

Dicionário de dados: um olhar de qualidade sobre os dados de recursos humanos da Universidade Federal da Integração Latino-Americana

Data dictionary: a quality look at human resources data from the Federal University of Latin American Integration

Claudio Roberto Marquette Mauricio¹, Eduardo César Dechechi², Willian Francisco da Silva³, Andréia da Cruz⁴

¹ Universidade Estadual do Oeste do Paraná (Unioeste), Foz do Iguaçu, Paraná, Brasil. ORCID: <http://orcid.org/0000-0003-2783-3988>

² Universidade Estadual do Oeste do Paraná (Unioeste), Foz do Iguaçu, Paraná, Brasil. ORCID: <http://orcid.org/0000-0001-6563-5435>

³ Universidade Estadual do Oeste do Paraná (Unioeste), Foz do Iguaçu, Paraná, Brasil. ORCID: <http://orcid.org/0000-0001-9181-5712>

⁴ Universidade Estadual do Oeste do Paraná (Unioeste), Foz do Iguaçu, Paraná, Brasil. ORCID: <http://orcid.org/0000-0002-5253-8676>

Autor para correspondência/Mail to: Cláudio Roberto Marquette Mauricio, crmmauricio@gmail.com

Recebido/Submitted: 20 de março de 2022; **Aceito/Approved:** 19 de outubro de 2023



Copyright © 2024 Mauricio, Dechechi, Silva, Cruz. Todo o conteúdo da Revista (incluindo-se instruções, política editorial e modelos) está sob uma licença Creative Commons Atribuição 4.0 Internacional. Ao serem publicados por esta Revista, os artigos são de livre uso para compartilhar e adaptar e é preciso dar o crédito apropriado, prover um link para a licença e indicar se mudanças foram feitas. Mais informações em <http://revistas.ufrpr.br/atoz/about/submissions#copyrightNotice>.

Resumo

Introdução: Este artigo é um estudo sobre a dicionarização dos dados pessoais de docentes de uma universidade federal, e tem como objetivo de criar um banco de dados para elaboração de dashboards operacionais. O escopo da pesquisa inclui a gestão e a governança de dados (GD), com foco nos aspectos legais e relativos aos metadados, à qualidade, privacidade, segurança e proteção desses ativos. **Método:** utilizou-se das pesquisas bibliográfica e documental em uma abordagem descritiva-exploratória, de natureza qualitativa. **Resultados:** realizou-se a compilação do arcabouço legal acerca do tratamento de dados no Brasil, no qual se destaca a Lei Geral de Proteção de Dados (LGPD); o levantamento e a compreensão de dados de recursos humanos (RH), a partir dos documentos e processos relacionados à carreira docente; e a criação de um dicionário de dados (DD). **Conclusão:** A literatura infere que a dicionarização em projetos que envolvem dados institucionais é de suma importância para garantir a usabilidade e a credibilidade do artefato produzido. O procedimento deve observar o ciclo de vida dos dados, partindo da exploração dos metadados, bem como as legislações vigentes. No entanto, apesar dos benefícios vislumbrados pelos estudiosos do tema, os metadados ainda são pouco considerados na prática. Assim, esta pesquisa se propôs a contribuir para a melhoria dos serviços públicos e para a ampliação da produção científica ao abordar e testar premissas teóricas insuficientemente empregadas no ambiente corporativo.

Palavras-chave: Dicionário de dados; Qualidade de dados; Levantamento de dados; LGPD.

Abstract

Introduction: This is a study on the dictionaryization of personal data of professors at a federal university, with the aim of creating a database for the development of operational dashboards. The scope of the research includes data management and governance (DG), focusing on legal aspects and metadata, quality, privacy, security, and protection of these assets. **Method:** The study utilized bibliographical and documentary research in a descriptive-exploratory, qualitative approach. **Results:** the compilation of the legal framework regarding data in Brazil was conducted, with a highlight on the General Data Protection Law; the collection and understanding of human resources (HR) data, based on documents and processes related to the teaching career; and the creation of a data dictionary (DD). **Conclusions:** The literature suggests that dictionaryization in projects involving institutional data is of utmost importance to ensure the usability and credibility of the artifact produced. The procedure should consider the data life cycle, starting with metadata exploration, as well as current regulations. However, despite the benefits envisioned by scholars in the field, metadata is still often overlooked in practice. Thus, this research aimed to contribute to the improvement of public services and to the expansion of scientific production by addressing and testing theoretical premises insufficiently applied in the corporate environment.

Keywords: Data dictionary; Data quality; Data collection; General Data Protection Law.

INTRODUÇÃO

Este estudo tem o objetivo de registrar o processo de dicionarização dos dados de recursos humanos (RH) da Universidade Federal da Integração Latino-Americana (UNILA), contemplando a LGPD. Trata-se de uma etapa do projeto de construção de um sistema de *Business Intelligence* (BI) com vistas à otimização da rotina administrativa e à facilitação do processo de tomada de decisão no âmbito do Instituto Latino-Americano de Arte, Cultura e História (ILAACH).

