

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

JÚLIO RICARDO BURLAMAQUI DOS SANTOS

Gestão de Dados de Referência: Um estudo de caso para a Universidade
Federal do Rio de Janeiro

RIO DE JANEIRO

2025

JÚLIO RICARDO BURLAMAQUI DOS SANTOS

Gestão de Dados de Referência: Um estudo de caso para a Universidade
Federal do Rio de Janeiro

Trabalho de conclusão de curso
de graduação apresentado ao
Instituto de Computação da
Universidade Federal do Rio de
Janeiro como parte dos
requisitos para obtenção do grau
de Bacharel em Ciência da
Computação.

Orientadora: Prof. Maria Luiza
Machado Campos

RIO DE JANEIRO

2025

CIP - Catalogação na Publicação

B961g Burlamaqui dos Santos, Júlio Ricardo
Gestão de Dados de Referência: Um estudo de caso
para a Universidade Federal do Rio de Janeiro /
Júlio Ricardo Burlamaqui dos Santos. -- Rio de
Janeiro, 2025.
73 f.

Orientadora: Maria Luiza Machado Campos.
Trabalho de conclusão de curso (graduação) -
Universidade Federal do Rio de Janeiro, Instituto
de Computação, Bacharel em Ciência da Computação,
2025.

1. gestão de dados. 2. dados de referência. 3.
dados mestres. 4. Universidade Federal do Rio de
Janeiro. 5. administração pública. I. Machado
Campos, Maria Luiza, orient. II. Título.

JÚLIO RICARDO BURLAMAQUI DOS SANTOS

Gestão de Dados de Referência: Um estudo de caso para a Universidade
Federal do Rio de Janeiro

Trabalho de conclusão de curso de
graduação apresentado ao Instituto
de Computação da Universidade
Federal do Rio de Janeiro como parte
dos requisitos para obtenção do grau
de Bacharel em Ciência da
Computação.

Aprovado em ____ de _____ de _____.

BANCA EXAMINADORA:

Maria Luiza Machado Campos, Doutora (UFRJ)

Sérgio Manuel Serra da Cruz, Doutor (UFRJ)

Luiz Cláudio Coutinho Cruz, Técnico Administrativo (PR2/UFRJ)

Dedico esse trabalho a todos os servidores públicos que atuam em benefício da sociedade, promovendo os princípios constitucionais da administração pública.

AGRADECIMENTOS

Gostaria de agradecer aos meus pais, Fábio Santos e Carla Burlamaqui, que sempre me apoiaram de todas as maneiras possíveis. Vós sois para mim como a rede de proteção de um funâmbulo: convosco, nada temo. Agradeço à Cecília e Diana Bulcão por terem me recebido de braços abertos assim que cheguei no Rio de Janeiro, quão felizando fui de ter sido acolhido durante os primeiros momentos tão longe de casa.

À minha orientadora, professora Maria Luiza Campos, obrigado por me fazer ver a importância da gestão de dados e da ontologia, assim como por todas as oportunidades que se desdobraram diante de mim e pelos anos trabalhados juntos; torço para que mais oportunidades como estas não cessem de surgir, e que seu nome não deixe de aparecer nos meus trabalhos futuros.

Meus agradecimentos à FAPERJ pelo financiamento no projeto Conecta UFRJ, que possibilitou meu contato inicial com a gestão de dados, e desenvolveu meu interesse nesta área. Este fomento viabilizou trocas importantes de conhecimento na minha trajetória acadêmica. Gostaria de agradecer também aos funcionários da UFRJ que se dispuseram a achar um tempo em suas lotadas agendas para trocar, de coração aberto, parte de seu conhecimento, fundamental ao meu trabalho.

Agradeço ainda ao professor Hugo Nóbrega por ter aguçado meu interesse pela Lógica e Computabilidade, bem como por ter sido um grande professor e tutor – me ajudando a navegar durante a graduação com primazia. E para a professora Juliana Valério por ter sido uma coordenadora que sempre se mostrou disponível para me auxiliar, com bastante entusiasmo, nas mais diversas situações.

Aos meus amigos, que tornaram o Rio de Janeiro meu segundo lar, todo meu amor fraternal. É inviável citar todos que compartilhei risadas durante a graduação, mas, agradeço, nomeadamente, a Felipe e Luca Zingali, Pedro Bessa, Arthur Newlands, João Pinto, Natasha Costa, Helena Ribeiro, Matheus Schlittler, Carol Carvalho, Lucas Laranja, Igor Murras, Daniel Levacov e Lucas Tanaka.

Por último, obrigado aos meus amigos extra marinhos que conheci durante o intercâmbio, época áurea da minha vida em todos os sentidos, especificamente Pedro Rodrigues, Gustavo Marcondes, Hyeri Kim, Tai-yun Yu, Daniel Arellano, Jaehyong Kim, José Otero e Andrés Velez. Danke schön!

“Os dados são o novo petróleo. Eles são valiosos, mas precisam ser refinados para se tornarem úteis.”

- Clive Humby

RESUMO

Este trabalho se propõe a examinar o grau de maturidade da gestão de dados mestre e de referência da Universidade Federal do Rio de Janeiro. Tendo como ponto de partida fundamentos e diretrizes do DAMA Data Management Body of Knowledge (DMBoK), amplamente utilizado em diversos sortes de organizações públicas e privadas. A monografia envolveu um levantamento de conceitos e caracterização destes dois tipos de dados. Por meio de análise documental e entrevistas estruturadas com servidores de sistemas da UFRJ, o presente trabalho apresenta uma visão aplicada dos conhecimentos do estado da arte levantados, bem como aqueles utilizados com sucesso em entidades do primeiro setor. Como contribuição, espera-se oferecer motivação e propostas de alternativas para mitigar os problemas de gestão dos dados de referência da UFRJ, através de uma cartilha de diretrizes e políticas recomendáveis para a melhoria da qualidade desses dados na universidade.

Palavras-chave: gestão de dados, dados de referência, dados mestre, Universidade Federal do Rio de Janeiro, administração pública.

ABSTRACT

This work aims to examine the maturity level of master and reference data management at the Federal University of Rio de Janeiro. It is grounded in the principles and guidelines of the DAMA Data Management Body of Knowledge (DMBoK), which is widely recognized in various public and private organizations. The research involved exploring concepts and characterizing these two data types. Through document analysis and structured interviews with UFRJ systems staff, this work provides an applied view of the state-of-the-art knowledge developed, as well as those successfully used in entities of the public sector. As a contribution, it is expected to offer proposals for alternatives to mitigate the management issues of reference data at UFRJ, through a guideline booklet of recommended policies to improve the quality of this data at UFRJ.

Keywords: data management, reference data, master data, Federal University of Rio de Janeiro, public administration.

LISTA DE ILUSTRAÇÕES

Figura 1 – Áreas de conhecimento na Gestão de Dados, segundo o DMBok

Figura 2 – Etapas para construção de um PDA

Figura 3 – Painel do Governo Federal das bases de dados abertas pela UFRJ

Figura 4 – Página dos dados abertos da UFRJ

Figura 5 – Página de dados abertos do CHS-UFRJ

LISTA DE WEBSITES

- ¹ ufrj.br – Página principal da UFRJ
- ² ufrj.br/pda – Página do Plano de dados Abertos da UFRJ
- ³ dados.gov.br – Repositório de Dados Abertos do Governo Federal
- ⁴ dadosabertos.ufrj.br – Repositório de Dados Abertos do Governo Federal
- ⁵ dados.unicamp.br – Página do Escritório de Dados e Apoio à Transformação da Unicamp
- ⁶ github.com/JulioBurlamaqui/TCC – Repositório pessoal do Github dedicado a este trabalho
- ⁷ github.com/Grupo-GRECO/Guias_Gestao_Dados – Repositório do Grupo de Engenharia do Conhecimento (GRECO) dedicado à promoção de gestão de dados na UFRJ

LISTA DE ABREVIATURAS E SIGLAS:

API – *Application Programming Interface* (Interface de Programação de Aplicações)

APF – Administração Pública Federal

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CDO – *Chief Data Officer* (Chefe-Executivo de Dados, ou Diretor de Dados)

CenSup – Censo da Educação Superior

CGU – Controladoria Geral da União

CHS – Complexo Hospitalar e da Saúde

CKAN – *Comprehensive Knowledge Archive Network* (Rede Abrangente de Arquivos de Conhecimento)

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico

CSV – *Comma-separated values* (Valores separados por vírgula)

DAMA – *Data Management Association* (Associação de Gestão de Dados)

DMBoK – *Data Management Body of Knowledge* (Corpo de Conhecimento de Gestão de Dados)

EDAT – Escritório de Dados e Apoio à Transformação

ESB – *Enterprise Service Bus* (Serviço de Barramento Corporativo)

ETL – *Extract, Transform, Load* (Extrair, Transformar, Carregar)

FAIR – *Findable, Accessible, Interoperable, and Reusable* (Encontrável, Acessível, Interoperável e Reutilizável)

FAPERJ – Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro

GDF – Governo do Distrito Federal

GRECO – Grupo de Engenharia do Conhecimento

IBM – International Business Machines

iGG – Índice integrado de Gestão e Governança Pública

iPaaS – *Integration Platform as a Service* (Plataforma de Integração como um Serviço)

LAI – Lei de Acesso à Informação (nº. 12.527/2011)

LGPD – Lei Geral de Proteção aos Dados (nº. 13.709/2018)

MDM – *Master Data Management* (Gestão de Dados Mestre)

NCE - Núcleo de Computação Eletrônica

ODS - Objetivos para o Desenvolvimento Sustentável

ONS - Operador Nacional do Sistema Elétrico

ONU - Organização das Nações Unidas

OSP-DF - Órgão de Segurança Pública do Distrito Federal

PDA - Plano de Dados Abertos

PNIFE - Plataforma Nacional de Infraestrutura de Pesquisa

PR1 - Pró-Reitoria de Graduação

PR2 - Pró-Reitoria de Pós-Graduação e Pesquisa

PR4 - Pró-Reitoria de Pessoal

PR5 - Pró-Reitoria de Extensão

PR6 - Pró-Reitoria de Gestão e Governança

PR7 - Pró-Reitoria de Políticas Estudantis

P&D - Pesquisa e Desenvolvimento

RDM - *Reference Data Management* (Gestão de Dados de Referência)

SERPRO - Serviço Federal de Processamento de Dados

SIGA - Sistema Integrado de Gestão Acadêmica

SiSU - Sistema de Seleção Unificada

SQL - *Structured Query Language* (Linguagem de Consulta Estruturada)

SEI - Sistema Eletrônico de Informações

TCU - Tribunal de Contas da União

TI - Tecnologia de Informação

TIC - Superintendência Geral de Tecnologia da Informação e Comunicações

UFMG - Universidade Federal de Minas Gerais

UFRJ - Universidade Federal do Rio de Janeiro

UNICAMP - Universidade Estadual de Campinas

SUMÁRIO

1. INTRODUÇÃO.....	16
1.1 ESTRUTURA.....	16
1.2 MOTIVAÇÃO.....	16
1.3 JUSTIFICATIVA.....	18
1.4 OBJETIVOS.....	20
1.4.1 GERAL.....	20
1.4.2 ESPECÍFICOS.....	20
1.5 METODOLOGIA.....	21
2. REVISÃO DE LITERATURA.....	23
2.1 CONCEITOS FUNDAMENTAIS SOBRE GESTÃO E GOVERNANÇA DE DADOS.....	24
2.1.1 FRAMEWORK DO DATA MANAGEMENT BODY OF KNOWLEDGE.....	24
2.1.2 GESTÃO DE DADOS MESTRES E DE REFERÊNCIA.....	28
2.2 PRÁTICAS DA GESTÃO DE DADOS MESTRES E DE REFERÊNCIA.....	31
2.3 TRABALHOS RELACIONADOS.....	33
3. SITUAÇÃO DA UFRJ.....	35
3.1 ENTREVISTA.....	36
3.1.1 PRÓ-REITORIA DE PÓS-GRADUAÇÃO.....	36
3.1.2 SISTEMA INTEGRADO DE GESTÃO ACADÊMICA.....	40
3.2 ANÁLISE DOCUMENTAL.....	43
3.2.1 PLANO DE DADOS ABERTOS.....	43
3.2.2 ANÁLISE DO TRIBUNAL DE CONTAS DA UNIÃO.....	48
3.3 SILOS DE INFORMAÇÃO.....	50
4. RECOMENDAÇÕES COM BASE NO ESTADO DA ARTE PARA UFRJ.....	52
4.1 MATURIDADE DE GESTÃO DE DADOS.....	52
4.2 RECOMENDAÇÕES.....	53
4.2.1 PAPÉIS EM RELAÇÃO AOS DADOS.....	53
4.2.2 ARQUITETURA DE GESTÃO DE DADOS MESTRES.....	54

4.2.3 GESTÃO DE DADOS DE REFERÊNCIA.....	56
4.2.4 GESTÃO DE METADADOS.....	58
4.3 EXEMPLOS BEM SUCEDIDOS.....	58
4.3.1 PRIMEIRO SETOR.....	58
4.3.2 SEGUNDO SETOR.....	60
5. CONCLUSÃO.....	62
5.1. CARTILHA DE DIRETRIZES DE DADOS DE REFERÊNCIA.....	62
5.2 LIMITAÇÕES.....	63
5.3. TRABALHOS FUTUROS.....	64
REFERÊNCIAS BIBLIOGRÁFICAS.....	66
APÊNDICE.....	71
QUESTIONÁRIO.....	71

1. INTRODUÇÃO

1.1 MOTIVAÇÃO

A garantia de dados públicos transparentes configura-se como um pilar fundamental para o fortalecimento de instituições democráticas e a promoção de governança eficaz, alinhando-se diretamente aos Objetivos de Desenvolvimento Sustentável (ODS) da Agenda 2030 da Organização das Nações Unidas (ONU).

