciências sociais do mundo.
- Na University of North Carolina o Odum Institute mantém o arquivo mais antigo de dados das ciências sociais fundado em 1920.
- The National Archives and Records Administration lidera as iniciativas de preservação de acervos de dados especializados governamentais.
- Na Universidade de Harvard existe a primeira aliança nacional de preservação de dados de pesquisa, Data-PASS, especializada no desenvolvimento de soluções e políticas de gestão de dados científicos (ALTMAN *et al.* 2009).

Na América Latina existem instituições emergentes que requerem dos seus pesquisadores que seus dados sejam publicados em repositórios abertos, caso esses projetos recebam algum tipo de financiamento.

No Brasil a agência de financiamento do estado de São Paulo FAPESP pede aos pesquisadores que solicitam auxílio financeiro pelo fundo do programa E-Science que apresentem seus planos de gestão de dados de pesquisa.

Mas ainda falta em iniciativas como essa a integração do processo de preservação em todo o ciclo de vida dos dados. Qualquer custo de armazenamento e gestão deve ser explicitado nos pedidos de financiamento, incluindo o local onde os dados serão disponibilizados e aplicadas as técnicas de curadoria, por anos após do término do projeto.

O Ibict é uma instituição de pesquisa da América Latina que realiza esforços para contribuir com o compartilhamento de dados e seu reuso. Essas atividades se iniciaram no Instituto com a implementação da Biblioteca Digital de Teses e Dissertações, a segunda maior biblioteca digital de seu tipo, e com a promoção da plataforma *Open Journal Systems* (OJS), sendo atualmente o Brasil o país com maior número de instalações dessa plataforma.

Os elementos relacionados à preservação digital de dados de pesquisa e à elaboração de uma política, podem ser divididos em três grupos: organizacionais, legais e técnicos. Apesar de terem suas próprias especificidades e seus elementos, esses grupos estão relacionados, da mesma forma que os elementos que os compõem.

3.1 Aspectos Organizacionais

Os aspectos organizacionais dizem respeito à gestão da instituição e são relacionados aos objetivos, ao envolvimento da comunidade e de pessoas especializadas nas atividades de preservação digital dos dados de pesquisa, além de tratar dos problemas administrativos e financeiros.

3.1.1 Objetivos da instituição

A implantação de uma política de preservação digital de dados de pesquisa depende da vontade e da disponibilidade das IES. Envolve, além de investimentos em infraestrutura e pessoal, uma mudança de cultura da instituição e, conseqüentemente, das pessoas, principalmente dos pesquisadores. Portanto, a preservação digital de dados de pesquisa deve estar inserida nos objetivos da instituição e conseqüentemente nas atividades de pesquisa dos pesquisadores ligados a ela.

Essa inserção da preservação digital nos objetivos da instituição deve começar pela elaboração da política, que além de consolidar o compromisso de preservar os dados de pesquisa dos pesquisadores da IES, também fortalece a continuidade das atividades relacionadas à preservação digital, independente das mudanças de gestores que ocorrem nas instituições.

3.1.2 Equipe multidisciplinar

Para realizar a gestão dos processos envolvidos na preservação digital dos dados de pesquisa, é necessária uma equipe multidisciplinar com profissionais que possuam as competências para cada uma das atividades envolvidas. Podemos citar Bibliotecários, Arquivistas, profissionais da Ciência da Informação, Informática, Sistemas de Informação, Direito, Administração da instituição, Docentes, Gestores acadêmicos e administrativos das IES e outros que a instituição julgar necessários.

A instituição deverá considerar a necessidade de: normas e padrões a serem adotados; analisar as questões técnicas relativas às TIC, decorrente das constantes mudanças e avanços; prever investimentos permanentes nas atividades de preservação digital; lidar com as questões legais, como as dos direitos autorais dos pesquisadores; acompanhar o aumento das informações digitais de pesquisas e a necessidade da preservação das mesmas; avaliar as questões sociais e culturais.

3.1.3 Responsabilidades

Nas IES, a preservação digital dos dados de pesquisa é responsabilidade do pesquisador (criador do objeto digital), da própria instituição e de todas as pessoas envolvidas nos processos.

A preservação dos dados de pesquisa começa com o pesquisador quando esses dados são coletados e armazenados. É necessário o armazenamento no formato e no local correto, além da descrição por metadados desses dados.