A motivação para a elaboração do BI surgiu da observação da rotina administrativa do Instituto, que recebe um grande volume de documentos e processos que em sua maioria são relacionados ao desenvolvimento da carreira do corpo docente. Por esta razão, a área também é acionada para responder às necessidades informacionais da gestão superior acerca do público mencionado. Para atendimento dessas demandas, o ILAACH extrai dados do Sistema Integrado de Gestão (SIG) e os complementa, quando necessário, com informações contidas em um banco de dados (BD) em formato de planilha, que visa a diminuição de inconsistências, os problemas de

atualização, acesso, falta de dados, entre outros. Porém essas fontes de dados são pouco intuitivas e acabam exigindo muito tempo e esforço para visualização e compreensão das informações necessárias para a tomada de decisão, por exemplo, a organização da equipe técnica de acordo com a quantidade e sazonalidade das demandas.

Com base neste diagnóstico, resultado da aplicação de questionário às gestoras da unidade, de análise documental e da observação participante, acredita-se que o BI pode contribuir para a otimização dos processos administrativos e decisórios do ILAACH por meio de painéis interativos (*dashboards*). De acordo com Neto (2021), o BI pode ser entendido como uma função da gestão de dados, assim como a GD é uma disciplina estratégica para as organizações em tempos de transformação digital.

Ao consultar a bibliografia disponível percebe-se que o termo “gestão de dados” começou a ser empregado no contexto organizacional por volta de 2010, como uma evolução da matéria até então conhecida como administração de dados (AD) (Rêgo, 2020). Esse processo foi promovido pela constatação de diversos problemas no uso dos sistemas ERP (*Enterprise Resource Planning*) e das ferramentas de BI (Rêgo, 2020). Na busca por solucionar essas questões, a gestão de dados passou a focar no ciclo de vida desses recursos (considerando os metadados), e no entrelaçamento de tecnologias da informação (TI) com os negócios, aspectos que não estavam contemplados no escopo da AD (Barbieri, 2020; Rêgo, 2020). Nesse sentido, a Escola Nacional de Administração Pública (2022, p. 9) acredita que é viável aprimorar a eficácia do ciclo de vida dos dados ao “analisá-lo a partir do olhar das novas tecnologias e da LGPD”. Dessa forma é possível promover um gerenciamento mais eficiente, além de atender aos direitos dos titulares dos dados, conforme estabelecido pela lei.

Portanto, este artigo engloba informações relacionadas à gestão de dados e GD, priorizando aspectos legais e relativos aos metadados, à qualidade, à privacidade, à segurança e à proteção desses ativos, para que sua utilização nos processos administrativos e de tomada de decisão possa ser mais eficiente.

Gestão de dados e GD

Para melhor compreensão do conceito de gestão de dados, pode-se lançar mão da obra de Rêgo, que versou sobre o tema em diferentes momentos. Em uma definição inicial o autor apresentou o termo como uma “disciplina responsável por definir, planejar, implantar e executar estratégias, procedimentos e práticas necessárias para gerenciar de forma efetiva os recursos de dados e informações das organizações, incluindo planos para sua definição, padronização, organização, proteção e utilização” (Rêgo, 2013, p. 48). Posteriormente, absorvendo os ensinamentos do guia de melhores práticas conhecido como *Guide to The Data Management Body of Knowledge* (DAMA-DMBOK), o autor incrementou o texto, afirmando que esta disciplina viabiliza um conjunto de onze funções, também denominadas áreas de conhecimento. De acordo com Rêgo (2020, p. 9), essas funções são:

(...) governança de dados; arquitetura de dados; modelagem e projeto de dados; armazenamento e operações de dados; segurança dos dados; integração e interoperabilidade; gestão de conteúdo e documentos; dados mestres e referenciais; *data warehousing e business intelligence*; gestão de metadados; qualidade de dados.

Nesta perspectiva, nota-se a progressão conceitual da gerência de dados à medida que incorpora conceitos correlatos, dentre os quais, destaca-se o da GD. Isso se deve à sua preocupação com a melhor organização e integração “dos dados e metadados das organizações, fazendo com que sejam aderentes às necessidades do negócio, únicos, íntegros, confiáveis, manuteníveis, conhecidos, performáticos, legíveis e disponíveis a quem realmente precisa ter o acesso” (Rêgo, 2013, p. 48). Partindo deste entendimento, Rêgo (2020, p. 9), afirma que a GD “representa a autoridade em relação à gestão de dados e ao controle de estratégias, políticas, padrões, processos, métricas e indicadores envolvidos com os ativos de dados”. Em outras palavras, seu papel é o de governar as outras funções (Rêgo, 2020).

Em complemento, Barbieri (2020) afirma que o conceito de GD evolui constantemente em amplitude, pluralidade e interdisciplinaridade, porém, sempre acaba convergindo para a qualidade de dados, em um sentido amplo desse constructo. O autor interpreta esse foco maior na qualidade como consequência do crescimento da GD, do surgimento de novas fontes de dados e das regulações, que se fazem cada vez mais presentes e exigentes no contexto da sociedade digital. Para ele, a GD passa por um processo de busca de maturidade nas empresas, envolvendo a gestão, valoração e produção dos dados, bem como a monitoração de seu uso e aspectos críticos de segurança, privacidade, ética e *compliance*.

O conjunto de boas práticas em GD apresenta outras abordagens como o DAMA-DMBOK *Guide*, o *Framework* 5W2H; o *Framework* da IBM; o Modelo EDM *council* e o Modelo *Data Management Maturity* (DMM). Todos esses modelos apresentam definições semelhantes de GD e remetem para o exercício de funções de gestão de dados, relacionando-as à qualidade (Neto, 2021). Esta, por sua vez, pode ser aferida por meio de várias dimensões. Entre as principais dimensões mencionadas por McGilvray (2008), Loshin (2009) e Barbieri (2020), estão: 1) linhagem/origem; 2) semântica; 3) estrutura; 4) completude; 5) consistência; 6) atualidade; 7) disponibilidade; 8) razoabilidade; 9) identificação; 10) segurança e privacidade; 11) especificação; 12) integridade; 13) duplicação; 14) acurácia; 15) consistência e sincronização; 16) temporalidade e disponibilidade; 17) facilidade de uso e

manutenibilidade; 17) cobertura; 18) qualidade de apresentação; 19) relevância, confiança e envelhecimento dos dados; e 20) transacionalidade.