Especificamente, duas podem ser citadas: a ODS 16.6 enfatiza a necessidade de instituições responsáveis e transparentes, as quais dependem da disponibilização e do acesso a informações precisas e verificáveis, assegurando a prestação de contas. Já a ODS 17.18 destaca a importância do fomento a capacitações técnicas e institucionais para a produção de dados desagregados e confiáveis, reforçando a sinergia entre a transparência informacional e o desenvolvimento sustentável (ONU, 2015).

No cenário nacional, temos o artigo 5º, inciso XIV, da Constituição Federal de 1988, que assegura a todos o direito de acesso à informação, garantia que se desdobra no inciso XXXIII do mesmo dispositivo, o qual explicita a obrigatoriedade de órgãos públicos disponibilizarem dados de interesse coletivo, ambos resguardados os casos de sigilo legal (BRASIL, 1988).

Ademais, o artigo 216, §2º, da Carta Magna, ao vincular a gestão do patrimônio cultural brasileiro à publicidade de dados, reforça a responsabilidade das instituições públicas em democratizar informações sobre seu acervo histórico, projetos e atividades de pesquisa, práticas que dependem de infraestruturas de dados robustas para evitar fragmentação ou perda de registros (BRASIL, 1988).

A Lei de Acesso à Informação (nº. 12.527/2011), por sua vez, consolida esse arcabouço normativo ao estabelecer mecanismos de transparência ativa e passiva, exigindo que as entidades da administração pública mantenham portais atualizados, com dados desagregados e acessíveis, em formatos interoperáveis (BRASIL, 2011).

Conforme o Brasil foi compreendendo a importância da gestão e governança de dados, surgiu a demanda por aplicá-la na administração pública, entendendo-se também sua utilidade nas instituições federais de ensino superior (Santos, Souza, 2022).

A Universidade Federal do Rio de Janeiro (UFRJ), enquanto autarquia pública de ensino e pesquisa, assume papel estratégico na concretização dos princípios constitucionais que regem a transparência e o acesso à informação, alinhando suas práticas institucionais não apenas às demandas contemporâneas de governança, mas também ao ordenamento jurídico brasileiro.

As universidades, ao implementarem repositórios digitais e sistemas de informação integrados, como o Portal da Transparência institucional e plataformas de dados abertos, não apenas cumprem exigências legais, mas também promovem *accountability* (responsabilidade), permitindo que a sociedade civil monitore recursos públicos, processos decisórios e impactos de políticas acadêmicas.

Ao buscar essa sinergia entre gestão técnica de dados e conformidade jurídica, a UFRJ poderia operar como modelo de institucionalidade democrática, na qual a precisão dos dados assegura a confiabilidade de indicadores de desempenho, e sua boa documentação facilita a contextualização necessária para o exercício pleno de direitos por cidadãos e pesquisadores.

Assim, as universidades podem transcender sua função educativa, posicionando-se como agentes ativos na construção de uma cultura de transparência. Dessa forma, a convergência entre governança transparente e infraestrutura de dados robusta revela-se não apenas estratégica, mas indispensável para a materialização de um desenvolvimento socioeconômico inclusivo e participativo.

1.2 JUSTIFICATIVA

Historicamente, a fragmentação e multiplicidade de sistemas desenvolvidos de forma independente foram fatores que contribuíram para a redundância e consequente inconsistência de dados nas organizações. Esse problema se manifesta através dos seguintes desafios: necessidade de constante sincronização entre setores e empresas; alta complexidade e esforço para manter e desenvolver sistemas que se comuniquem; e, necessidade crescente de infraestruturas robustas para armazenar e processar grandes volumes de dados (INMON, STRAUSS, NEUSCHLOSS, 2010).

Conforme esses dados tornaram-se percebidos nas empresas de maneira ubíqua, houve o recrudescimento da valorização dos dados como um ativo essencial no apoio a decisões. Dessa maneira, elas sentem-se compelidas a unificar esses dados espalhados por seus sistemas em uma “versão única dos dados” (Khatri, Brown, 2010). É nesse cenário que se torna cada vez mais crítica a demanda por gestão de dados e, como parte desta, a caracterização de um tratamento diferenciado para dados mestres e de referência.

Os dados mestre são aqueles que representam informações precisas e fundamentais para as transações críticas e o funcionamento de uma organização. Em contrapartida, os dados de referência são conjuntos de valores que fornecem informações adicionais para que uma organização opere de forma mais eficiente, definindo e categorizando outros dados (DAMA, 2017).

Além disso, de acordo com Hildebrand *et al.* (2025), o número de exigências regulatórias e legais que as empresas e setor público devem seguir está crescendo continuamente, necessitando também ser capazes de provar que estão em conformidade com essas normas. Isso envolve a capacidade de armazenamento, coleta e disponibilização de dados específicos.

Nesse ensejo, a UFRJ, ao adotar sistemas de gestão baseados em dados mestres, como registros acadêmicos unificados e cadastros institucionais; e em

dados de referência, como nomenclaturas e fontes padronizadas, fortalecerá a integridade, a consistência e a rastreabilidade das informações, assegurando conformidade com tais preceitos legais.

Ainda, de acordo com Barata (2015), o sucesso de pesquisas acadêmicas depende da qualidade dos dados. Essa qualidade é garantida por uma governança de dados que estabelece responsabilidades, métricas e procedimentos claros para a coleta, armazenamento e recuperação das informações, permitindo que as instituições cumpram seus objetivos.

Rêgo (2013) salienta alguns dos ganhos que a gestão de dados traz para as organizações, ressaltando porém que muitos desses ganhos são intangíveis e difíceis de serem mensurados, mas sendo notadamente sentidos nas operações cotidianas da instituição. Ele cita exemplos, como a mudança de cultura, na qual os dados e informações são reconhecidos como ativos estratégicos pelos funcionários e há a criação do hábito do uso de indicadores de processos, qualidade e desempenho dos dados e informações.

Essas mudanças levam à redução do tempo, riscos, falhas e custos no desenvolvimento de novos sistemas, bem como na quantidade de informações redundantes. Por fim, isso resulta em um aumento da produtividade, promovendo um vocabulário unificado dos dados em toda a organização (RÊGO, 2013).

1.3 OBJETIVOS

1.3.1 GERAL

O objetivo deste trabalho é compreender o grau de maturidade da gestão e governança de dados, em específico, de dados mestres e de referência, na Universidade Federal do Rio de Janeiro, bem como oferecer propostas de diretrizes e políticas recomendáveis para a melhora do processo de gestão e governança desses dados na universidade.

No entanto, não nos propomos neste trabalho a realizar uma avaliação formal de maturidade, na linha do Capability Maturity Model (CMM) aplicado à gestão de dados, pela amplitude e complexidade envolvidos nesse tipo de processo (CCMI Institute, 2019).

1.3.2 ESPECÍFICOS

O primeiro objetivo específico é o de realizar um levantamento de conceitos e caracterizar a gestão e governança de dados mestres e de referência, principalmente, sob a ótica do *Data Management Body of Knowledge* (DAMA-DMBoK) da *Data Management Association* (DAMA), e demais conceitos correlatos.

Em seguida, nos propomos a analisar a situação atual da Universidade Federal do Rio de Janeiro em relação à gestão de dados por meio de entrevistas estruturadas com servidores de sistemas internos; e análise documental dos Planos de Dados Abertos (PDA) da UFRJ e do relatório do Tribunal de Contas da União (TCU) sobre a transparência desta instituição.

Finalmente, o trabalho objetiva a proposição de alternativas e diretrizes para mitigar os problemas de gestão de dados de referência da UFRJ, com base nos conhecimentos do estado da arte e exemplos de sucesso noutras instituições, por meio da elaboração de um guia.

1.4 METODOLOGIA

Além da revisão bibliográfica, esse trabalho também traz um estudo de caso sobre alguns sistemas da Universidade Federal do Rio de Janeiro, no qual recolhemos dados a fim de entender seus desafios e obstáculos na sincronização entre seus dados mestres e referência, bem como um levantamento busca em outras instituições por iniciativas que busquem implementar a gestão de dados

eficazmente. Os métodos escolhidos para coleta das informações do estudo de caso foram entrevista estruturada de servidores e análise de documentos importantes que descrevem o funcionamento e relação entre alguns dos sistemas de bancos de dados da UFRJ.

Essa decisão possui respaldo em outros trabalhos acadêmicos que se propuseram a investigar a maturidade da gestão de dados de demais entidades públicas e privadas, como Barata (2015), que compara a maturidade da gestão de dados de três organizações brasileiras de grande porte por meio da análise de entrevistas feitas pelo autor; ou Ribeiro (2014), que realizou uma pesquisa sobre a gestão de dados corporativos no SERPRO (Serviço Federal de Processamento de Dados) por meio de entrevistas e análises documentais.

De acordo com Duarte (2004), para o êxito de uma entrevista acadêmica, é necessário que o elaborador tenha claramente seus objetivos definidos e assimilados, além de que conheça com alguma profundidade o tema sobre o qual pretende abordar na entrevista, se valendo de uma revisão bibliográfica acerca do objeto de estudo previamente, por exemplo. A autora ainda menciona algumas boas práticas que devem ser seguidas ao realizar uma entrevista, como segurança e autoconfiança, pontualidade e ensaios da entrevista antes de executá-la.

Flick (2010) salienta que as perguntas devem ao mesmo tempo ampliar o escopo, de maneira que não enviesem a resposta e permita ao entrevistado compartilhar sua visão, e devem ser específicas o suficiente para manter a estrutura e coesão, sem perder o foco.

Por se tratar de um estudo de caso informal, especificamente voltado para levantamento de situações exemplo que tipicamente ocorrem na UFRJ e outras universidades, buscou-se entrevistar pessoas fortemente envolvidas com esses sistemas, mas não necessariamente seus responsáveis. Espera-se com isso, dar andamento à discussão e percepção da necessidade da implantação de uma política de governança e gestão de dados na universidade, contribuindo para um

cultura orientada a dados e para a identificação de responsabilidades e papéis associados.

Por sua vez, segundo Júnior *et al* (2021), a pesquisa documental é um método que se baseia exclusivamente em documentos – isto é, declarações escritas, oficialmente reconhecidas – para obter informações e compreender fenômenos, utilizando técnicas de captação, análise e interpretação de um conjunto heterogêneo de documentos.

Kripka, Scheller e Bonotto (2015) nos conduzem a delimitar o universo que pretendemos investigar. Elas afirmam que a escolha do documento deve ser feita tendo em mente o problema específico que se visa a solucionar. Sendo assim, as perguntas que o pesquisador formula ao documento são tão importantes quanto o próprio documento.

1.5 ESTRUTURA

Este trabalho se apresenta organizado em cinco capítulos. O primeiro deles compreende a apresentação da motivação que instigou a luz desta monografia, seus objetivos geral e específicos a serem cumpridos, as razões que justificam a realização dele, e metodologia conduzida. No capítulo seguinte, foram compilados os conhecimentos presentes no estado da arte da gestão de dados, suas práticas, bem como foram discutidos alguns trabalhos relacionados a este e suas contribuições conceituais. No terceiro capítulo, encontra-se o cerne do trabalho, isto é, o resultado das entrevistas com os servidores de sistemas da UFRJ, e de funcionários de outra entidade da administração indireta; e, ainda, da análise documental. No penúltimo capítulo, foram discutidas as implicações destes achados, na visão da literatura levantada. Na conclusão, resumimos o trabalho, propomos um guia de boas diretrizes e políticas de gestão de dados de referência e refletimos as limitações do trabalho.

2. REVISÃO DE LITERATURA

Como todo conceito científico, existem várias definições aplicáveis aos conceitos aqui estudados e debatidos. Portanto, é prudente primeiro examinar de que maneira essas definições se sobrepõem, se complementam ou mesmo se conflitam para poder traçar estratégias e estudar mecanismos ou ferramentas que visam a apoiar a gestão de dados.

Usualmente, qualquer estudo sobre gestão de dados inicia-se com o *Data Management Body of Knowledge* (DAMA-DMBoK), por constituir-se como a principal referência na área. O DMBoK consiste em um compêndio abrangente de temas considerados fundamentais para a gestão de dados (DAMA, 2017). Trata-se de uma obra resultante dos esforços da *Data Management Association* (DAMA, Associação de Gestão dos Dados), uma organização internacional sem fins lucrativos dedicada à promoção e ao avanço do conhecimento em gestão de dados.

Neste livro, temos uma definição de gestão de dados que a descreve com base nas atividades desempenhadas por ela. Assim, temos que a gestão de dados é o “desenvolvimento, execução, e supervisão dos planos, políticas, programas e práticas que entregam, controlam, protegem e melhoram o valor dos ativos de dados e informações durante seu ciclo de vida” (DAMA, 2017, p. 17, tradução nossa).

No entanto, essa é uma definição ampla e abrangente e, por isso, o DMBoK aprofunda essa ideia, nos dando um rol exemplificativo de atribuições da gestão de dados. Apesar de sua forte relação com a Computação, ela não é constituída apenas de habilidades técnicas, mas também apresenta características multidisciplinares, requerendo conjuntamente habilidades oriundas de outras áreas do saber (DAMA, 2017).

Embora acreditemos que definir o conceito de dado seja desnecessário – por sua ubiquidade – e, ao mesmo tempo, desafiador, por sua qualidade

polissêmica, gostaríamos de trazer a distinção entre dado e fato, que por seu turno, é distinto de informação. Dados são, simultaneamente, uma interpretação do objeto que ele almeja representar, e um objeto a ser interpretado (SEBASTIAN-COLEMAN, 2013). Isso significa que eles não representam a verdade ou a essência de um objeto – o fato em si – sendo, na realidade, uma forma de representação, que é carente de contexto. Contexto é, pois, o sistema de representações que permite a obtenção da informação através do dado.

Essa natureza interpretativa dos dados tem como consequência a multiplicidade das representações. Pessoas diferentes fazem escolhas distintas sobre como representar conceitos, criando assim várias maneiras corretas de expressar as mesmas ideias. Isso faz com que os dados assumam formas variadas. Um exemplo simples são as diversas maneiras de representar datas de calendário, mesmo sendo um conceito com uma definição amplamente aceita. Conceitos mais complexos, como "aluno" ou "laboratório", tornam o processo de representação e gerenciamento ainda mais complexo, pois a granularidade e o nível de detalhe necessários nem sempre são óbvios (DAMA, 2017).