A Instituição Financiadora deve definir claramente o que espera do pesquisador e das IES sobre os processos de preservação digital.

Nas IES, além dos pesquisadores, a preservação digital é uma responsabilidade compartilhada pela equipe multidisciplinar, definida na política e nas normas.

As instâncias superiores da IES devem responsabilizar-se pela: definição dos objetivos e diretrizes para a preservação digital; pela montagem da equipe multidisciplinar que irá administrar os processos; planejamento orçamentário; garantia da continuidade dos programas de preservação digital.

A equipe multidisciplinar deve responder pela definição das políticas de preservação digital dos dados de pesquisa, pelo acompanhamento dos processos envolvidos e por definir os critérios de seleção e descarte.

As instâncias jurídicas da IES devem resguardar a instituição e os pesquisadores, nos aspectos legais, seguindo as normas e leis existentes, orientando as questões dos direitos autorais e servindo de assessoria para a equipe multidisciplinar.

O pessoal de TI deve montar e dar suporte a toda infraestrutura tecnológica necessária, além dos procedimentos de segurança para acesso e preservação dos dados de pesquisa, tais como: rotinas de cópia de segurança (*backup*), aplicação periódica das estratégias de preservação digital, criação de sistemas de informação, acesso *on-line* aos dados de pesquisa, entre outras.

Os bibliotecários e arquivistas definirão os metadados necessários e outras atividades relacionadas às suas atribuições e realizarão as atividades relacionadas à gestão documental.

Todos deverão trabalhar em conjunto para, além de preservar os dados de pesquisa, produzirem e disponibilizarem os produtos e serviços mais adequados para busca, recuperação e acesso aos dados de pesquisa.

3.1.4 Recursos financeiros / Investimento

10

Como a preservação digital envolve investimentos permanentes em tecnologia, infraestrutura e capacitação de pessoal, é necessária para a preservação digital de dados de pesquisa uma política permanente de investimento por parte da instituição, onde os custos da preservação façam parte do orçamento anual.

O investimento necessário dependerá de quais e quantos dados de pesquisa serão preservados, por quanto tempo e como será disponibilizado o acesso a eles.

3.1.5 Atos administrativos

Após a definição da política de preservação digital de dados de pesquisa, a IES deve, através de atos administrativos (normas, portarias e outros), implantar a política e definir uma estrutura que tenha como objetivo gerir os processos envolvidos, definindo os órgãos executores e as pessoas responsáveis.

Esses atos jurídicos contribuem para a inserção da cultura de preservação digital de dados de pesquisa na cultura organizacional e auxiliam os pesquisadores na elaboração dos planos de gestão de dados de pesquisa, a partir do momento que

a instituição define e documenta qual a infraestrutura e os serviços e produtos que ela oferece para a preservação e acesso aos dados do pesquisador.

3.2 Aspectos Legais

Os aspectos legais estão relacionados às questões legais, ao respeito aos direitos autorais, às leis existentes em nível internacional e nacional e a atos administrativos da IES, buscando garantir a legalidade dos processos de preservação digital para a instituição e para o pesquisador.

3.2.1 Direitos Autorais

Preservar implica em copiar para garantir o acesso ao longo do tempo na plataforma adequada. Dessa forma a IES deve garantir que a autenticidade e a integridade dos dados de pesquisa e o direito de autor do pesquisador responsável por esses dados.

Nesse sentido, a política de preservação digital dos dados de pesquisa deve estar amparada em leis que garantam a propriedade intelectual ao pesquisador e sua autenticidade, resguardando a instituição de problemas jurídicos e atendendo os requisitos das instituições financiadoras.

3.3 Aspectos Técnicos

Os aspectos técnicos são compostos da seleção do que deve ser preservado, da utilização de modelos, padrões e iniciativas, da montagem da infraestrutura tecnológica adequada, da utilização de repositórios digitais, da aplicação de estratégias de preservação, do uso do suporte adequado e da representação por metadados.

3.3.1 Seleção e descarte

A seleção implica em definir o que deve ser preservado, ou seja, selecionar as informações digitais a serem preservadas de acordo com a missão e os objetivos da instituição, as necessidades do pesquisador, da comunidade interna e externa e de critérios claramente definidos.