Os Metadados

Conforme a introdução, o gerenciamento do ciclo de vida dos dados e metadados - *Data Life Cycle Management* desempenha um papel fundamental na mitigação de problemas relacionados ao uso de sistemas ERP e de BI (Barbieri, 2020; Rêgo, 2020). Isso implica na necessidade de compreender detalhadamente os metadados e a linhagem dos dados, ou seja, entender como esses dados fluem através dos diversos sistemas, desde sua origem até o momento em que são descartados ou armazenados (Barbieri, 2020).

Mas o que são dados e metadados e qual a relação entre eles? A compreensão desses conceitos e da dinâmica que os une passa pelo entendimento da hierarquia existente entre dados, informação, conhecimento e sabedoria. Laudon e Laudon (2013, p. 12), definem os dados como “uma sequência de fatos ainda não analisados, representativos de eventos que ocorrem nas organizações ou no ambiente físico, antes de terem sido organizados e arranjados de uma forma que as pessoas possam entendê-los e utilizá-los”. A informação é um conjunto de dados interligados e relacionados que entregam valor dentro de um contexto (Laudon & Laudon, 2013). Já o conhecimento, é caracterizado como a capacidade de compreender as coisas, incluindo informações, por meio da razão ou da experimentação e experiência (Barbieri, 2020). O último elemento desta relação, a sabedoria, é compreendido como a utilização eficiente e eficaz do conhecimento (Rêgo, 2013).

No entendimento de Barbieri (2020), nesse encadeamento conceitual, ao sair da camada de dados, passa-se às camadas de informação, conhecimento e sabedoria, por meio do enriquecimento dos dados com seus respectivos metadados, também conhecidos como “dados sobre dados”. Segundo Arakaki e Arakaki (2020), esta é a definição mais difundida e aceita no campo da ciência da informação, cuja literatura foi encorpada ao longo dos anos por autores como Méndez Rodríguez (2002), Zeng e Qin (2008, 2016), Alves (2014) Alves (2010, 2014), Alves e Santos (2013), Pomerantz (2015) e Joudrey, Taylor e Wisser (2018).

Os conjuntos de metadados são chamados de catálogos ou DDs, e armazenam uma gama de informações essenciais, incluindo: nomes de relações e de atributos, especificações de domínios, descrições de índices, definições de restrições, visões e procedimentos (Alves, 2014; Lancaster, 1979). De forma similar, a Escola Técnica da Universidade Vale do Rio Doce (2014) entende que o DD é uma documentação do Diagrama Entidade Relacionamento, no qual estão relacionados os seguintes elementos:

a) entidade – pessoa, objeto ou lugar pelo qual temos interesse em guardar informações; b) atributos – características da entidade; c) classe – característica do atributo, que pode ser: simples; composto; multivalorado (a exemplo do número de telefone, que pode ser celular e fixo); ou determinante (aquele que pode ser usado como chave, como CPF, código, matrícula); d) domínio – pode ser numérico, textual, em formato de data ou booleano; e) tamanho – define a quantidade de caracteres necessários para armazenar o conteúdo; e f) descrição – serve para dar informações adicionais ao atributo.

Segundo Alves (2014), o conceito de DD tem sua origem na disciplina de modelagem de banco de dados e sua relevância está atrelada à necessidade de padronização do ambiente informatizado. Para Barboza e Freitas (2018), o artefato é uma composição de tabelas consultivas para a construção de outras tabelas do Sistema Gerenciador de Banco de Dados (SGBD). Recentemente, o site da Aquarela Analytics (2022) definiu o DD como um documento que descreve significado, origem, relacionamento e uso dos dados de forma estruturada.

De forma complementar, as duas últimas fontes mencionadas enfatizam a importância do DD para o bom funcionamento da base de dados ao afirmarem que o dispositivo garante a integridade, a consistência e a padronização dos dados à medida que o BD cresce e outros desenvolvedores passam a trabalhar no projeto. Além disso, a Aquarela Analytics (2022) assegura que esse elemento é um ponto de partida consensual, objetivo e livre de ambiguidades, que torna viável a reconstrução do cenário em que os dados foram coletados, levando a uma notável melhoria na qualidade das avaliações realizadas a partir deles.

No entanto, ainda que a literatura aponte a relevância da gestão dos conjuntos de metadados (DDs) nas organizações, estas ainda não a empregam de forma consistente. Nesse sentido, em pesquisa realizada pela Fumsoft em parceria com a DAMA Brasil, em 2012, com o objetivo de mapear o posicionamento das empresas em todas as áreas de conhecimento da gestão dados, Barbieri (2020, p. 216) relata que obteve a seguinte impressão:

O processo ou corpo de conhecimento de metadados foi aquele em que praticamente todas as respostas ficaram fora do viés positivo, ou seja, houve a predominância absoluta da resposta “A atividade não é realizada”, com um grau médio de “não sei” de quase 8%. A única resposta com viés positivo foi o entendimento dos metadados, na clássica imagem de que se entende a importância daquilo como requisito, mas não se implanta.