2.1 CONCEITOS FUNDAMENTAIS SOBRE GESTÃO E GOVERNANÇA DE DADOS

2.1.1 FRAMEWORK DO DATA MANAGEMENT BODY OF KNOWLEDGE

O DMBok estabelece um *framework* estruturado para a compreensão da gestão de dados, subdividindo-a em dez áreas distintas, conforme a Figura 1, cada uma delas abordada em um capítulo específico, escrito por autores diferentes. Dentre essas áreas, a que orientará a discussão aqui apresentada é a gestão de dados mestres e de referência.

No entanto, outras áreas adjacentes também têm relevância ao tópico tratado, a título da arquitetura de dados, que atua como a base, definindo o projeto e a estrutura dos sistemas que vão abrigar e gerenciar esses dados, alinhando a infraestrutura tecnológica aos objetivos de negócio. Além disso, a

gestão de metadados engloba a criação, manutenção e utilização de metadados, que contextualiza os dados mestres e de referência (DAMA, 2017).

Para que a visão unificada dos dados mestres seja útil, é fundamental uma boa gerência de qualidade de dados, que se concentra em garantir que esses dados sejam precisos, completos, consistentes e atualizados, bem como em gerar métricas para avaliar essas características. Assim, ela pode assegurar uma boa integração e interoperabilidade dos dados, que é a área responsável por descrever os processos técnicos de movimento e consolidação de informações entre diferentes sistemas e aplicações (Barbieri, Farinelli, 2013).

Todavia, as demais áreas também impactam direta e indiretamente na gestão de dados mestre e de referência, e foram, em medidas diferentes, exploradas no trabalho. Essas quatro subáreas mencionadas que se relacionam com a gestão de dados mestres e de referência são unidas e viabilizadas pela governança de dados. Ela está situada no centro da mandala do DMBok, conforme Figura 1, para representar como é o elo que mantém todas as demais disciplinas da gestão de dados.

O DMBok (2017) define a governança de dados como o exercício da autoridade e controle da gerência sobre o ativo dos dados. Sua função é a de garantir as melhores práticas e políticas possíveis na gestão de dados. O foco dela é o de entender profundamente as regras de negócio a fim de traçar estratégias sobre como os dados, processos e as pessoas devem ser conduzidos. Ainda, a governança de dados deve se preocupar com a conformidade às normas legais e métricas internas e externas.

Figura 1 – Áreas de conhecimento da Gestão de Dados, segundo o DMBok



Fonte: DAMA (2017), p. 36, tradução nossa

O DMBok tem uma visão mais descritiva e normativa, buscando padronizar a área e estabelecer uma partilha entre domínios. Para entendermos de quais maneiras esses domínios se relacionam, precisamos entender os conceitos fundamentais antes. Podemos enxergar essa gama de dados como uma escada, na qual os que estão embaixo sempre dão contexto a todos que estão acima. Assim, ascenderemos esta escada, entendendo cada conceito, que se relaciona com os demais.

O metadado é uma espécie de dado que contém informações técnicas sobre os dados em si, isso inclui, mas não é limitado por, regras de negócio, limitações e regra dos dados, estrutura lógica e física dos bancos de dados, representação conceitual dos dados, conexões entre atributos, características intrínsecas do

surgimento do dado, explicação dos conceitos e usos, entre outros. Note que todo tipo de dado é passível de carregar consigo metadados, incluindo metadados. Dessa forma, ele contextualiza todas as outras categorias de dados (DAMA, 2017).

O metadado possui um papel vital na gestão de dados, porque permite a avaliação na qualidade dos dados ao fornecer informações sobre quais são os dados que tem, onde e como eles estão armazenados, onde e como eles se originam, qual seu ciclo de vida, ou qual seu significado. O DMBok é categórico ao dizer que “sem metadados, uma organização não poderá gerir seus dados de forma alguma” (DAMA, 2017, p. 418, tradução nossa).

A gestão de metadados se relaciona com praticamente todas as demais áreas, ao viabilizar o processamento, integração, manutenção, auditoria, segurança e aferição de qualidade e modelagem de arquitetura (DAMA, 2017).

Precisamos alertar que apesar do DMBok falar extensivamente sobre metadados, os princípios FAIR (Findable, Accessible, Interoperable, and Reusable) não são mencionados em seu corpo. Isto se dá pelo fato da segunda edição do DMBok ter sido lançada em 2017, enquanto estes princípios que norteiam os dados abertos, embora concebidos em 2016, eram muito restritos à ciência aberta, somente ganhando atenção geral em 2018 (Wilkinson, 2016).

No degrau acima, encontramos o dado de referência, que o DMBok (2017) descreve como aquele que caracteriza e classifica outros dados do ponto de vista contextual para a organização, desse modo, sendo frequentemente representado por códigos e listagens. Os dados de referência desempenham um papel estratégico na melhoria da qualidade dos dados e na simplificação da integração e compartilhamento de informações. Ao estabelecer definições, classificações e relacionamentos comuns para elementos de dados por meio de códigos e valores predefinidos, os dados de referência garantem a consistência em todos os sistemas.

Subindo mais, encontramos os dados mestres, que são dados acerca de entidades essenciais ao negócio. Essas entidades são objetos do mundo real com a qual a organização se relaciona para concretização de suas ações (inclusive ela

mesma). Os dados mestres são usados para dar contexto aos dados transacionais. Seus formatos e valores possíveis tipicamente são definidos pelas regras de negócio da empresa, mas, no geral, podem cair em quatro categorias: pessoas (físicas ou não), produtos e serviços, estruturas burocráticas e localidades (DAMA, 2017).

A ideia chave para entender dados mestres na perspectiva do DMBok (2017) é a de “*golden record of truth*” (um registro áureo da verdade, numa tradução livre), isto é, uma representação única e soberana de uma determinada entidade. Os autores argumentam que é possível que a mesma entidade possua diferentes versões em diferentes setores de uma organização, e isso pode gerar conflitos, redundâncias e inconsistências. Dessa maneira, se faz mister que todas essas versões sejam consolidadas num único cadastro, o *golden record*.

No topo da escada situa-se o dado transacional. Uma falha do DMBok é usar este termo sem defini-lo. No entanto, podemos recorrer a outra fonte, que traz uma conceituação satisfatória dele. McGilvray nos entrega de forma bastante sucinta e eficaz: “Dados transacionais descrevem um evento ou transação interna ou externa que ocorre enquanto uma organização conduz seus negócios.” (MCGILVRAY, D., 2008, p. 42, tradução nossa).

2.1.2 GESTÃO DE DADOS MESTRES E DE REFERÊNCIA

A partir desses conceitos, conseguimos entender agora a visão do DAMA sobre gestão de dados mestres e de referência. Apesar de frequentemente tratadas como uma única disciplina devido ao objetivo comum de melhorar a qualidade e integração dos dados entre diferentes fontes, aplicações e tecnologias, essas duas áreas possuem definições e desafios distintos.

Desse modo, ele define dados mestres e de referência como dados que fornecem informações essenciais para transações da instituição, porém, o que difere um do outro é que dados de referência são usados para classificar ou categorizar outros dados, como nacionalidade, status, estado civil. Ao passo que

dados mestres representam as entidades fundamentais para essas transações (DAMA, 2017).

A gestão de dados mestres (MDM - *Master Data Management*, do inglês) visa a criar, integrar, manter e usar dados mestres dentro da instituição, estabelecendo e mantendo os registros áureos e assegurando seu uso pela instituição como um todo. A gestão de dados de referência (RDM - *Reference Data Management*, do inglês) foca em criar e manter uma lista de valores válidos para os campos, além de explicar seus significados de uma maneira mais compreensível e organizar as relações entre os dados, especialmente suas hierarquias. Veja que estes não são um trabalho pontual, mas que exigem esforços contínuos (DAMA, 2017). Esse caráter de conservação continuada é reforçado por outros autores, como Bill Inmon *et al.* (2010).

A ausência de uma gestão formal de dados de referência, todavia, implica que diferentes sistemas e departamentos de uma organização podem empregar terminologias e classificações distintas para o mesmo conceito. Por exemplo, em uma instituição de ensino, "aluno ativo" pode ter nuances diferentes entre os sistemas de matrícula, biblioteca ou assistência estudantil, levando a inconsistências nos dados mestres de um aluno.

Essa falta de padronização dificulta a integração de sistemas e a geração de relatórios unificados - conferindo díspares resultados para uma mesma métrica, bem como no comprometimento da tomada de decisões, que se baseiam em informações desatualizadas e conflitantes. Sem contar com a possibilidade de retrabalho, consumindo tempo e recursos.

A gestão de dados de referência, portanto, atua como o alicerce para a interoperabilidade interna e externa, permitindo que todos tenham o mesmo léxico e um linguajar comum. Uma gestão eficiente dos dados de referência permite evitar problemas decorrentes de informações conflitantes e redundantes, que são comuns quando diferentes departamentos desenvolvem suas próprias versões de dados.

A perspectiva de McGilvray (2021) é bastante similar à do DMBok, porém, sua abordagem se fundamenta no que ela chama de “categorias de dados”, ou seja, agrupamentos de dados que possuem certas características ou atributos. Ela argumenta que entender as relações e dependências entre esses grupos nos ajuda a direcionar os esforços de melhorar a qualidade dos dados.

Assim, ela é mais enfática e explícita em dizer que os dados de referência podem vir de fontes internas e externas à instituição e que esses agrupamentos servem não só para dar contexto aos dados transacionais, mas também aos dados mestres (McGilvray, 2021). Podemos ilustrar este conceito com um exemplo: uma loja (dado mestre) possui um endereço (dado de referência), e um pedido nela (dado transacional) de um produto (dado mestre) possui um *status* (dado de referência).

A autora traz uma noção bastante simples mas bem eficiente em explicar dados mestres: “*são os substantivos que usamos para fazer as ações*” (MCGILVRAY, D., 2021, p. 41, tradução nossa, grifo da autora). Com isso, ela retoma aquela ideia do DAMA de que os dados mestres são as entidades que se relacionam com a organização e dá contexto para os dados transacionais, trazendo exemplos bastante ilustrativos, como clientes, fornecedores, colaboradores, contas, locais, produtos, entre outros (McGilvray, 2021). Completando sua analogia com a gramática sobre dados mestres, os dados transacionais seriam os verbos, a ação que está sendo feita.

Um conceito idêntico ao *golden record* também é levantado por McGilvray (McGilvray, 2021), embora tenha outro nome: registro mestre (*master record*, em inglês). Apesar de áureo (*golden*, do inglês) traduzir a imagem de precioso e inequívoco, o nome *master record* nos parece mais apropriado pois se relaciona diretamente com a ideia de dados mestres, e portanto será o nome utilizado daqui para frente neste projeto.

Outro conceito muito relacionado com registro mestre e que é fundamental para uma boa gestão de dados mestres é o chamado sistema de registros. Ele deve ser um sistema oficial no qual os dados são criados nele – ou copiados de

outro - e mantidos, respeitando uma série de regras de negócio. Seu sentido teleológico é o de servir como uma autoridade absoluta respectiva a um determinado domínio de dados, resolvendo inconsistências, se houver. Deve ser a partir dele que os usuários de dados os consomem (DAMA, 2017).

2.2 PRÁTICAS DA GESTÃO DE DADOS MESTRES E DE REFERÊNCIA

O DMBOK (2017) cita alguns do que ele chama de “princípios” da gestão de dados. Alguns são bem intuitivos pela própria natureza e objetivo da MDM e RDM, como que devem estar aptos a serem compartilhados e interoperados dentro, e potencialmente fora, da organização; ou que seus registros devem ser sempre replicados através do sistema de registros e estes devem ser a versão soberana do dado; ou ainda que a qualidade deles deve ser monitorada continuamente.

Todavia, outros são não triviais ou levam a políticas interessantes de geri-los. Por exemplo, nenhum setor ou sistema em específico terá propriedade exclusiva aos dados, porque eles são entendidos para serem compartilhados por toda a organização, assim sendo, um ativo da organização como um todo. Apesar disso, devemos ter alguma entidade que seja responsável pelos dados, se não um setor ou sistema. Por isso, outro princípio é a instituição formal de alguns papéis hierárquicos que lidam com os dados, investidos de poderes-deveres.

O primeiro deles é a figura do proprietário do dado, aquele responsável pela gestão, controle e manutenção de um ativo de dados dentro de um domínio específico. Ele definirá as regras de acesso, uso e modificação do conjunto de dados, bem como terá decisão final para deliberar sobre assuntos que competem ao conjunto de dados. Essa atribuição bem definida de competência não é uma mera formalidade, já que será incumbido a ele o dever de garantir conformidade às regulações, além de revisar e atualizar as políticas de uso e manutenção. Note que ele é dono dos dados no sentido de ser responsável, não de possuir o ativo.

O proprietário do dado deve subordinação a apenas uma entidade, o diretor dos dados (CDO, *chief data officer*, do inglês) que corresponde a maior liderança no que tange aos dados. Suas atribuições são as de definir estratégias gerais para gestão e governança dos dados e promover uma cultura consciente do manejo deles, bem como garantir a sua qualidade e conformidade às legislações federais, como Lei Geral de Proteção dos Dados (LGPD) e a Lei Acesso à Informação (LAI). Ainda, deve estar antenado às inovações tecnológicas e burocráticas, garantindo modernidade e eficiência à entidade.

Subordinado ao proprietário dos dados está o curador dos dados, seu papel é o da manutenção da cultura de boas práticas de gestão dos dados. Suas atribuições se materializam em manter o catálogo de dados e metadados atualizado, classificar níveis de acesso, garantir a proteção de dados pessoais, auxiliar na compreensão e melhoria do uso de dados, monitorar o ciclo de vida dos dados, incentivar o reuso, garantir a adoção de registros de referência e gerenciar a qualidade dos dados. Eles são essenciais para garantir que os dados estejam limpos, consistentes, precisos, atualizados e compartilháveis.