Dessa forma, os pesquisadores têm a definição de quais dados de pesquisa a instituição se propõe a preservar, quais os formatos e quais metadados serão utilizados.

O descarte tem como objetivo definir por quanto tempo os dados de pesquisa devem ser preservados e sua forma de descarte. O tempo que um dado de pesquisa precisa ser preservado deve estar de acordo com as questões legais e com sua frequência de uso. Na maioria dos casos, os dados de pesquisa deverão ser de guarda permanente.

3.3.2 Padrões e modelos

O uso de padrões de preservação digital para os dados de pesquisa, como por exemplo padrões de metadados, auxiliam na organização dos dados, na interoperabilidade, na troca de experiências e de conhecimento e permitem que as instituições sejam auditadas e certificadas.

O modelo de preservação digital recomendado atualmente é o Open Archival Information Systems (OAIS), desenvolvido pelo Consultative Committee for Space Data Systems (CCSDS), definido através da norma ISO de n. 14.721:2012, que descreve o enquadramento conceitual para um repositório digital genérico, aberto a todas as comunidades e com garantias de confiabilidade.

3.3.3 Metadados de preservação digital

Metadados de preservação digital são informações estruturadas que documentam as estratégias de preservação digital aplicadas a um objeto digital (dados de pesquisa) ao longo de sua vida, com o objetivo de assegurar o acesso a longo prazo.

Dessa forma, para auxiliar na garantia de autenticidade dos dados de pesquisa preservados, é necessário documentar nos metadados de preservação todas as estratégias aplicadas aos dados de pesquisa durante sua vida.

3.3.4 Suporte

Historicamente, alguns pesquisadores guardavam seus dados de pesquisa em suporte próprios. Com as mudanças e evoluções dos suportes e seu tempo de vida, é necessário avaliar a deterioração e a obsolescência dos mesmos, realizando as mudanças de suporte quando necessário.

3.3.5 Autenticidade, Integridade e Proveniência

A Autenticidade é a capacidade de identificar elementos que permitam definir se o dado de pesquisa é autêntico ou não, tais como, ferramentas para validar a integridade das sequências de bits e informações de proveniência.

A Integridade é a capacidade de garantir que o dado de pesquisa não foi alterado ao longo de sua existência e cobre abordagens como a encriptação, as assinaturas digitais e a verificação de fixidez (*checksum*).

A Proveniência é a documentação do processo do ciclo de vida dos dados de pesquisa, descritas nos metadados.

Esses são elementos que devem ser definidos em uma política de preservação de dados de pesquisa, pois são fundamentais para garantir sua perenidade.

3.3.6 Infraestrutura tecnológica

A infraestrutura tecnológica garante o arquivamento, a preservação, a integridade, a segurança e o acesso aos dados de pesquisa. As principais necessidades são:

- Infraestrutura de *hardware* e *software*;
- Links de acesso de alta velocidade para acesso aos dados de pesquisa;
- Infraestruturas de preservação e de acesso, tais como Archivematica e Dataverse;
- Medidas de integridade, verificação regular dos dados de pesquisa, procedimentos de recuperação de desastres e de segurança;
- Mais de um backup e em lugares de armazenamento diferentes;
- Programa permanente de capacitação para o pessoal técnico.

3.3.7 Repositório Digital

Um repositório digital pode ser definido como “um complexo que apoia o gerenciamento dos materiais digitais, pelo tempo que for necessário, e é formado por elementos de hardware, software e metadados, bem como por uma infraestrutura organizacional e procedimentos normativos e técnicos.” (CONARQ, 2015). No caso de dados de pesquisa uma das ferramentas utilizadas é o Dataverse.

3.3.8 Estratégias de preservação digital para dados de pesquisa

É necessário planejar e avaliar periodicamente, para cada tipo de dados de pesquisa, as tecnologias existentes no momento e determinar qual a melhor estratégia a ser implementada. As principais estratégias de preservação são:

- Migração: transforma os dados de pesquisa de um formato obsoleto para um formato atual;
- Atualização de suporte: transfere um dado de pesquisa de um suporte para outro;
- Emulação: preserva o material no formato de arquivo original e desenvolve ferramentas de software que simulam o software original necessário para acessar o material digital;
- Arquivamento de software / hardware: a organização preserva o hardware e software originais para poder acessar o material digital no seu ambiente original.