Ao analisar a situação, Barbieri (2020) pondera que, embora os metadados sejam frequentemente subestimados, os problemas resultantes da negligência desses ativos justificam sua inclusão em projetos que envolvem dados organizacionais. O autor argumenta que em uma era de transformação digital conhecer os dados e seus metadados

com profundidade, catalogando-os para garantir entregas precisas ao destinatário correto e no momento adequado, representa o primeiro passo para estabelecer um plano de identificação de riscos e medidas de mitigação, a fim de alcançar conformidade com a LGPD.

MÉTODO E MATERIAIS

Para a elaboração deste trabalho foram realizadas pesquisas bibliográfica e documental, de forma a caracterizar uma abordagem metodológica descritiva-exploratória, de natureza qualitativa. Concluído o embasamento teórico que delineou a investigação, iniciou-se a pesquisa documental, evidenciada no arrolamento da legislação nacional aplicável à GD, e na busca das normativas institucionais, realizada para demonstrar os mecanismos de controle disponíveis para proteção dos dados no âmbito da UNILA.

Na sequência, realizou-se o levantamento e o entendimento dos dados de RH da Universidade. Para isso, optou-se pelo preenchimento da planilha disponibilizada pela Controladoria Geral do Estado do Paraná (CGE) para implementação da LGPD nas instituições públicas. A referida planilha é composta por duas abas, sendo uma para detalhamento do(s) processo(s) e outra para registro e caracterização dos dados. Nesta pesquisa, apenas a segunda aba foi utilizada, pois os processos relacionados à gestão de pessoas na UNILA já foram mapeados. No documento foram inseridos dados e metadados referentes a quatro processos relacionados à gestão da carreira docente do Magistério Superior Federal, escolhidos por conveniência: estágio probatório, retribuição por titulação, progressão/promoção docente e aceleração da promoção. Estas informações, portanto, passaram a compor o DD de RH.

RESULTADOS

O conteúdo está organizado em três etapas, sendo que a primeira demonstra o aparato legal identificado, a segunda, detalha o processo de levantamento e compreensão dos dados e a terceira apresenta o DD construído a partir do método investigativo.

Do amparo legal

A literatura utilizada infere que conhecer as regulações às quais a instituição está sujeita é um ponto importante na definição das dimensões de qualidade de dados (Barbieri, 2020). A partir do levantamento realizado nesta pesquisa, identificou-se na esfera federal, um aparato legal que abrange: proteção de dados pessoais, estrutura da segurança da informação e cibernética, governança, e compartilhamento de dados. No âmbito da UNILA, foram identificadas duas portarias e uma resolução que visam atender às normativas vigentes em território nacional. A tabela 1 fornece uma descrição detalhada dos resultados encontrados:

Instrumento Legal	Data	Órgão Emissor	Ementa
Lei nº 13.709 (2018)	14/08/2018	Presidência da República	Lei Geral de Proteção de Dados (LGPD).
Lei nº 13.853/2019	08/07/2019	Presidência da República	Altera a Lei nº 13.709, de 14 de agosto de 2018, para dispor sobre a proteção de dados pessoais e para criar a Autoridade Nacional de Proteção de Dados; e dá outras providências.
Portaria nº 93/2019	26/09/2019	Presidência da República	Aprova o Glossário de Segurança da Informação.
Decreto nº 10.046/2019	09/10/2019	Presidência da República	Dispõe sobre a governança no compartilhamento de dados no âmbito da administração pública federal e institui o Cadastro Base do Cidadão e o Comitê Central de Governança de Dados.
Instrução Normativa nº 1/2020	27/05/2020	Presidência da República	Dispõe sobre a Estrutura de Gestão da Segurança da Informação nos órgãos e nas entidades da administração pública federal.
Instrução Normativa nº 2/2020	24/07/2020	Presidência da República	Altera a Instrução Normativa nº 1, de 27 de maio de 2020, que dispõe sobre a Estrutura de Gestão da Segurança da Informação nos órgãos e nas entidades da administração pública federal.
Instrução Normativa nº 4/2020	27/03/2020	Presidência da República	Dispõe sobre os requisitos mínimos de Segurança Cibernética que devem ser adotados no estabelecimento das redes 5G.
Instrução Normativa nº 3/2021	31/05/2021	Presidência da República	Dispõe sobre os processos relacionados à gestão de segurança da informação nos órgãos e nas entidades da administração pública federal.
Instrução Normativa nº 5/2021	31/08/2021	Presidência da República	Dispõe sobre os requisitos mínimos de segurança da informação para utilização de soluções de computação em nuvem pelos órgãos e pelas entidades da administração pública federal.
Portaria nº 260/2021/GR (2021)	19/07/2021	UNILA	Institui o Comitê de Governança Digital da UNILA.
Portaria nº 261/2021/GR (2021)	19/07/2021	UNILA	Disciplina a Estrutura de Gestão de Segurança da Informação da UNILA.
Portaria nº 3/2022/GR (2022)	25/07/2022	UNILA	Estabelece a Política de Segurança da Informação – POSIN, da UNILA.