Abaixo deste, estão os guardiões (ou zeladores) dos dados, que se responsabilizam pela implementação técnica e operacional das políticas de dados, incluindo segurança, armazenamento, *backup* e recuperação de dados. Eles promovem os recursos tecnológicos, fazendo a manutenção dos dados em ambientes gerenciados. São importantes para pôr em prática a execução das políticas, garantindo a qualidade e consistência dos dados à nível operacional. Eles atuam como *proxy* dos cargos que estão acima deles, manifestando suas diretrizes (DAMA, 2017).

O último princípio evidencia a característica volátil das informações, que muda através do tempo. Embora elas possam mudar, a empresa deve estar atenta a sempre atualizar os dados para ter um entendimento sempre acurado e corrente dos dados. Devido ao fato de os dados de referência terem uma natureza classificatória, seus valores não são tão variáveis quanto os dos dados mestres, que abrangem valores mais diversos, podendo incluir informações como nomes,

endereços, datas ou valores monetários. Por isso, alguns autores recomendam que a gestão de dados de referência restrinja as opções de classificação a valores predefinidos – afinal, não é comum surgirem novos estados civis com frequência – em vez de permitir que os usuários insiram dados livremente (DAMA, 2017), (MCGILVRAY, 2021), (INMON, *et al.*, 2010).

Porém, é possível que haja alteração nos dados de referência, e quando eles são alterados, deve haver um processo bem definido para tal: as mudanças devem ser comunicadas e aprovadas antes de serem implementadas. Em convergência, quando os dados mestres são alterados, as alterações devem ser bem documentadas e facilmente reversíveis. Essas práticas garantem maior consistência e qualidade dos dados, facilitando sua análise e uso no futuro (DAMA, 2017).

2.3 TRABALHOS RELACIONADOS

Em sua dissertação de mestrado, Silva (2023) analisa a gestão de dados, incluindo todas as suas áreas, dos órgãos de Segurança Pública do Distrito Federal (OSP-DF) através da análise de doze documentos governamentais, tanto da Administração Pública Federal (APF), quanto do Governo do Distrito Federal (GDF) compreendidos entre os anos de 2017 e 2022. O trabalho apresenta o desenvolvimento de um *framework* como proposta para mitigar os problemas e desafios apresentados durante o corpo de seu trabalho com base na revisão de literatura. Este *framework* é construído em quatro áreas: governança, qualidade, gestão e *compliance* (conformidade).

Ambas, revisão de literatura e base metodológica, serviram de ponto de partida para este trabalho, principalmente a abordagem de identificação e categorização de mecanismos de governança, em especial, ao considerar que os dois versam sobre temas correlatos. Além disso, o *framework* desenvolvido ao final aborda e levanta pontos que nortearam a confecção da cartilha entregue

juntamente com esta monografia, em particular a definição de responsabilidades e padrões de qualidade.

Outro artigo que impulsionou este trabalho foi o do Carlos Barbieri (2013), que trata de um enorme mapeamento de empresas brasileiras acerca de como elas se posicionam em relação à gestão e governança de dados. Essa publicação consistiu na realização de um questionário respondido por 76 empresas de diversos ramos e tamanhos. Ele monta um panorama do estado de avanço dessas organizações através de gráficos e tabelas aliados de comentários pertinentes.

A partir deste panorama, fica claro que a deficiência na gestão de dados não é um problema exclusivo das universidades, ou sequer restrito a administração pública indireta, como ilustrado pela dissertação anterior, mas também afeta diversas entidades do segundo setor, independentemente de seu porte. Em particular, a área mais crítica identificada é a gestão de dados mestres e de referência (gráfico 11). Nele, vemos que todas as atividades relativas a este aspecto não eram realizadas por mais da metade das empresas.

Este aspecto, que será abordado com mais profundidade ao longo deste trabalho, representa um ponto de partida fundamental para estruturar a gestão de dados de forma abrangente dentro da organização.

3. SITUAÇÃO DA UFRJ

Como mencionado na seção de metodologia, o estudo de caso da Universidade Federal do Rio de Janeiro sobre seu nível de maturidade na gestão de dados mestres e referência foi realizado através de dois meios investigativos.

O primeiro deles foi a entrevista estruturada de pessoas envolvidas em sistemas que possam representar bons exemplos de algumas das questões consideradas relevantes no escopo deste trabalho. Os sistemas escolhidos foram o Sistema Integrado de Gestão Acadêmica (SIGA) e o da Pró-Reitoria de Pós-Graduação (PR2).

A razão por trás dessa decisão reside no fato de que o SIGA compreende os dados principais da UFRJ relacionados aos alunos e professores – entidades essenciais ao negócio de uma instituição de ensino superior, sendo assim, um forte candidato para estudarmos os dados mestres. Além disso, o interesse pelo sistema da PR2 nasceu nos avanços do projeto Conecta UFRJ que tem como objetivo o mapeamento de competências dentro da universidade. Nele, lidamos com muitos dados da UFRJ, e ficou evidente os desafios de interligar os dados oriundos de diferentes fontes internas e externas à universidade. Ao colaborar com o Conecta, observamos algumas oportunidades de melhora na gestão dos dados mestres e de referência.

A segunda parte do estudo de caso se deu por uma análise documental, especificamente nos documentos que dizem respeito ao Plano de Dados Abertos da UFRJ, à exemplos dos PDAs da UFRJ em suas diversas versões, relatórios referentes aos planos consecutivos de abertura de dados e que sofrem acompanhamento de suas execuções por parte da Controladoria Geral da União (CGU) acerca da transparência da UFRJ. O projeto do Plano Geral de Dados Abertos para o Poder Executivo consiste em definir planos bienais, estipulando metas e estratégias de abertura de dados para diversas entidades da administração pública, o que inclui cada Universidade Federal brasileira (BRASIL, 2016). Além

deles, analisamos uma análise do Tribunal de Contas da União sobre transparência no setor público.

Através da leitura dos PDAs que a UFRJ publicou até o presente momento, podemos ter um entendimento da evolução do comprometimento que esta instituição dá à transparência e gestão de dados, e analisar o avanço da disponibilidade e maturidade de seus dados.

3.1 ENTREVISTA

Como mencionado na seção de metodologia, o estudo do caso foi concretizado através de entrevistas estruturadas de servidores nos sistemas de gerenciamento de dados da Universidade Federal do Rio de Janeiro que podem fornecer *insights* e informações relevantes sobre como a UFRJ realiza a gestão de seus dados, especialmente dados mestres e de referência e entender de forma geral como esses sistemas se relacionam.

A entrevista estruturada foi cuidadosamente formulada com base em questionamentos e reflexões suscitadas tanto pelos materiais acadêmicos trazidos na revisão bibliográfica quanto no meu conhecimento tácito desenvolvido durante meus anos formativos de graduação, em particular nas matérias de Processos de Software, Organização de Dados, Modelagem de Sistemas, Banco de Dados e Governança e Gestão de Dados, além do meu trabalho na iniciativa Conecta UFRJ, que ofereceu constante contato com a pletera de sistemas que ela possui.

3.1.1 PRÓ-REITORIA DE PÓS-GRADUAÇÃO

A primeira entrevista atentou-se em explorar alguns aspectos de dados mantidos em um sistema da Pró-Reitoria de Pós-Graduação, mais especificamente dados sobre os laboratórios da UFRJ e registro para diplomação

dos alunos de mestrado e doutorado. Aqueles são exclusivamente encontrados no sistema da PR2, no entanto, estes possuem versões noutros sistemas, como no SIGA. Esta questão é relevante no contexto desse trabalho pois se há mais de uma versão contida em diferentes sistemas, elas devem ser consistentes e para isso, o desejável seria que esses sistemas pudessem interoperar. Neste sentido, a PR2 também recebe dados de outros sistemas, o que faz com que a consistência entre eles seja constantemente necessária, fomentando a interoperabilidade e a confiabilidade dos dados.

Ao curso da entrevista, entendemos que a troca de dados entre sistemas e áreas é feita mediante solicitação e transcrição manual, com pouca ou nenhuma automação ou uso de APIs (*Application Programming Interface*). Isto torna o processo propenso a erros e com pouca agilidade, requerendo esforço considerável para sua manutenção e efetividade de seu uso. O método oficial de transmissão de dados entre os sistemas é, na maior parte das vezes, baseado em processos do SEI (Sistema Eletrônico de Informações), mas a ausência de padronização para o tráfego de dados e a dependência deste sistema podem representar um gargalo na eficiência e eficácia dos sistemas. Isto porque o SEI não é uma ferramenta de integração, ou seja, não foi feito para trocar dados estruturados e automatizar a comunicação direta entre as bases ou sistemas de dados. Além disso, há um potencial risco de segurança em relação aos dados sensíveis e à falta de rastreabilidade sobre quem modificou o que e como.

Outra questão crítica é o fato de que dados como a listagem de cursos de graduação vêm do Ministério da Educação (MEC), para garantir consistência com o que está registrado em instâncias superiores oficiais externas. Dessa forma, a UFRJ precisa solicitar a outra organização o acesso aos seus próprios dados mestres. Isso é devido ao fato de que esta fonte é considerada mais confiável pela UFRJ do que seus próprios dados internos, tornando-se assim, a fonte oficial, mesmo que os dados que estão lá sejam abastecidos pela própria universidade. Isto representa uma perda de autonomia estratégica e é uma evidência de que a

universidade ainda não tem uma efetiva governança e gestão de seus dados mestres.

Analisando especificamente os dados sobre laboratórios, exclusivamente mantidos pela PR2, enxergamos que os dados nascem internamente à universidade, através de formulários preenchidos por professores que chefiam os laboratórios. Este formulário é então verificado pelo coordenador do centro, e então é encaminhado à PR2, que insere os dados no sistema. Além de novamente funcionar por replicação manual dos dados, encontramos outros desafios, como a ausência de definição formal de laboratório e seus requisitos, que ficam à cargo discricionário (e manual) do coordenador em aprovar ou não, gerando uma falta de aferição de qualidade objetiva.

Pode-se identificar neste exemplo figuras importantes da gestão de dados, como o dono do dado – manifestado, *de facto*, na figura do coordenador do centro – que é aquele com jurisdição para definir as regras de negócio, poder decisório e possuem uma visão ampla e difusa sobre o domínio (DAMA, 2017); e o guardião dos dados – representado, novamente *de facto*, pelo professor que chefia o laboratório – que é responsável por preencher, coletar e manter os dados, e são aqueles que possuem uma visão específica e detalhada sobre o domínio (DAMA, 2017).

No entanto, no caso da UFRJ, por não ter ainda estabelecido formalmente uma área de governança de dados, faltam definições oficiais dos papéis e suas responsabilidades em relação aos recursos de dados da universidade. Assim como foi implementado em outras universidades, é relevante adotar a criação de um setor ou unidade formal em sua estrutura encarregada da governança e gestão de dados, formalizando, assim, perante sua instituição, a importância de uma atenção maior aos recursos de dados da universidade. Estes setores são também os responsáveis por estabelecer e controlar padrões claros de qualidade e as regras que devem ser observadas.

Outrossim, não existem incentivos para que os professores atualizem o sistema, sendo a atualização um evento totalmente volitivo por parte dos

docentes. Quando os dados são atualizados, os novos apenas substituem os anteriores, não havendo qualquer gestão de dados legados ou manutenção de histórico, dificultando análises de tendências, auditorias e compatibilidade entre os registros legados e os dados atuais.

Ainda com relação ao conceito de laboratório, surgiram durante a discussão sobre os dados atualmente sendo mantidos, novas demandas externas associadas a laboratórios. Uma notável é o debate sobre definição formal do próprio laboratório em si. Esta é uma definição que transcende o mero rigor formalista ao envolver questões legais e orçamentárias, como o recebimento de bolsas. Para chegar a uma definição para a UFRJ, podemos nos voltar a outras definições à nível federal, como a da Plataforma Nacional de Infraestrutura de Pesquisa (PNIPE) e a do Censo da Educação Superior (CenSup).

A PNIPE cria uma definição estrutural, ao falar de infraestrutura de pesquisa, que é um gênero de lugar, com laboratório como sua espécie. Uma infraestrutura de pesquisa é descrito como aquele espaço físico que possui equipamentos e instrumentos com o objetivo de realizar pesquisa e desenvolvimento (P&D), mas também, permite que seja, alternativamente, qualquer outra estrutura que tenha “recursos baseados em conhecimentos [ou] tecnologia da informação e comunicação [, ou ainda,] outra infraestrutura de natureza singular utilizada em atividades de P&D” (BRASIL, 2025).

Já o CenSup é mais pragmático, definindo o laboratório como “um espaço físico, utilizado por alunos da graduação, equipado com recursos, instrumentos, materiais ou tecnologia destinados ao ensino, pesquisa ou extensão”. E ainda, complementa com uma definição menos literal: “seu principal objetivo é proporcionar um ambiente controlado para a realização de atividades práticas, experimentais ou de desenvolvimento acadêmico, complementando o aprendizado teórico” (MEC, 2025).

Desta maneira, vemos que há uma grande sobreposição entre os dois conceitos, cada uma com suas abordagens e particularidades. Suas divergências, porém, são benéficas para criar a própria definição usada pela universidade.

Nenhuma definição oficial poderá abarcar os dois conceitos na sua integralidade. É necessária que a definição própria acomode-os tendo como parâmetro uma revisão da situação atual do sistema da PR2.

No geral, essa primeira entrevista serviu como preâmbulo dos problemas enfrentados por outros sistemas da UFRJ. É importante salientar a intrínseca noção dos problemas existentes nos sistemas por parte dos servidores, que entendem a natureza dos desafios enfrentados quanto à qualidade dos dados por eles mantidos, mas sem estar ao seu alcance a autoridade e agenciamento devido para sua solução.