4 Um modelo processual para gestão da preservação digital de dados de pesquisa

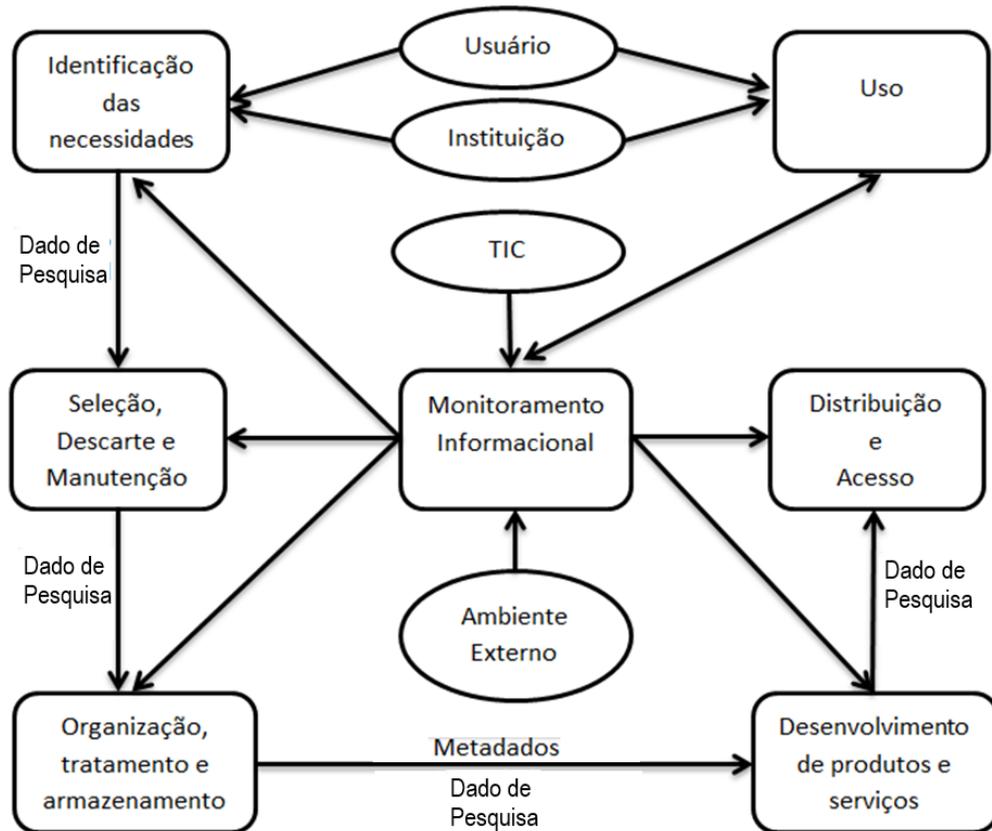
A preservação digital de dados de pesquisa muda a cultura organizacional a partir do momento que implementa novos processos e influencia as atividades, tanto dos docentes, quanto da instituição. Dessa forma demanda vontade institucional, tempo de implantação, planejamento e o envolvimento dos pesquisadores e dos grupos envolvidos.

Os processos possuem um fluxo de atividades que devem integrados e contemplar os elementos da preservação digital de dados de pesquisa. Devem ser definidos os responsáveis pelas atividades relacionadas de cada um dos processos.

Para demonstrar os processos apresentamos um modelo denominado Modelo Processual de Preservação Digital para Dados de Pesquisa, com processos cíclicos e correlatos.

O modelo deve ser dinâmico e se adequar, quando necessário, às mudanças causadas pelos avanços das TICs e pelas mudanças ocorridas na IES.

Figura 3. Modelo processual de preservação digital para dados de pesquisa



Fonte: GRÁCIO, 2012.

O modelo é composto de processos, entidades e fluxos. Os processos, representados por retângulos com os vértices arredondados, contemplam as atividades e os elementos da preservação digital. As entidades, representadas por uma elipse, são pessoas (docentes, funcionários ou usuários externos), instituições (IES, gerência etc.) ou situações (mudança tecnológica, crise financeira, cultura etc.) que interagem com os processos. Os fluxos são simbolizados por uma linha e uma seta indicando a direção do fluxo, e representam as conexões existentes entre um processo e outro, e entre uma entidade e um processo.