Tabela 1. Legislação sobre Governança de Dados

Ao analisar a tabela 1 nota-se, que o foco no tratamento de dados pessoais foi incluído na LGPD em meados de 2019, por meio de uma modificação da Lei nº 13.709/2018. Essa emenda objetivou adequar a legislação à realidade do tratamento de dados na era digital, segundo Barbieri (2020), havia um claro descompasso entre a legislação e a velocidade de desenvolvimento das tecnologias. Portanto, a lei precisou evoluir para ser aplicada em tecnologias como *big data*, internet das coisas e inteligência artificial (Barbieri, 2020).

Além disso, também é possível notar pela cronologia de publicação dos dispositivos elencados, e pelos conteúdos tratados por cada um deles, que a LGPD impulsionou uma maior pressão governamental sobre as organizações dos setores público e privado, no sentido de efetivarem a implementação de mecanismos de segurança capazes de prever e aplacar possíveis incidentes. A partir dessa atualização regulamentar, aqueles que incorrerem em mau uso de dados pessoais, estão sujeitos a sanções administrativas e danos à sua reputação, conforme apontam os trabalhos de Barbieri (2020) e Lima (2021). No entanto, há uma margem procedimental para que as organizações definam, customizem, reajustem e demonstrem êxito em sua jornada de conformidade legal (Lima, 2021). Isso porque a LGPD não estipula métodos para o manuseio de dados pessoais, mas exige o cumprimento de finalidades.

Partindo desse entendimento, o Guia Orientativo da Autoridade Nacional de Proteção de Dados (Aquarela Analytics, 2022) menciona que o tratamento de dados pessoais em órgãos públicos deve estar atrelado à uma necessidade específica (finalidade), previamente informada aos titulares dos dados, a fim de se obter permissão para o procedimento. Paralelamente, a Escola Nacional de Administração Pública (2022), afirma que a identificação da base legal é um fator importante sobre o tratamento desse tipo de dados. As bases legais previstas na legislação, de acordo com o guia supracitado, podem ser conferidas na tabela 2.

Base legal	LGPD	Características
Consentimento	Art. 5º, XII e art. 11, I (dados sensíveis)	pressupõe uma escolha entre autorizar e recusar o tratamento dos dados pessoais, incluindo a possibilidade de revogar o consentimento a qualquer momento. Adicionalmente, no caso de dados sensíveis, o consentimento deve ser fornecido “de forma específica e destacada, para finalidades específicas”.
Legítimo interesse	Art. 7º, IX	aplicável ao tratamento de dados pessoais de natureza não sensível para atendimento de interesses legítimos do controlador ou de terceiros, “exceto no caso de prevalecerem direitos e liberdades fundamentais do titular”.
Obrigação legal regulatória	Art. 7º, II	é apropriada quando o tratamento de dados pessoais é realizado de forma compulsória ou quando for necessário para o cumprimento de obrigações e atribuições legais do poder público.
Execução de políticas públicas	Art. 7º, III	a “administração pública” pode realizar “o tratamento e uso compartilhado de dados necessários à execução de políticas públicas previstas em leis e regulamentos ou respaldadas em contratos, convênios ou instrumentos congêneres”.

Tabela 2. Bases legais aplicáveis ao tratamento de dados

Fonte: Adaptado de “Tratamento de dados pelo Poder Público” [Guia orientativo], ANPD (2022, pp. 6-12).

No caso em tela, convém a exploração mais detalhada da obrigação legal regulatória, de acordo com [Garcia \(2020\)](#), esta finalidade ampara a coleta e tratamento de dados pelo poder público sem consentimento expresso dos seus titulares, desde que haja persecução do interesse público, e que seja para execução de suas competências legais ou cumprimento de suas atribuições. No entanto, isso não exclui a necessidade de transparência no processo.

Finalizadas as considerações sobre a LGPD, suas implicações e contribuições, passa-se a analisar os atos normativos emitidos pela UNILA. A Portaria nº 260/2021/GR trata da criação de um comitê de governança digital, com competência deliberativa sobre assuntos relativos à GD, implementação das ações de governo digital e uso dos serviços e recursos de tecnologia da informação e comunicação (TIC). A Portaria nº 261/2021/GR visa estabelecer atribuições e estruturar a gestão de segurança da informação na Unila, bem como instituir equipe de prevenção, tratamento e resposta a incidentes cibernéticos. A Resolução nº 03/2022/CGirc, apesar de não revogar a Portaria nº 261/2021/GR, aborda a mesma temática, porém de forma mais ampla, incluindo o tratamento de informações pertencentes a UNILA ou sob sua responsabilidade, assim como a promoção, conscientização e capacitação de RH em segurança da informação. Nesse sentido, convém destacar os seguintes pontos da Resolução:

Art. 9º Os ativos de informação, de acordo com sua criticidade, devem ser identificados, inventariados, mantidos atualizados, possuírem um responsável e serem protegidos contra acessos indevidos.

Parágrafo único. Ativos de informação não inventariados, não gerenciados, ou de uso particular, não poderão acessar a rede administrativa da instituição, exceto em condições excepcionais tratadas em norma complementar específica. (...)

Art. 11. A informação armazenada em estações de trabalho, dispositivos móveis ou mídias externas é de responsabilidade do usuário, cabendo ao mesmo adotar as medidas necessárias para evitar a perda de dados. (...)

Art. 34. É vedado o uso de aplicações e repositórios não homologados para processar, armazenar ou publicar informações (...) salvo casos de informação sem restrição de acesso e que tenham sido adotadas medidas de preservação contra perda de dados e aprisionamento tecnológico.