3.1.2 SISTEMA INTEGRADO DE GESTÃO ACADÊMICA

Outra série de entrevistas bastante elucidadora foram as que concernem o Sistema Integrado de Gestão Acadêmica (SIGA). Este se mostra como um sistema mais central, contendo o ciclo de vida do discente, desde o primeiro instante, antes de ingressar na universidade, como candidato à vaga pelo SiSU (Sistema de Seleção Unificada), até depois da formação, e portanto, é um forte candidato a ser o baluarte de uma versão unificada de dados, ou seja, do sumo registro mestre, tornando-se um sistema de registro.

Além disso, ele também serve de fonte de dados para sistemas de outros setores da universidade com base em diretrizes definidas por entidades superiores, como Divisão de Ensino da PR1 (Pró-Reitoria de Graduação), PR2 e PR5 (Pró-Reitoria de Extensão) ou o MEC. O fato do SIGA já alimentar outros sistemas demonstra uma experiência em ser uma fonte confiável e soberana, bem como atesta sua capacidade de governança funcional em manter e seguir regras de negócio estabelecidas por outrem. Isso demonstra novamente um avanço deste sistema no quesito da maturidade da gestão de dados, que o faz de maneira segura e tempestiva.

Essa atenção se revela no desenvolvimento e uso de APIs próprias para integração dos dados, como é o caso das APIs utilizadas no projeto extensionista

Caronaê, que consiste em um aplicativo de troca de caronas restrito aos alunos e funcionários da UFRJ, dessa forma, o SIGA precisa fornecer dados que verifiquem os usuários como membros desta instituição, sendo dados que envolvem tanto os domínios de alunos de graduação e pós-graduação (PR1 e PR2), como de servidores – que está sobre a tutela da PR4 (Pró-Reitoria de Pessoal); ou nas catracas de ingresso ao NCE (Núcleo de Computação Eletrônica), que apenas permitem a entrada de alunos e servidores cadastrados através de API feita pelo SIGA.

Apesar disso, é observável uma falta de interoperabilidade bidirecional, já que o SIGA é o único provedor de APIs. Sendo assim, quando eles precisam de um dado interno gerado por outro sistema, não existe API que faça essa integração, restando a opção perniciosa da replicação manual, com potencial de introduzir inconsistências difíceis de resolver sem um processo de MDM formalizado, especialmente ao ponderar o SIGA como um candidato a ser um sistema de registros.

Não obstante, dados externos à UFRJ também se mostram desafiadores quanto à interoperabilidade e integração. A exemplo do MEC, que não possui API própria para extração e obtenção de seus dados utilizados pela UFRJ. Por isso, a obtenção destes dados é feita através de uma planilha CSV (*comma-separated values*), cujo acesso é concedido apenas ao superintendente da PR1, e portanto, a chegada dos dados ao sistema é mediata à solicitação; ou, de maneira análoga, as informações dos cursos de pós-graduação, oriundos da plataforma Sucupira (Maia, 2020).

A falta de uma *pipeline* automatizada de ingestão dos dados de referência externos é um obstáculo sistemático que está fora do controle da instituição e seus servidores, mas que não exime a responsabilidade da universidade de mitigá-la, reforçando a dificuldade de se adaptar aos padrões externos não padronizados.

Outro fator levantado nas entrevistas que se mostrou bastante significativo é a falta de compartilhamento de conceitos globalmente em toda a organização.

Na ausência de definições, de responsáveis por elas e de sua divulgação de forma sistemática e ampla entre as diversas unidades, cada sistema utiliza sua perspectiva particular sobre determinado elemento de informação.

Podemos exemplificar, para dados de referência, as listagens de municípios do Brasil, obtidos pelos sistemas de forma independente, em vez de, ao menos, serem compartilhadas e unificadas em toda a instituição; ou dados sobre servidores, advindos da PR4, que são copiados manualmente em outros sistemas, sem vinculação entre as versões, existindo a possibilidade de ficarem desatualizados e inconsistentes; ou ainda, caso um aluno mude seus dados cadastrais no SIGA, as novas informações não são atualizadas automaticamente nos demais sistemas. Para dados mestres, podemos citar o conceito de aluno, que pode ser entendido de diversas maneiras, como aquele matriculado, ou aquele que possui matrícula ativa, ou aquele que está cursando alguma matéria, etc.

Aliada a esta carência, o escopo dos domínios de dados é dúbio, levando a sobreposições e conflitos entre os sistemas, e a potencialização dos silos de informação. Nomeadamente, podemos referenciar o SIGA, que possui dados cadastrais sobre contas bancárias, que também podem ser encontrados no sistema da PR7 (Pró-Reitoria de Políticas Estudantis), ou mesmo a já mencionada sobreposição entre o sistema da PR2 e o SIGA, a respeito da pós-graduação.

No SIGA, foi apontado como desafios para melhoria da gestão dos dados mestres e dos dados de referência a falta de centralização, definição clara dos papéis e domínios, mapeamento das fontes dos dados e problemas de sincronização e compatibilidade tecnológica. Isso demonstra como a equipe é autoconsciente das carências que os cercam, já que estes foram justamente os pontos elencados até então durante o trabalho.

3.2 ANÁLISE DOCUMENTAL

3.2.1 PLANO DE DADOS ABERTOS

A UFRJ, como uma das maiores e mais antigas instituições de ensino superior do país, se dispôs, em 2017, a disponibilizar seus dados, salvo os sigilosos e pessoais. No primeiro documento de PDA da UFRJ, que estabelecia planos para o biênio de 2017 a 2018, foram elencados 6 (seis) grandes conjuntos de dados, cada um com seus respectivos subconjuntos de dados, marcados como prioridade para serem abertos ao final do ano de 2017. Além disso, foi traçada uma estratégia para fê-los, bem como métricas de qualidade dos dados (UFRJ, 2017).

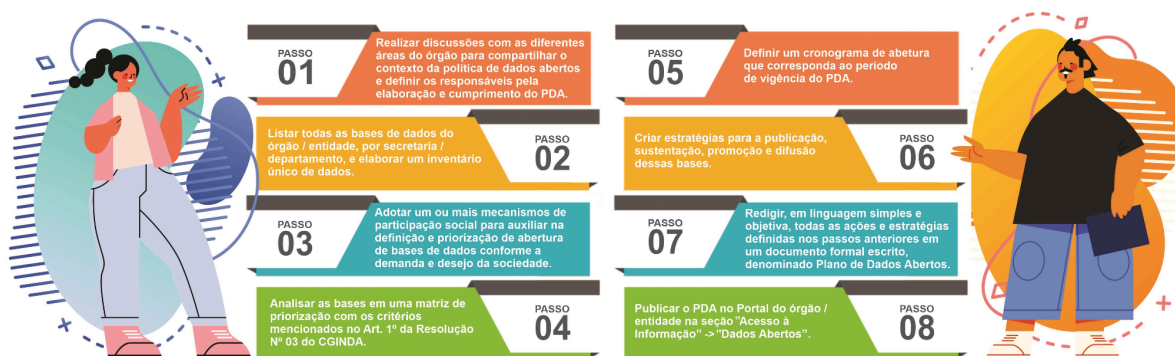
Seriam eles: dados cadastrais, acadêmicos, administrativos, da ouvidoria, contábeis e hospitalares. Notem que todos esses dados pertencem ou se relacionam diretamente à categoria de dados mestres, por se mostrarem partes essenciais dos negócios que a UFRJ conduz. Além disso, de acordo com a LGPD, os dados de referência sobre alunos (pessoas) devem ser disponibilizados de maneira agregada ou anonimizados (BRASIL, 2018). Dessa forma, não basta a UFRJ apenas abrir seus dados, deve ser feito um trabalho prévio de anonimizar ou agregar os dados de referência.

Os dados seriam divulgados na página principal da universidade¹, na página dedicada ao PDA da UFRJ², e seriam encaminhados para o MEC e para Portal Brasileiro de Dados Abertos, além de virem a constar no repositório de dados em CKAN (*Comprehensive Knowledge Archive Network*) do Governo Federal³ e da própria UFRJ⁴ (UFRJ, 2017). O CKAN é um portal *open-source* destinado à arquivagem e distribuição de dados abertos. Ele foi criado pela Open Knowledge Foundation, e é utilizado por organizações privadas e governamentais ao redor do mundo para trazer transparência e divulgar dados.

Após isso, um novo PDA foi formulado, compreendendo o prazo de 2020 a 2022. Dessa vez, foram declaradas 147 (cento e quarenta e sete) bases de dados

para serem abertas, cada uma atrelada a um responsável de alguma unidade, com previsão de abertura, bem como exposta uma frequência de atualização (UFRJ, 2020). Todos esses fatores são muito importantes na elaboração de um PDA, de acordo com o guia do Ministério Público abaixo.

Figura 2 - Etapas para construção de um PDA



Fonte: Governo Federal

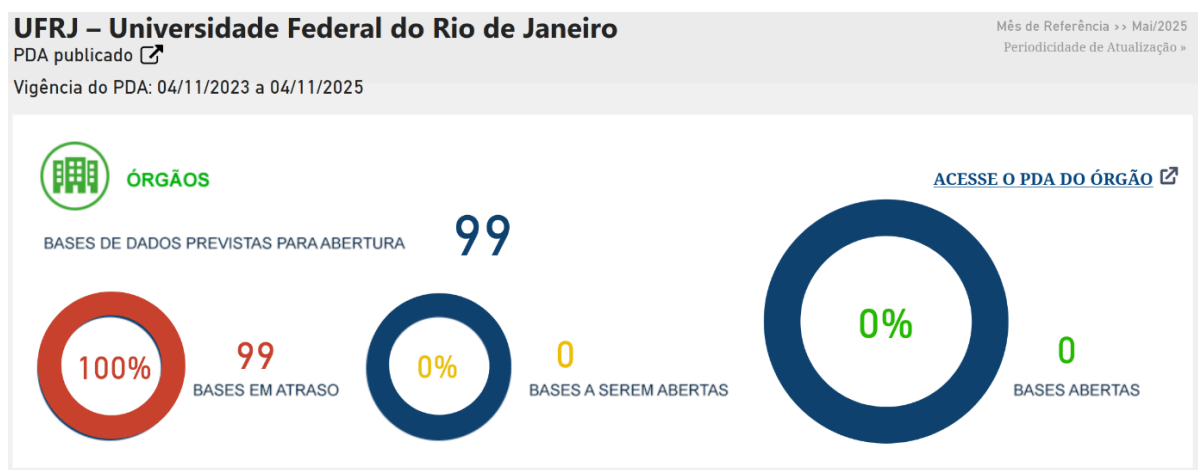
(<https://www.gov.br/governodigital/pt-br/estrategias-e-governanca-digital/sisp/guia-do-gestor/pda>)

Um relatório a respeito dos desafios e sucessos encontrados na execução do PDA de 2020 foi redigido pela UFRJ. Nele, a UFRJ se congratula por ter aberto todas as bases prometidas. Mas, ao mesmo tempo, elenca como o maior desafio a completar seu objetivo a pandemia global de COVID-19 (UFRJ, 2022).

Após isso, um novo PDA foi elaborado, compreendendo os anos de 2023 a 2025. Neste mais recente, 94 (noventa e quatro) bases de dados foram prometidas a serem abertas. Todas as 94 estão incluídas nas 147 bases de dados do PDA anterior (UFRJ, 2023). No relatório que a UFRJ publicou analisando o seu desempenho na realização do PDA de 2023 a 2024, é informado que 74 dessas bases foram abertas de forma integral, e se compromete a continuar os esforços para abrir as demais ao longo do prazo estabelecido (UFRJ, 2025).

Entretanto, no painel do Governo Federal de indicadores de abertura de dados de seus órgãos, a UFRJ atualmente consta com 0 de 99 (zero de noventa e nove) bases abertas, o que configura como 0% de sua meta, conforme ilustrado abaixo.

Figura 3 – Painel do Governo Federal das bases de dados abertas pela UFRJ

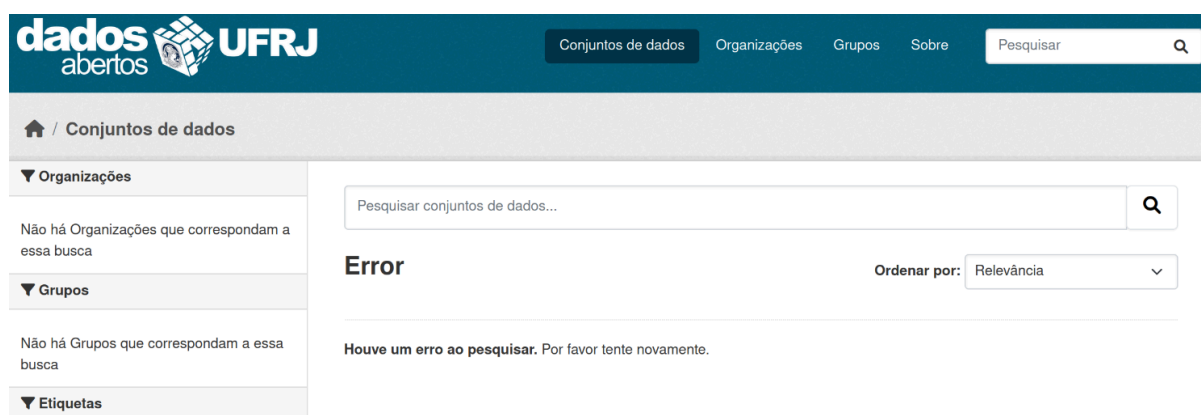


Fonte: Controladoria Geral da União (<https://centralpaineis.cgu.gov.br/visualizar/dadosabertos>)

Ainda sobre o painel do Governo Federal, precisamos reportar que algumas outras universidades, de porte similar ao da UFRJ, pesquisadas não possuíam PDAs em vigor, com exceção da Universidade Federal de Minas Gerais (UFMG). Ela não só tem um Plano de Dados Abertos em vigor durante a pesquisa, como também cumpriu a abertura de todas as 19 (dezenove) bases prometidas. Isso pode sugerir que estabelecer metas mais realistas (dezenove, em vez de noventa e nove) possam auxiliar na factibilidade de seu cumprimento.