As TICs e a Cultura Organizacional interagem com todo o modelo, pois a instituição deve estar atenta às mudanças e às inovações que ocorrem nas tecnologias e na instituição.

A seguir, detalhamos cada processo envolvido no modelo.

4.1 Identificação das necessidades

O processo de identificação das necessidades envolve a definição das políticas e normas, a montagem das equipes de trabalho, a definição de quais dados de pesquisa devem ser preservados.

4.2 Seleção, descarte e manutenção

A partir do processo de identificação das necessidades de preservação, o processo de seleção, descarte e manutenção dos dados de pesquisa deve selecionar o que deve ser preservado, o que deve ser descartado e como fazê-los, a partir das políticas e normas definidas.

O processo se inicia quando o pesquisador transfere seus dados de pesquisa e os metadados associados para custódia da IES. Baseado nas políticas, nas normas e nos critérios de seleção, a solicitação é analisada. Se aprovada, o pesquisador deve encaminhar à IES uma autorização em que consta que, a partir desse momento, seus dados de pesquisa passam a fazer parte da custódia da instituição e que esta se torna responsável por sua preservação, garantindo sua autenticidade e aplicando as estratégias quando necessário.

A definição dos dados de pesquisa a serem descartados passa pela definição de uma política que deve envolver a instituição, o pesquisador e as instituições financiadoras, além da análise do uso por parte da comunidade.

4.3 Organização, tratamento e armazenamento

A organização envolve a montagem e administração da infraestrutura técnica e de pessoal, que irá atender às necessidades físicas e lógicas relacionadas ao hardware e ao software. A infraestrutura deverá atender também as necessidades de armazenamento, busca e recuperação dos dados de pesquisa, empreendendo ações tais como: capacitação e aperfeiçoamento para o pessoal técnico envolvido no processo; atualização da infraestrutura de acesso e armazenamento, do software e da rede de computadores.

O tratamento dos dados de pesquisa envolve atividades em 3 situações: na submissão e inserção na infraestrutura de preservação da IES; no descarte; na aplicação das estratégias de preservação.

Na submissão e inserção, o tratamento envolve complementar os metadados encaminhados pelo pesquisador, adequar o formato dos dados de pesquisa se necessário e armazenar esses dados na infraestrutura de preservação e de acesso.

No descarte, o tratamento envolve, após detectar a necessidade, definir o procedimento a ser adotado e a descrição por metadados.

Para os dados de pesquisa já preservados e armazenados, será necessária, ao longo de seu ciclo de vida, a aplicação de estratégias de preservação digital que os mantenha acessível. As atividades envolvem:

- Verificar a necessidade da aplicação de estratégias;
- A partir da constatação, identificar as tecnologias a serem utilizadas;

- Determinar e aplicar as estratégias e os suportes a serem utilizados;
- Descrever os metadados de preservação;
- E armazenar os dados no novo formato e seus metadados.

4.4 Desenvolvimento de produtos e serviços

A partir do momento em que os dados de pesquisa a serem preservados estão organizados, armazenados e descritos é necessário criar e disponibilizar ferramentas que recuperem esses dados da infraestrutura de acesso e atendam às necessidades de busca e recuperação dos pesquisadores. Os produtos e os serviços devem atender e se adequar, quando necessário, as necessidades dos usuários e as mudanças e avanços nas TICs existentes.

4.5 Distribuição e acesso

O processo de acesso verifica os privilégios de acessos dos pesquisadores interessados nos dados de pesquisa, define os formatos de distribuição e gera o pacote de disseminação da informação que será encaminhado.

O processo de distribuição define a forma que os dados de pesquisa serão encaminhados ao pesquisador, utilizando canais de comunicação eficientes, que atendam as mais variadas formas de acesso, formato e suporte.

4.6 Uso

O processo de uso avalia se os dados de pesquisas e os produtos e serviços relacionados a eles estão atendendo adequadamente os pesquisadores. Como o objetivo principal da preservação digital é o acesso contínuo aos objetos digitais, o processo de Uso da informação é tratado como um processo separado devido à sua importância, mesmo estando fortemente relacionado ao monitoramento informacional.

A avaliação pode ser realizada utilizando ferramentas automáticas, como contadores de acesso, e formulários de avaliação encaminhados para os pesquisadores que utilizam os dados preservados.