Art. 35. No processo de tratamento da informação, deve ser garantida a privacidade, a disponibilidade, a integridade e a confidencialidade dos dados pessoais em todo o seu ciclo de vida, em qualquer formato de armazenamento e suporte.

Como se vê, há regras claras de gestão de dados na Universidade, mas sua implementação não está totalmente consolidada. Um exemplo disso é que a Resolução foi aprovada há mais de um ano, e até o momento, a Instituição não dispõe de um inventário de dados, conforme disposto no art. 9º. Como a LGPD não impõe métodos nem prazos para as organizações se adequarem, situações como essa podem se arrastar por tempo indefinido. Essa omissão da LGPD acaba também por dificultar a aplicação de penalidades pelos órgãos fiscalizadores ([Garcia, 2020](#)).

Ainda em relação à POSIN, pontua-se a necessidade de considerar o ciclo de vida dos dados, disposta no art. 35º, o que endossa o disposto na literatura de suporte e reforça a pertinência do conhecimento dos dados e metadados, bem como a elaboração do DD proposto. Por fim, ressalta-se que o BD que alimentará os painéis interativos deverá estar contido em ambiente institucionalmente homologado e capaz de oferecer as condições adequadas de segurança, privacidade, disponibilidade, integridade e confidencialidade, exigidas no trato de dados pessoais.

Do levantamento e compreensão dos dados

A planilha da CGE é estruturada de forma a permitir o registro dos dados envolvidos no projeto e de seus metadados, viabilizando a compreensão dos fluxos e demais aspectos exigidos pela LGPD. Nela foram inseridas as seguintes informações: processo(s) relacionado(s) ao dado, setor de início do processo, documento no qual o dado deve estar contido, tipo de coleta necessário, identificação/nome do dado, se o dado é sensível ou não, finalidade/base legal, necessidade que fundamenta o tratamento do dado, ciclo de vida, informações sobre compartilhamento com outras áreas internas e/ou externas, formas de tratamento, meio de armazenamento e mecanismos de controle disponíveis para proteção do dado.

Feito isso, foram identificados treze tipos de dados e/ou metadados pessoais nos processos em análise. Como essas informações são inseridas nos requerimentos pelos próprios docentes, que os encaminham ao ILAACH por e-mail, considera-se que a coleta é realizada de forma direta. No entanto, é comum que os professores consultem o Instituto durante o preenchimento da requisição, pois muitas vezes desconhecem ou não lembram de determinadas informações. Nesses casos, os servidores técnicos recorrem ao SIG e à planilha setorial, que podem apresentar inconsistências, como duplicação de matrícula SIAPE ou marcos temporais divergentes. Para evitar problemas dessa ordem, sugere-se utilizar o Portal do Sistema Integrado de Administração de Pessoal (SIAPEnet), do Governo Federal, como fonte oficial de coleta de dados dessa natureza.

Quanto aos dados sensíveis, não foi identificada a necessidade de coletá-los para realização dos trâmites administrativos em foco neste trabalho, uma vez que estão relacionados aos aspectos mais íntimos de um indivíduo, como: origem racial ou étnica; convicção religiosa; opinião política; filiação a sindicato ou a organização de caráter religioso, filosófico ou político; saúde; vida sexual; e dados genéticos ou biométricos ([Aquarela Analytics, 2022](#)).

Com relação à base legal que fundamenta o tratamento dos dados levantados, aplicou-se a obrigação legal regulatória decorrente de normas de organização, pelos seguintes motivos: necessidade de identificar servidores e serviços solicitados em determinado período; contabilizar prazos legais; e identificar os meios adequados de contato com o público-alvo. Apenas um dos treze dados elencados ensejava a declaração de consentimento antes de ser manuseado e optou-se por excluí-lo do DD e da base de dados que suportará os *dashboards*.

Quanto ao ciclo de vida dos dados em pauta, o [Conselho Nacional de Arquivos \(2001\)](#), determina que documentos relacionados ao assentamento de pessoal devem ser mantidos em arquivo corrente ou setorial por cinco anos, devendo então, ser encaminhados para o arquivo central (fase intermediária), onde permanecerão por 47 (quarenta e sete) anos e finalmente, poderão ser eliminados.

Os dados elencados no contexto deste estudo são compartilhados internamente, entre o instituto de lotação do docente (no caso o ILAACH), a Comissão Permanente de Pessoal Docente (CPPD) e a Pró-Reitoria de Gestão de Pessoas (PROGEPE). Não há necessidade de compartilhamento de dados com entes externos à UNILA.

Dentre as formas de tratamento dos dados coletados no âmbito do ILAACH, predominaram a recepção, o acesso, a avaliação da informação, a utilização dos dados e a comunicação.

Para finalizar esta etapa, registraram-se os mecanismos de controle utilizados institucionalmente, os quais podem ser conferidos na descrição do arcabouço legal.

Do DD

A partir da compreensão dos dados e metadados, iniciou-se a elaboração do DD de RH, relacionado à entidade 'carreira docente'. Uma vez descartado o dado que ensejava autorização expressa para tratamento (telefone para contato), a tabela representativa do DD, apresentada na tabela 3, foi estruturada com cinco colunas, preenchidas de acordo com as recomendações da [Escola Técnica da Universidade Vale do Rio Doce \(2014\)](#):

a) código – ordenação dos atributos, de 1 a 12; b) atributos – interessado, SIAPE, data de início de exercício no Sistema Federal de Ensino Superior, classe, nível, data da última progressão/promoção, regime de trabalho, área da CAPES, data de início de exercício na UNILA, avaliação (I, II, III ou etapa única), avaliador, e-mail institucional. c) domínio – foram identificadas informações do tipo textual (6), numérico (2) e data (3); d) classe – foram identificados atributos simples (10), determinante (1) e multivalorado (1); e) descrição – de acordo com regulamentos e documentos oficiais.