Além disso, ao acessar o CKAN da UFRJ destinado à divulgação e acesso aos dados abertos, nada se encontra disponível, conforme ilustra a imagem abaixo, senão uma mensagem de erro. Considerando a possibilidade da página se encontrar fora do ar momentaneamente, foram feitas outras tentativas de acessar o *website* ao longo de três meses e todas proviram o mesmo resultado.

Figura 4 – Página dos dados abertos da UFRJ



Fonte: Universidade Federal do Rio de Janeiro (<https://dados.ufrj.br/dataset/>)

Dessa maneira, não foi possível, em nossa pesquisa, confirmar ou negar as informações apresentadas nos relatórios da UFRJ a respeito dos Planos de Dados Abertos. Figurando-se uma incógnita a respeito de onde estariam os dados abertos que a universidade alega ter disponibilizado, quando logram-se de êxito absoluto.

Para finalizar esta seção sobre os dados abertos da UFRJ, precisamos reconhecer o sucesso na abertura dos dados relacionados ao Complexo Hospitalar Universitário da Universidade do Rio de Janeiro (CHS). Os conjuntos de dados do CHS estão incluídos nos conjuntos de dados que a UFRJ prometeu disponibilizar. Podemos checá-los no CKAN específico do CHS, que está exposto em figura abaixo, todavia, seu avanço não consta no painel do Governo Federal sobre dados abertos da UFRJ. Portanto, a simples vinculação ou indexação dos dados abertos pelo CHS daria um avanço de 13 (treze) conjuntos de dados à UFRJ.

Figura 5 – Página de dados abertos do CHS-UFRJ

The screenshot displays the 'dados abertos CHS UFRJ' website. The header includes navigation links for 'Grupos', 'Organizações', 'Conjuntos de dados', and 'Sobre', along with a search bar. The main content area shows '13 conjuntos de dados encontrados'. A sidebar on the left lists various organizations and groups with their respective dataset counts. The main list of datasets includes 'Residência CHS', 'Produção Assistencial do INDC', and 'Relatório de Gestão das Unidades do CHS 2024'. Each dataset entry shows available file formats like XLS, XLSX, and PDF.

Organizações	Grupos
Instituto de Pueric... (3)	Assistência (4)
Complexo Hospitalar... (2)	Gestão (2)
Instituto de Gineco... (2)	Ensino, Pesquisa e ... (1)
Instituto de Neurol... (2)	
Hospital Universitá... (1)	
Instituto de Atençã... (1)	
Instituto de Psiqui... (1)	
Maternidade-Escola (1)	

13 conjuntos de dados encontrados

Ordenar por: Relevância

Residência CHS
Este conjunto de dados não tem descrição
XLS XLSX PDF

Produção Assistencial do INDC
DADOS ABERTOS ASSISTENCIAS DO INDC 2022_2023_2024
XLS XLSX

Relatório de Gestão das Unidades do CHS 2024
Este conjunto de dados não tem descrição
XLSX

Fonte: Complexo Hospitalar e de Saúde da UFRJ (<https://dadosabertos.ch.ufrj.br/dataset/>)

Acreditamos, com base no que se evidenciará, que isso se deve pelo fato de, apesar da Universidade Federal do Rio de Janeiro ter todos esses dados, seu acesso a eles – e subsequente disponibilização – é dificultado pela ausência de um banco de dados analítico, em razão de um nível de gestão de dados mestres e de referência ainda embrionário, bem como à falta de uma área de governança de dados que proponha políticas e estratégias, além de definir papéis e responsabilidades para seus dados críticos.

Como mencionado, os dados que estão para serem abertos são dados mestres, que se utilizam de dados de referência como contextualizadores. Dessa forma, para eles estarem prontos para a divulgação, todos os dados de referência devem estar disponíveis e consistentes entre si, para que seja possível traçar comparações entre eles. Sem a viabilização destes dados de referência, a transparência perde um de seus sentidos, a análise histórica e auditoria popular e governamental, estando assim, em desconformidade às normas legais nacionais.

3.2.2 ANÁLISE DO TRIBUNAL DE CONTAS DA UNIÃO

Ademais, outra série de documentos analisados, que se relacionam com o Plano de Dados Abertos do Poder Executivo foram aqueles editados pelo Tribunal de Contas da União que urgiam às universidades públicas para que melhorassem suas gestões de dados. Com essa prática, era esperado mais transparência nas instituições de ensino superior do país, que se traduziriam em mais auditabilidade para o TCU.

Neste cenário, o órgão de controle externo conduz estudos acerca da maturidade de TI (tecnologia da informação) de diversas entidades da administração direta e autárquica desde 2017. Apesar deste trabalho se dedicar à gestão de dados, e não à gestão de TI, ambas são intimamente relacionadas, mesmo sendo distintas. A gestão de tecnologia da informação se volta a *hardware*, *software* e arquiteturas técnicas de funcionamento dos dados, enquanto a gestão de dados foca-se no entendimento dos dados como ativo. Dessa forma, a TI serve como alicerce para a gestão de dados (DAMA, 2017).

Esses estudos do TCU são elaborados com base no iGG (índice integrado de Gestão e Governança Pública). Seu score é calculado através de um questionário de autoavaliação composto por uma análise de quatro áreas: governança pública, e capacidade de gestão de pessoas, de tecnologia da informação, e de contratos (TCU, 2017).

O TCU esclarece que, apesar de sua natureza numérica sugerir a possibilidade de comparações objetivas entre a maturidade de cada instituição, o iGG não tem como objetivo classificar ou hierarquizar as organizações públicas. Assim, a comparação deve ser feita com cautela, porque as instituições podem não ser diretamente comparáveis. E ainda, ele alerta que a função deste índice é o de expor pontos de melhoria da instituição, dessa forma, o foco não deve ser a imediata majoração do valor do iGG, mas a melhora das carências evidenciadas. A última ressalva é que o questionário é autoavaliativo, sem que haja um processo

de validação completo das informações, implicando que pode haver imprecisões nos resultados (TCU, 2018).

Após três edições dos relatórios de nível de maturidade em gestão pública, o Tribunal de Contas publicou um compilado analisando os resultados encontrados. Apesar de não podermos enxergar os detalhamentos de cada instituição, vemos que o TCU colocou a UFRJ na 26ª posição de 63 universidades participantes (172ª de 716, no geral) no mérito de adesão às práticas de governança e gestão públicas definidas pelo órgão de controle externo (TCU, 2021). Nesse posto, foram avaliados tanto a maturidade geral quanto a evolução de 2017 a 2021. Nesta classificação, vemos o CHS, que está na posição 235ª de 716 no posto geral.

Esses dados nos sugerem que a UFRJ tem melhorado bastante sua gestão e governança de TI, modernizando processos e se adequando às políticas estabelecidas pelo Governo Federal. Esta constatação tem sua importância ao entender a TI como fundamento da gestão de dados, servindo como um substrato sólido no qual ela pode se firmar e crescer. Assim, a universidade tem estrutura suficiente para desenvolvê-la, bastando formalizar sua definição e implementar os passos necessários.

A partir dos dados analisados podemos concluir ainda que a Universidade Federal do Rio de Janeiro não enfrenta um desafio isolado na transparência pública, mas sim um reflexo das dificuldades comuns a toda a administração pública. Embora existam deficiências, a universidade se destaca, mantendo-se no primeiro quadrante entre todas as instituições e na primeira metade das universidades (TCU, 2021). Assim, sua experiência ilustra um cenário generalizado, no qual a superação desses obstáculos é um processo coletivo e contínuo, e não uma falha particular da UFRJ.

Este ponto de vista é reforçado por Zorzal (2015), que analisa, conforme metodologia própria, a transparência de cinco universidades federais brasileiras, uma de cada região do país, exclusive a Universidade Federal do Rio de Janeiro. Ela conclui que nenhuma delas atingiu o nível alto de transparência, com as mais

altas empatando no nível médio-alto. Isto denota a dificuldade compartilhada por todas as universidades, e mostra a necessidade de mais pesquisa e trabalho na área da gestão de dados.

3.3 SILOS DE INFORMAÇÃO

Um dos maiores desafios da gestão de dados mestres é a integração de diferentes setores de uma organização de modo que atenda aos objetivos comuns dessa organização. O DMBok nos alerta sobre os perigos de adicionar cada vez mais sistemas isolados: por mais que individualmente eles sejam simples, a complexidade se revela na necessidade de criar mecanismos elaborados de compartilhamento dos dados entre eles. A complexidade se traduz no custo de compreender e manter as relações entre os sistemas, que por sua vez, cada vez mais evidencia as vantagens de bancos de dados corporativos. Para esses sistemas não bem integrados, o livro apelida de silos de informação, evocando a ideia daqueles reservatórios gigantes de grãos que não se comunicam.

Todavia, não existe um problema inerente em ter um sistema distribuído de armazenamento de dados. O ponto principal é que urge uma arquitetura de integração eficaz que controle o acesso compartilhado, a replicação e o fluxo dos dados mestres e de referência, a fim de garantir a qualidade e consistência deles, visando à interoperabilidade. Sem essa arquitetura, inevitavelmente haverá dados inconsistentes e redundantes.

Uma abordagem interessante sugerida é a de analisar alguns poucos dados em um domínio por vez para evitar complexidade excessiva. Com essa abordagem incremental, garantimos que os dados não só estão consistentes e unificados dentro de um sistema, mas entre todos os sistemas. (DAMA, 2017).

Rêgo (2013) reforça essa noção, estreitando a relação entre uma má gestão dos dados mestres e de referência e a criação dos silos de dados, trazendo como consequências o histórico de sistemas e processos descentralizados e

independentes, funcionários das aplicações não alinhados com necessidades de informação e pouco foco em informações não estruturadas, acarretando em um “número crescente de disparidades de conceitos e valores nas aplicações [...] e apresentando duplicações e representações redundantes [...] sobre os mesmos conceitos” (RÊGO, Bergson L., 2013, p. 182).

4. RECOMENDAÇÕES COM BASE NO ESTADO DA ARTE PARA UFRJ

4.1 MATURIDADE DE GESTÃO DE DADOS

As informações levantadas pelas análises documentais sugerem a existência de silos de informação dentro da Universidade Federal do Rio de Janeiro. Convergentemente, os dados levantados em exemplos discutidos em nossas entrevistas apontam que os sistemas da UFRJ ainda carecem de mecanismos de interoperabilidade, atuando assim, em muitos casos, como silos de informação.

No contexto de implementação de arquitetura de gestão de dados, a situação da universidade se enquadra no “pré-MDM”, como chama Rêgo o estágio em que “as aplicações não estão integradas do ponto de vista de uma gestão de dados mestre[s]” (RÊGO, Bergson L., 2013, p. 185). Ele é caracterizado pela existência de silos de informação, nos quais os seus dados estão contidos dentro dos escopos particulares de cada aplicação e geralmente não são compartilhados (RÊGO, 2013).

Rêgo (2013) ressuscita a noção levantada pelo DMBok na seção anterior ao explicitar que a mera presença de sistemas distribuídos que não compartilham dados não é uma característica intrinsecamente ruim. Na verdade, caso uma transação se dê integralmente dentro da conjuntura de um sistema, esses dados não precisam se submeter a um compartilhamento. No entanto, ele reforça que dados altamente replicados através de grande extensão de aplicações da organização – como professores, disciplinas, e alunos, no caso de uma instituição de ensino – precisam ser governados a fim de mitigar problemas de qualidade de dados e integração.

Ele vai além e nos sugere diversas formas pelas quais podemos adotar uma gestão de dados mestres. Porquanto desejamos resguardar a estrutura atual de sistemas, visando a simplificação na implementação e baixa onerosidade, devemos preferir arquiteturas que permitam sistemas distribuídos, sem a necessidade de unificar todos os dados em um único banco de dados

centralizado. Decerto, um dos objetivos primários da gestão de dados principiante deve ser a dissolução dos silos de informação a fim de melhorar a qualidade dos dados, a transparência, a auditabilidade e eficiência da universidade.

4.2 RECOMENDAÇÕES

Inicialmente, ressaltamos a importância de ser institucionalizado e conferido poderes a um comitê gestor de dados na UFRJ, talvez ligados à Pró-Reitoria de Gestão e Governança (PR6). Ele assumiria o papel de diretor de dados em nossa autarquia. Ao comitê estaria a incumbência e responsabilidade de demarcar formalmente os domínios, desenvolver planos de objetivos para melhora da MDM de curto, médio e longo prazo, apontar os cargos subordinados de gestão de dados, cultivar uma cultura de dados, etc. Resumidamente, ser a força motriz que fomenta a gestão de dados dentro da UFRJ, bem como a autoridade máxima no que tange à gestão e governança de dados, com suas prerrogativas e sujeições (DAMA, 2017).

Essa medida é adotada com sucesso em outras universidades que desenvolvem sua gestão de dados a mais tempo que a UFRJ, como a Universidade Estadual de Campinas (UNICAMP), que possui um escritório dedicado a esta área e apresenta uma boa transparência frente às demais universidades brasileiras. Esta medida, inclusive, é incentivada pela coordenadora geral da UNICAMP para ser replicada em outras instituições de ensino superior, adequando-se às particularidades delas (Atvars, 2020).

4.2.1 PAPÉIS EM RELAÇÃO AOS DADOS

A demarcação formal dos domínios bem como a atribuição de papéis em relação aos dados supracitadas são importantes na consolidação de uma

estratégia de gestão e governança de dados. Além de contribuírem para uma semântica globalmente definida através de um glossário comum a toda instituição. Ele deve ser claro, rigoroso, exaustivo em conceitos e abrangente em sinônimos. Uma semântica bem definida é mais do que uma lista de termos e definições, pois os conceitos devem estar associados a metadados, regras de negócio, e demais informações que os contextualizam dentro da organização (DAMA, 2017).