4.7 Monitoramento informacional

O processo monitoramento informacional realiza o acompanhamento e a adequação constante dos processos que abrangem o modelo às necessidades dos pesquisadores.

Esse processo pode ser realizado por meio do levantamento das necessidades das entidades envolvidas no modelo e da observação, análise e avaliação permanente dos elementos relacionados à preservação digital, tendo como principal resultado atos administrativos que registram e encaminham as mudanças necessárias.

5 Considerações

A exigência de um plano de gestão, que envolve também a preservação de dados de pesquisa, por parte de algumas instituições de financiamento, cria a necessidade das instituições planejarem uma infraestrutura de preservação digital dos dados de pesquisa que observe essa exigência. A preservação digital de dados de pesquisa irá trazer benefícios para o pesquisador e para a instituição, pois esses dados poderão ser validados e reutilizados pela comunidade acadêmica.

Entretanto, para atender essa demanda de preservação digital, as instituições devem elaborar uma política de preservação digital de dados de pesquisa, que permita implementar os processos envolvidos nas atividades de preservação digital.

O modelo proposto define o fluxo dos processos e os elementos de preservação digital que devem ser contemplados na implementação da política.

Observa-se que a preservação digital de dados de pesquisa deve tratar dos aspectos administrativos, técnicos e legais, onde os principais desafios são a inserção da preservação digital nos objetivos da instituição, a mudança na cultura organizacional, tanto dos pesquisadores como das instituições e a montagem da infraestrutura de preservação digital e de acesso aos dados de pesquisa. A participação dos docentes é fundamental nesse processo.

A instituição deve, através da gestão dos processos, garantir a autenticidade, integridade, confiabilidade e o acesso aos dados de pesquisa preservados, o que dá credibilidade para a instituição. Estabelecer parcerias com outras instituições auxiliam na utilização das melhores ferramentas e também nos melhores fluxos de trabalho.

Referências

ALTMAN, Micah *et al.* Digital preservation through archival collaboration: the data preservation alliance for the Social Sciences. **The American Archivist**, v. 72. doi: 10.17723/aarc.72.1.eu7252lhnrp7h188.

CONARQ. Conselho Nacional de Arquivos. **Diretrizes para a implementação de repositórios arquivísticos digitais confiáveis** – RDC-Arq. Rio de Janeiro, 2015. Disponível em: http://www.conarq.gov.br/images/publicacoes_textos/diretrizes_rdc_arq.pdf. Acesso em: 04 jun. 2019.

CCSDS. Consultative Committee dor Space Data Systems. **Reference model for an Open Archival Information System (OAIS)**. Washington, 2012. 135 p. Disponível em: <https://public.ccsds.org/pubs/650x0m2.pdf>. Acesso em: 29 jul. 2019.

DIGITAL Curation Centre. **DCC Curation Lifecycle Model**. Disponível en: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>. Aceso em: 04 fev. 2019.

FUNDAÇÃO de Amparo à Pesquisa do Estado de São Paulo. **Plano de Gestão de Dados - FAPESP**. Disponível em: <http://www.fapesp.br/gestaodedados>. Acesso em: 04 fev. 2019.

GRÁCIO, J. C. A. **Preservação digital na gestão da informação: um modelo processual para as instituições de ensino superior**. São Paulo: Cultura Acadêmica, 2012.

INGRAM, Caroline. **How and why you should manage your research data: a guide for researchers**. JISC 2019. Disponível em: <https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data>. Acesso em: 28 jul 219.

NATIONAL SCIENCE BOARD. **Long-lived digital data collections: enabling research and education in the 21st Century**. 2005. Disponível em: <https://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf>. Acesso em:30 jul. 2019.

SAYÃO, Luis Fernando; SALES, Luana Farias. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, v.22, n.3, p. 179-191, set./dez. 2012. Disponível em: <http://www.periodicos.ufpb.br/ojs2/index.php/ies/article/download/12224/8586>. Acesso em: 23 fev. 2013.

SILVA, Fabiano Couto Corrêa da. Infraestructuras y políticas internacionales de desarrollo para gestión de los datos de investigación. **Biblios** [online] 2016, (Sin mes): Disponível em: <http://www.redalyc.org/articulo.oa?id=16146347004>. E-ISSN: 1562-4730. doi: <https://doi.org/10.5195/biblios.2016.286>. Acesso em: 04 fev. 2019.