Entidade: Carreira Docente				
Código	Atributo	Domínio	Classe	Descrição
1	Interessado	Texto	Simple	Nome do servidor solicitante da do serviço. Nos formulários este atributo também poderá ser encontrado como 'nome do avaliado'.
2	SIAPE	Número	Determinante	Matrícula composta por sete dígitos, utilizada para comprovar vínculo com o órgão federal.
3	Data de início de exercício no Sistema Federal de Ensino Superior	Data	Simple	Data de início das atividades docentes no Sistema Federal de Ensino Superior.
4	Classe	Texto	Simple	Classe da carreira do Magistério Superior Federal.
5	Nível	Número	Simple	Nível da Classe da carreira do Magistério Superior Federal.
6	Data da última progressão/promoção	Data	Simple	Data da publicação da última portaria de progressão ou promoção do servidor.
7	Regime de trabalho	Texto	Simple	Quantidade de horas de trabalho semanal do servidor.
8	Área da CAPES	Texto	Simple	Uma das 49 áreas de avaliação agregadas pela CAPES por critério de afinidade.
9	Data de início de exercício na UNILA	Data	Simple	Data de início das atividades docentes na UNILA.
10	Avaliação (I, III ou etapa única)	Texto	Simple	Estágio de avaliação, considerando que cada servidor passa por 3 avaliações, ou é avaliado em etapa única, nos casos excepcionais.
11	Avaliador	Texto	Multivalorado	Nome da chefia imediata. Em casos de ausência do titular do cargo, podem avaliar o servidor: o substituto legal da chefia, a chefia imediatamente superior.
12	E-mail institucional	Texto	Multivalorado	Correio eletrônico institucional

Tabela 3. Dicionário de dados de RH

Destaca-se que o atributo 'interessado' também pode ser encontrado nos formulários de solicitação de serviço como 'nome do avaliado'. Porém, para efeitos da construção do BD, convencionou-se a utilização padronizada do termo 'interessado'. Em relação ao atributo 'e-mail institucional', nos documentos institucionais se utiliza o termo 'endereço de e-mail', sem mencionar o tipo de conta (profissional ou pessoal). Assim, considerando que o e-mail institucional deve ser acessado por todos os servidores diariamente e é um dado público, que independe do consentimento formal para seu tratamento, optou-se pela utilização do e-mail corporativo.

Por fim, salienta-se que esta é a primeira versão do DD, pois à medida que os trabalhos avançam e o BD cresce, torna-se necessário atualizar seu catálogo de dados.

CONSIDERAÇÕES FINAIS

Este trabalho representa os primeiros passos em direção à construção de um serviço de BI pensado para auxiliar os gestores do ILAACH/UNILA no planejamento das tarefas administrativas do setor, bem como no provisionamento de informações coerentes e tempestivas às instâncias superiores, sem deixar de garantir os direitos dos titulares dos dados.

De acordo com a literatura, o êxito em projetos de dados organizacionais está atrelado à gestão e governança desses ativos, que se acumulam em velocidade cada vez maior devido ao advento da internet e dos avanços tecnológicos promovidos por ela. Nessa era de *big data*, as regulamentações têm ocupado lugar de destaque ao pressionar as organizações a realizarem um tratamento de dados seguro e ético, principalmente no manuseio de informações pessoais. Nesse sentido, a LGPD se destaca em território nacional, como um marco legal que influenciou no surgimento de dispositivos complementares com vistas à uma efetiva implementação da GD em ambientes corporativos. No entendimento dos autores consultados, tudo começa pelo conhecimento dos dados e dos metadados associados ao negócio, que devem ser rastreados em todas as etapas do seu ciclo de vida, catalogados e controlados, visando minimizar possíveis prejuízos advindos de uma má gestão.

Nesse cenário, entende-se que a dicionarização de elementos que compõe sistemas informatizados em projetos que envolvam dados institucionais é de suma importância para que se possa extrair informações corretas, adequadas, atualizadas e em *compliance* com as regulamentações vigentes, garantindo a usabilidade e credibilidade do artefato de trabalho. No entanto, a aposta no gerenciamento de metadados para assegurar a qualidade desses

projetos, ainda reside no campo teórico, carecendo de esforços organizacionais para sua efetivação. Sendo assim, esse trabalho se propôs a colocar em prática a teoria estudada, por meio da catalogação dos dados de RH dos doentes do ILAACH, contextualizando-os no ambiente universitário pelo registro de fluxos, características e mecanismos de controle empregados durante seu tratamento.

Acredita-se que o presente trabalho contribuirá para a melhoria dos serviços públicos ao passo que se propõe a elaborar uma solução tecnológica de gerenciamentos de dados, fundamentada em premissas atuais de gestão de dados, de forma a concretizar recomendações ainda pouco praticadas institucionalmente. O fato de a LGPD ser uma conquista recente da sociedade brasileira, ainda em fase de implementação e consolidação, somado à escassez de práticas relacionadas à melhor compreensão dos dados a partir do contexto trazido pelos seus respectivos metadados, também abre caminho para futuras investigações, a partir desta. entende-se que os resultados da pesquisa vão de encontro aos preceitos teóricos difundidos na academia pelos autores citados no artigo. A presente pesquisa, no entanto, não teve objetivo de esgotar o assunto, mas sim de complementar os estudos já realizados. Assim, espera-se que as nuances abordadas possam contribuir para o aprimoramento e maior eficiência da gestão pública e para o incremento da literatura científica.