Por sua vez, o glossário contribuirá para a uniformidade dos conceitos, *ergo*, consistência entre os dados, apoiando a integração dos sistemas. Essa medida ajudaria a mitigar alguns dos problemas encontrados, na qualidade de definições ambíguas, incongruências de conceitos entre sistemas, definições extraoficiais, entre outros. Além de sanar esses problemas, traria o bônus de instrumentalizar a busca documental por meio de conceitos-chave de maneira universal dentro dos dados da UFRJ (DAMA, 2017).

A definição formal de papéis em relação aos dados também terá outros efeitos positivos, visto que, apesar de eles existirem hoje, não lhes são conferidos poderes e deveres *de jure*. Por conseguinte, mecanismos protocolares de soluções podem ser traçados, construindo uma rede de comunicação entre papéis de mesma hierarquia para que seja eliminada a dependência de soluções *ad hoc*.

Tais mecanismos devem definir padrões claros de qualidade e completude para os donos dos dados julgarem a admissão das submissões com mais propriedade e estrutura comum. Para tanto, seria importante um canal de comunicação entre os donos dos dados, *vide* os coordenadores, para que criação de um consenso acerca de temas afins.

4.2.2 ARQUITETURA DE GESTÃO DE DADOS MESTRES

Outro elemento singular que deve ser definido pelo comitê de gestão de dados é qual arquitetura de MDM deve ser implementada, levando em conta o sistema de registros. A literatura classifica alguns (três ou quatro, dependendo do

autor) tipos de arquiteturas, sendo nossa recomendação para a UFRJ a consolidação.

Neste estilo, os dados de diversas fontes são centralizados em um *hub*-MDM. O *hub* realiza a limpeza, a padronização e a deduplicação para criar registro mestre de alta qualidade. No entanto, as atualizações desses registros mestres geralmente não são enviadas de volta para os sistemas de origem. Os sistemas de origem continuam sendo a fonte de autoridade para a entrada inicial dos dados (DAMA, 2017).

Argumentos em favor desta implementação são o fato de que é um ponto de partida realista, se tratando de um estilo intermediário entre o registro e os mais centralizados. Sendo assim, ele não é um estilo particularmente complexo de implementar, não havendo necessidade de fazer mudanças estruturais para tanto, o que não resultaria num tempo de transição prolongado, paralisando o SIGA neste ínterim. Apesar disso, ela oferece diversos benefícios imediatamente, como a dissolução dos silos de informação, integrando os sistemas (RÊGO, 2013).

Ademais, a arquitetura consolidada é ideal para análises temporais e dados analíticos, adequando-se bem às necessidades da universidade, levando em consideração que ele unifica as versões, consolidando os dados no *hub*. No entanto, ele o faz sem precisar desestruturar os sistemas já existentes, oferecendo menor custo e trabalho de implementação. Por fim, ele ainda oferece possibilidade facilitada de, caso seja relevante no futuro, evoluir para uma arquitetura mais robusta, como coexistência ou centralização, porque já teria sido feita a consolidação dos dados mestres e de referência limpos, tratados e deduplicados.

Ainda nesta nota, se decidirmos implementar a arquitetura consolidada, precisaremos definir um dos sistemas para ser o *hub* de um domínio, atuando como o sistema de registros deste domínio específico. Os achados da análise documental e entrevistas sugerem que o melhor candidato para esta posição é o SIGA, devido à maturidade de gestão de dados apresentada, seja culturalmente, ou pelo fato de desenvolver soluções baseadas em APIs, ou ainda, por, de certa

forma, já consolidar muitos dos dados da UFRJ em seu sistema. Acreditamos que o SIGA deva ser o sistema de registros de muitos domínios – como aluno, professor, matéria, diploma de graduação, extensão, bolsa, centros e coordenadores –, se não de todos da universidade.

Isto se relaciona diretamente com a dependência do SEI para a troca de informações. Compreendemos que seria mais proveitoso, aliado à implementação da arquitetura consolidada, substituí-lo por um sistema especializado na integração de bases de dados, como são, por exemplo, o ESB (*Enterprise Service Bus*), que atua como um barramento centralizante para o fluxo de dados, garantindo consistência e segurança na integração (CHAPPELL, 2004); uma iPaaS (*Integration Platform as a Service*) que oferecesse um conjunto de ferramentas para construir, implementar e gerenciar as integrações entre os dados, aplicações e sistemas; ou um sistema de ETL (*Extract, Transform, and Load*), típicos de gestão de dados mestres.

Dessa forma, seria estabelecido um arcabouço tecnológico capaz de suportar a governança de dados proposta, permitindo a execução de políticas de qualidade e consistência de forma automatizada. A transição de uma troca de informações baseada em documentos para uma baseada em APIs e protocolos definidos seria crucial para a interoperabilidade, garantindo que os dados mestres e de referência sejam consumidos e atualizados de maneira eficiente e rastreável, sem depender de replicação manual.

4.2.3 GESTÃO DE DADOS DE REFERÊNCIA

Um dos problemas encontrados, comum aos sistemas estudados, é a dificuldade de perpetuar a atualização das fontes de dados de referência. Apesar da estabilidade destes, eles não são imutáveis: novos países podem surgir, logradouros podem mudar de nome, etc. Para essa atualização, Inmon, Strauss e Neuschloss (2010) sugerem dois extremos: o primeiro consiste no acompanhamento regular e periódico das tabelas dos dados de referências, já o

segundo, na atualização instantânea a cada mudança feita. Eles argumentam que existem muitas alternativas de design entre esses dois extremos, e as organizações devem escolher um método que equilibre complexidade, disponibilidade de recursos e a necessidade de completude lógica na gestão de seus dados de referência.

Por isso, cabe aos donos dos dados de cada sistema estudarem suas necessidades e regras de negócios e achar a balança ideal para a frequência de atualizações das fontes de dados de referência para cada um de seus domínios. Para então estabelecer o meio pelo qual estes dados serão atualizados, desenvolvendo, caso seja necessário, APIs especializadas para fazerem a coleta e substituição destas informações. Além disso, é preciso desenvolver um banco de dados, como um *data warehouse*, que acumule informações desatualizadas, para viabilizar análises históricas, compreensão de dados legados e contexto histórico (INMON, STRAUSS, NEUSCHLOSS, 2010).

No início da implementação, é importante entender as necessidades específicas que são resolvidas com esses dados em cada contexto, para então identificar como será esse registro único, e finalmente, procurar e manter o dado original e soberano. Às vezes, ele é encontrado com facilidade, devendo a tarefa principal ser a de impor seu uso em todos os setores (DAMA, 2017).

Em suma, entendemos estes problemas listados como sintomas da inexistência de estratégias e políticas definidas de gestão de dados, e não com a falta de competência individual dos funcionários. Na realidade, por meio das entrevistas, enxergamos um profundo entendimento por parte deles dos óbices que enfrentam em seu labor diário. É notório a ausência de posicionamentos de cima para baixo no tocante à gestão de dados. Os funcionários são impotentes para realizar mudanças estruturais, que só podem ser feitas por quem possui poder decisório real dentro da instituição (RÊGO, 2013).

4.2.4 GESTÃO DE METADADOS

Para haver uma boa gestão de dados mestres e de dados de referência, é imperativo haver também uma boa gestão de metadados. Uma maneira clássica proposta pelo DMBOK para manter metadados é usar um catálogo de dados, que acompanha os bancos de dados. O catálogo de dados é, em essência, um código que armazena os metadados. Nele, pode-se encontrar informações a respeito de onde os dados se encontram, como eles são armazenados, onde e como eles foram gerados, data de criação, entre muitas outras informações. A existência do glossário também impactará positivamente o catálogo, ao fornecer explicações quanto aos termos lá listados (DAMA, 2017).

Além disso, diferentes tipos de dados geram diferentes tipos de metadados. A universidade precisa estar apta para gerenciar toda esta sorte de dados e metadados. Para tanto, deve ser feita uma análise sobre de quais formas diferentes os dados são gerados na UFRJ, para então estruturar como armazená-los (DAMA, 2017). Nesse sentido, esforços de mapear as competências da UFRJ, como o Conecta UFRJ, devem ser continuados e incentivados.

4.3 EXEMPLOS BEM SUCEDIDOS

Por fim, resolvemos investigar, fora da academia, práticas e estratégias que triunfaram na implementação de políticas de gestão de dados mestres e de referência, para servir como parâmetros e metas que qualquer organização que almeje maturidade nesta área deva, ao menos, se inspirar.

4.3.1 PRIMEIRO SETOR

Em busca de uma implementação de políticas de gestão de dados com êxito na administração pública indireta, entrevistamos funcionários da gestão de dados do Operador Nacional do Sistema Elétrico (ONS). Esta conversa ofereceu

um bom contraponto às conduzidas sobre a Universidade Federal do Rio de Janeiro, demonstrando não só que é possível atingir um alto nível de maturidade na gestão de dados no setor público, como também foi possível desenhar abstratamente um mapa de como atingi-lo, salientando outras características importantes da gestão de dados que foram, até então, negligenciadas nesta monografia.

Com base na entrevista, o sucesso do ONS pode ser atribuído a três sólidos pilares: fomento da cultura de dados desde o princípio, forte patrocínio da alta gestão, e uma equipe de arquitetura de dados dedicada à documentação e mapeamento de competências. Esses três alicerces são basais em muitos dos conhecimentos passados na entrevista. Tal como ao expor a importância de uma boa gestão de metadados, o que se traduz num esforço contínuo de documentação, e numa padronização consistente por entre os domínios.

Uma dica simples e eficaz que nos é passada, é o uso de mnemônicos como convenção de nomenclatura unificadamente por toda a entidade, de maneira que garanta a fácil compreensão por qualquer funcionário, independentemente do setor em que ele esteja, ou do conhecimento *a priori* que ele tenha do domínio. Essa estrutura de padrões técnicos deve ser bem documentada e conhecida, e seu uso reforçado para assegurar uma interoperabilidade e qualidade dos dados.

Ademais, é reiterado o papel singular dos arquitetos de dados, salientando sua evolução, saindo de uma função de desenvolvedor de sistemas centralizados, para orquestrador e integrador de sistemas descentralizados, muitas vezes em nuvem. Enxergamos a necessidade da UFRJ em adotar estes pilares e boas práticas, incorporando-as às nossas.

Além disso, observamos o funcionamento da gestão de dados da Universidade Estadual de Campinas, que há mais de oito anos investem recursos nesta área. Nela, foi fundado o Escritório de Dados e Apoio à Transformação (EDAT), que possui uma série de responsabilidades, como as de “dispor os dados, as informações e os relatórios dos Portais da universidade, inclusive o Transparência Unicamp, seguindo modelos de dados abertos (sempre que

possível) e atendendo, simultaneamente, a Lei de Acesso à Informação e a Lei Geral de Proteção dos Dados Pessoais” (Atvars, 2020).

Uma das incumbências do EDAT foi a centralização dos dados da UNICAMP em um *data lake* (análogo ao *data warehouse*), com seus metadados e a composição de um catálogo que viabilizasse o encontro desses dados. Essa é uma das medidas apresentadas nesta monografia e que conseguimos ver o resultado positivo dela na prática. Atualmente, a UNICAMP consegue oferecer transparência de seus dados, não apenas científicos, mas também institucionais, bem como oferecer painéis deles (UNICAMP, 2025).

4.3.2 SEGUNDO SETOR

Adjunto ao lançamento de um *software* de gestão de dados mestres e de referência, a International Business Machines (IBM) lançou um manual de uso com várias dicas e políticas de MDM e RDM usadas por eles mesmo em sua empresa. Com base nele, pegamos algumas ideias que podem ajudar instituições a saírem da inércia nesta área.

A IBM estabelece um processo de cinco passos para se implementar uma gestão de dados de referência. A primeira fase é a de analisar o estado atual, identificar os papéis de dados, encontrar e classificar os problemas e montar prioridades. Em resumo, o que foi feito no capítulo 2 desta monografia. A segunda etapa é a de projetar a solução conceitualmente, isto é, entender quais regras de negócio podem ser atendidas da maneira desejada e como, e antever a factibilidade das decisões. A etapa seguinte é a de configurar, ou seja, ajustar às regras de negócio que não puderam ser alcançadas. A penúltima fase é a implementação do que foi proposto, conduzindo a carga inicial de dados de referência limpos e padronizados ao *hub* consolidado. A etapa final consiste em operar e realizar manutenções no que foi feito até então (IBM, 2013).

A *Big Blue* ainda nos dá algumas dicas para gestão de dados de referência, como o uso de pastas para hierarquizar a informação, e sempre manter um

histórico de versões, para analisar mudanças, dar contexto a dados legados e viabilizar possibilidade de reverter alguma alteração, se necessário. Eles ainda ratificam a ideia do DMBOK de formalizar o processo decisório sobre as alterações nos dados mestres e de referência. Por fim, um conselho que se relaciona com o histórico de versões é o que eles chamam de *soft deletion*: não apagar fisicamente dados de imediato, mas deixá-los numa seção à parte, por segurança e por possibilitar auditorias (IBM, 2013).

5. CONCLUSÃO

5.1. CARTILHA DE DIRETRIZES DE DADOS DE REFERÊNCIA

Agora que entendemos melhor a situação da gestão de dados atual da UFRJ, em particular de dados mestres e de referência, pegamos conselhos e dicas sobre implementação inicial de uma gestão de dados de referência. Precisamos coletivamente promover ações que visam a sua imediata melhora, considerando não somente as imposições legais, sejam elas a fim de transparência, controle, auditoria ou *accountability* (responsabilidade), mas também as vantagens estruturais e culturais para a própria instituição, como maior facilidade para definir políticas, encontrar inconsistências, reduzir custos, e aumentar produtividade e conhecimento eficientemente.

No entanto, devido ao escopo do trabalho e limitações de tempo e recursos, entendemos que devemos dar preferência à resolução da gestão de um dos tipos de dados, deixando a outra aberta para que trabalhos futuros debruçem-se sobre ela, tomando as informações aqui contidas como ponto de partida. Notem que apesar deste parágrafo se referir especificamente à gestão dos dados mestres e de referência, os problemas expostos neste trabalho abrangem outras áreas da gestão de dados, pois elas são interligadas, conforme à mandala do DMBok (Figura 1).