REFERÊNCIAS

- Alves, W. P. (2014). *Banco de dados*. São Paulo: Saraiva.
- Aquarela Analytics. (2022). *Guia orientativo de tratamento de dados pessoais pelo poder público*. Recuperado de <https://www.cge.pr.gov.br/Pagina/Cartilhas-da-Lei-Geral-de-Protecao-de-Dados-LGPD>
- Arakaki, A. C. S., & Arakaki, F. A. (2020). Dados e metadados: conceitos e relações. *Ci. Inf*, 49(3), 34-45. Recuperado de <https://doi.org/10.18225/ci.inf.v49i3.5504>
- Barbieri, C. (2020). *Governança de dados*. Rio de Janeiro: Editora Alta Books.
- Barboza, F., & Freitas, P. (2018). *Modelagem e desenvolvimento de banco de dados*. Porto Alegre: Grupo A.
- Conselho Nacional de Arquivos. (2001). *Classificação, temporalidade e destinação de documentos de arquivo relativos às atividades-meio da administração pública*. Rio de Janeiro: Arquivo Nacional.
- Escola Nacional de Administração Pública. (2022). *Gestão do conhecimento no setor público: Etapas para a implementação da gestão do conhecimento*. Recuperado de <https://mooc38.escolavirtual.gov.br/course/view.php?id=12379>
- Escola Técnica da Universidade Vale do Rio Doce. (2014). *Aula 5 - dicionário de dados*. Recuperado de <https://pt.slideshare.net/janynnegomes/aula-5-41778197>
- Garcia, L. R. (2020). *Lei geral de proteção de dados (lgpd): Guia de implantação*. São Paulo: Editora Blucher. Recuperado de <https://app.minhabiblioteca.com.br/books/9786555060164>
- Lancaster, F. W. (1979). *Information retrieval systems: Characteristics, testing, and evaluation*. Nova Jersey: John Wiley & Sons.
- Laudon, K. C., & Laudon, J. P. (2013). *Sistemas de informação gerenciais*. Londres: Pearson Prattice Hall.
- Lei nº 13.709. (2018). *Lei geral de proteção de dados pessoais (lgpd)*. Recuperado de https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm
- Lima, A. P. M. C. D. (2021). *Lgpd aplicada*. Rio de Janeiro: Grupo GEN.
- Loshin, D. (2009). *Master data management*. Massachusetts: Morgan Kaufmann.
- McGilvray, D. (2008). *Executing data quality projects: Ten steps to quality data and trusted information*. Massachusetts: Morgan Kaufmann.
- Neto, A. P. D. A. (2021). *Governança de dados*. São Paulo: Editora Saraiva.
- Portaria nº 260/2021/GR. (2021). *Universidade federal da integração latino-americana*. Recuperado de <https://l1nq.com/B5YFM>
- Portaria nº 261/2021/GR. (2021). *Universidade federal da integração latino-americana*. Recuperado de <https://acesse.dev/klbwH>
- Portaria nº 3/2022/GR. (2022). *Universidade federal da integração latino-americana*. Recuperado de <https://l1nq.com/kh20N>
- Rêgo, B. L. (2013). *Gestão e governança de dados: Promovendo dados como ativo de valor nas empresas*. Rio de Janeiro: Brasport.
- Rêgo, B. L. (2020). *Simplificando a governança de dados*:

governe os dados de forma objetiva e inovadora. Rio de Janeiro: Brasport.

Como citar este artigo (APA):

Maurício, Claudio Roberto Marquette, Dechechi, Eduardo César, Silva, Willian Francisco da, Cruz, Andréia da (2024). Dicionário de dados: um olhar de qualidade sobre os dados de recursos humanos da Universidade Federal da Integração Latino-Americana. *AtoZ: novas práticas em informação e conhecimento*, 13, 1 – 12. Recuperado de: <http://dx.doi.org/10.5380/atoz.v13.90347>

Questionário de Dados

[illegible]

NOTAS DA OBRA E CONFORMIDADE COM A CIÊNCIA ABERTA

CONTRIBUIÇÃO DE AUTORIA

Papéis e contribuições	Claudio Roberto Marquette Mauricio	Andréia da Cruz	Eduardo César Dechechi	Willian Francisco da Silva
Concepção do manuscrito	X	X	X	X
Escrita do manuscrito		X		
Metodologia	X	X		X
Curadoria dos dados		X		
Discussão dos resultados	X	X	X	X
Análise dos dados		X		

EQUIPE EDITORIAL

Editora/Editor Chefe

Paula Carina de Araújo (<https://orcid.org/0000-0003-4608-752X>)

Editora/Editor Associada/Associado Júnior

Karolayne Costa Rodrigues de Lima (<https://orcid.org/0000-0002-6311-8482>)

Editora/Editor de Texto Responsável

Suzana Zulpo (<https://orcid.org/0000-0003-2440-9938>)

Seção de Apoio às Publicações Científicas Periódicas - Sistema de Bibliotecas (SiBi) da Universidade Federal do Paraná - UFPR

Editora/Editor de Layout

Tânia Mara Mazon Barreto (<https://orcid.org/0000-0002-0314-4486>)