McGilvray (2021) argumenta que uma das raízes de problemas na gestão de dados mestres pode advir de erros nos dados de referência que contextualizam esse dado mestre. Dessa forma, a autora sugere que um projeto que tem como objetivo melhorar a gestão de dados pode economizar tempo e recursos ao incluir dados de referência nas avaliações iniciais de qualidade, em vez de apenas focar nos dados mestres. Isso se dá pelo fato de que esse foco pode ser mais fácil de resolver, e, não obstante, possui um impacto instantâneo na melhor da gestão de dados como um todo, já que como os dados de referência, mestres e

transicionais se fundamentam naquele. Melhorar os dados de referência avança a gestão de todos os demais.

Portanto, propomos a formulação de uma série de guias sobre gestão de dados. Nós desenvolvemos o primeiro deles, um manual ou cartilha com diretrizes abrangentes para se lidar especificamente com dados de referência na UFRJ – que pode ser encontrado em sua integridade no meu repositório pessoal Github⁶, dedicado a este trabalho, bem como no repositório Github dedicado à promoção da gestão de dados da UFRJ do Grupo de Engenharia do Conhecimento (GRECO)⁷. Esse guia tem inspiração em outros similares, compostos por demais instituições de ensino superior brasileiras que cremos, aliados ao levantamento do TCU de 2021 mencionado anteriormente, estar num nível mais maduro de gestão, bem como possui a incorporação dos conhecimentos aprendidos noutras instituições públicas e privadas. Além disso, ele tem um foco maior nos problemas que conseguimos identificar ao longo das pesquisas de campo, documentais e entrevistas.

5.2 LIMITAÇÕES

Considerando as limitações de tempo e escopo, este trabalho trata apenas de compreender o estado atual da Universidade Federal do Rio de Janeiro no tocante à gestão de dados mestres e referência, utilizando os conhecimentos do estado da arte, e de sugerir políticas e diretrizes que podem melhorar a gestão de dados de referência pelos setores da UFRJ com base na literatura de gestão de dados e no que outras entidades da administração pública indireta e empresas de grande porte que obtiveram sucesso na sua gestão e governança de dados fazem. Sem, no entanto, se preocupar com sua implementação, que compete exclusivamente à decisão discricionária do magnífico reitor ocupante.

Além disso, outras áreas interligadas à gestão e governança de dados que não foram estudadas neste trabalho foram a avaliação formal de maturidade, exposto previamente; e a governança de dados científicos e de pesquisa, que se

relaciona com ciência aberta e princípios FAIR por estarem fora do escopo estipulado.

Ainda, apesar de debatermos sobre metadados com alguma frequência neste trabalho, não nos aprofundamos nas arquiteturas e estratégias de gestão de metadados. Foi falado brevemente sobre ela por considerarmos fulcral à gestão de dados mestres e de referência, mas não nos debruçarmos em demasia para não sairmos do tema proposto.

Finalmente, apesar da extensiva busca não ter sido capaz de encontrar os dados abertos que a UFRJ alega ter disponibilizado, a ausência de prova não se configura como prova de ausência. É possível que eles estejam no repositório da UFRJ destinado a este fim, que, todavia, encontra-se fora do ar. Esta monografia não afirma, tampouco nega a abertura dos dados pela Universidade Federal do Rio de Janeiro.

Desta maneira, entendemos que as limitações deste trabalho representam oportunidades de trabalhos futuros para pesquisadores e profissionais que desejam melhorar a gestão de dados desta ou outra instituição. Nomeadamente, a replicação adaptada da cartilha aqui elaborada noutra entidade da administração pública; ou uma contínua evolução à gestão de dados da UFRJ, como avanços na implementação das diretrizes ou na elaboração de soluções que tangem os dados mestres, ou até progresso em subáreas da gestão de dados intocadas neste trabalho, à título de gestão de documentos ou segurança dos dados.

5.3. TRABALHOS FUTUROS

Finalmente, gostaríamos de mencionar a possibilidade de aplicação dos trabalhos citados na seção de Trabalhos Similares no contexto da UFRJ, com mais tempo e escopo ampliado se tratando de temas decorridos durante mestrado e doutorado. Esses desenvolvimentos futuros devem fortalecer ainda mais as informações geradas acerca da universidade, e com isso tomadas de decisões mais assertivas, bem como otimização de processos e diminuição dos custos. Ultimamente, as instituições, ao se tornarem mais transparentes, permitindo adequação maior às

normas e contribuindo com o conhecimento público em nível superior, contribuirão às fundações do estado democrático de direito, solidificando-as, ao mesmo passo que as fomenta.

REFERÊNCIAS BIBLIOGRÁFICAS

ATVARS, Teresa D. Z. **Escritório de Dados da Unicamp**. Campinas, 2020. Disponível em <https://metricas.usp.br/escritorio-de-dados-da-unicamp/>. Acesso em: 08 ago. 2025.

BARATA, André Montoia. **Governança de dados em organizações brasileiras: uma avaliação comparativa entre os benefícios previstos na literatura e os obtidos pelas organizações**. 2015. Dissertação (Mestrado) – Universidade de São Paulo, São Paulo, 2015. Disponível em: teses.usp.br/teses/disponiveis/100/100131/tde-28072015-215618/. Acesso em: 24 mar. 2025.

BARBIERI, Carlos. **ANÁLISE DA PESQUISA: O perfil das empresas brasileiras em gestão e governança de dados**. 2013. FUMSOFT.

BARBIERI, Carlos; FARINELLI, Fernanda. **Uma visão sintética e comentada do Data Management Body of Knowledge (DMBOK)**. 2013. 10.13140/RG.2.1.1561.2561.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. 1988. Brasília, DF: Presidente da República, [2025]. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em 02 fev. 2025.

BRASIL. **Lei nº 12.527**, de 18 de Novembro de 2011. 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; [...]. Brasília, DF, [2011]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 15 fev. 2025.

BRASIL. **DECRETO Nº 8.777, DE 11 DE MAIO DE 2016**. Institui a Política de Dados Abertos do Poder Executivo federal; [...]. Brasília, DF, [2016]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm. Acesso em: 12 jun. 2025.

BRASIL. **Lei nº 13.709**, de 14 de Agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF, [2018]. Disponível em:

https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em 22 jun. 2025.

BRASIL. **Plataforma Nacional de Infraestrutura de Pesquisa (PNIPE):** Dúvidas Frequentes, 2025. Disponível em <https://pnipe.mcti.gov.br/suport/faq>. Acesso em 14 jul. 2025.

CCMI INSTITUTE. CAPABILITY MATURITY MODEL INTEGRATION INSTITUTE. **DATA MANAGEMENT MATURITY (DMM) MODEL**. ISACA, 2019.

CHAPPELL, David A. **Enterprise service bus: Theory in practice**. "O'Reilly Media, Inc.", 2004.

DAMA. DATA MANAGEMENT ASSOCIATION INTERNATIONAL. **DAMA-DMBOK: Data management body of knowledge**. Technics Publications, LLC, 2017.

DUARTE, Rosália. **Entrevistas em pesquisas qualitativas**. Educar em revista, n. 24, p. 213-225, 2004.

FLICK, Uwe. **Qualitative Sozialforschung**. 2010.

Wilkinson, M. D. *et al.* **The FAIR Guiding Principles for scientific data management and stewardship**. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

HILDEBRAND, K.; GEBAUER, M.; HINRICHS, H.; MIELKE, M. **Daten- und Informationsqualität**. 6^a ed. Wiesbaden, Alemanha. Springer Vieweg, 2025.

INMON, William H.; STRAUSS, Derek; NEUSHLOSS, Genia. **DW 2.0: The architecture for the next generation of data warehousing**. 2010. Elsevier.

IBM. INTERNATIONAL BUSINESS MACHINES CORPORATION. **A Practical Guide to Managing Reference Data with IBM InfoSphere Master Data Management Reference Data Management Hub**. 2013.

JUNIOR, E. B., OLIVEIRA, G, S. SANTOS, A. C. O., SCHNEKENBERG, G. F. **Análise documental como percurso metodológico na pesquisa qualitativa**. Cadernos da FUCAMP, v. 20, n. 44, 2021.

KHATRI, Vijay; BROWN, Carol V. **Designing data governance**. Communications of the ACM, v. 53, n. 1, 2010.

KRIPKA, Rosana; SCHELLER, Morgana; BONOTTO, Danusa Lara. **Pesquisa Documental: considerações sobre conceitos e características na Pesquisa Qualitativa**. CIAIQ2015, v. 2, 2015.

MAIA, Maria Aniolly Queiroz. **SISTEMA DE INFORMAÇÃO DA PÓS-GRADUAÇÃO BRASILEIRA: AVALIAÇÃO DA PLATAFORMA SUCUPIRA**. 2020. Tese (Doutorado). Universidade Federal de Minas Gerais, Belo Horizonte, 2020.

MCGILVRAY, D. **Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information**. 2nd ed. Morgan Kaufmann, 2021.

MEC. MINISTÉRIO DA EDUCAÇÃO. **Coleção de Manuais de Preenchimento do Censo da Educação Superior 2024: Módulo Instituição de Educação Superior (IES)**, 2025. Disponível em https://download.inep.gov.br/pesquisas_estatisticas_indicadores_educacionais/ce_nso_da_educacao_superior/orientacoes/modulo_instituicao_de_educacao_superior_ies_2024.pdf. Acesso em 01 jul. 2025.

ONU. Organização das Nações Unidas. **Objetivos de Desenvolvimento Sustentável**. 2015. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 24 mar. 2025.

RÊGO, B. L., **Gestão e Governança de Dados: Promovendo dados como ativo de valor nas empresas**. ed. Brasport Livros e Multimídia Ltda, 2013

RIBEIRO, Ana Paula Tolentino. **Cenário de gestão de dados no SERPRO comparado à referência conceitual do guia DMBOK**. 2014.

SANTOS, D. J.; SOUZA, K. R. **A governança nas instituições de ensino superior públicas brasileiras**. Revista Ibero-Americana de Estudos em Educação, p. 1532–1557, 1 jul. 2022.

SEBASTIAN-COLEMAN, Laura. **Data Quality and Measurement**. MK Series on Business Intelligence. 2013.

TCU. TRIBUNAL DE CONTAS DA UNIÃO. **Levantamento de Governança e Gestão Públicas** 2017. 2017. Disponível em: <https://iesgo.tcu.gov.br/wp-content/uploads/sites/12/igg2017/ministerio/496-MPD-G-Ministerio-do-Planejamento-Desenvolvimento-e-Gestao.pdf>. Acesso em: 16 jun 2025.

TCU. TRIBUNAL DE CONTAS DA UNIÃO. **Esclarecimentos sobre uso dos resultados.** 2018. Disponível em

https://www.trt7.jus.br/files/aceso_informacao/transparencia/acoes_de_controle/TCU/esclarecimentoparausodosdados.pdf. Acesso em 16 jun 2025.

TCU. TRIBUNAL DE CONTAS DA UNIÃO. **Perfil Integrado de Governança Organizacional e Gestão Públicas** - 2021. Brasília, DF: TCU, 2021. Disponível em:

<https://iesgo.tcu.gov.br/wp-content/uploads/sites/12/igg2021/arquivos/Relatorio-consolidador-IGG2021-equipe.pdf>. Acesso em 16 jun 2025.

UNICAMP. UNIVERSIDADE ESTADUAL DE CAMPINAS. **Missão, Visão, Valores** - Escritório de Dados e Apoio à Transformação. 2025. Disponível em:

<https://dados.unicamp.br/missao-visao-valores/>. Acesso em: 08 ago. 2025

UFRJ. UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, **PDA** – Plano de Dados Abertos 2017 – 2019. 2017. Disponível em

https://ufrj.br/wp-content/uploads/2020/08/pdav2_-_ufrj.pdf. Acesso em: 12 jun. 2025.

UFRJ. UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, **Plano de Dados Abertos.** 2020. Disponível em

https://ufrj.br/wp-content/uploads/2021/11/PDA_UFRJ_rev_05112021.pdf. Acesso em: 12 jun. 2025. (Referente aos anos 2020-2022).

UFRJ. UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, **RELATÓRIO PLANO DE DADOS ABERTOS PERÍODO: 2020-2022.** 2022. Disponível em:

<https://ufrj.br/wp-content/uploads/2023/05/Relatorio-PDA-2020-2022-2.pdf>.

Acesso em: 12 jun 2025.

UFRJ. UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, **Plano de Dados Abertos** – UFRJ. 2023. Disponível em:

https://ufrj.br/wp-content/uploads/2023/04/PDA_UFRJ_rev_11-02-2023-1.pdf.

Acesso em: 12 jun 2025. (Referentes aos anos de 2023 a 2025).

UFRJ. UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, **RELATÓRIO PLANO DE DADOS ABERTOS PERÍODO: 2023-2024.** 2025. Disponível em

https://ufrj.br/wp-content/uploads/2025/04/Revisado.Relatorio_PDA_2023-2024_e_ditado.docx-1.pdf. Acessado em 12 jun 2025.

ZORZAL, Luzia. **Transparência das informações das universidades federais:** Estudo dos relatórios de gestão à luz dos princípios de boa governança na Administração Pública Federal. 2015. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, 2015.

APÊNDICE

QUESTIONÁRIO

Perguntas Gerais

1. Governança:

- a. Quais são os dados mestres e de referência você entende que são usados exclusivamente pelo seu sistema, poderia trazer exemplos? O que é feito para manter o padrão deles?
- b. Quais são aqueles que existem noutros sistemas? Como é mantida a consistência entre eles?
- c. Como os cargos são divididos dentro do seu sistema? Existe alguém que seja formalmente responsável por manter uma consistência desses dados dentro do seu sistema? Há diálogo com outros sistemas para que essa consistência seja padronizada?
- d. Quais fontes dos dados de referência externas ao seu sistema mas internos à UFRJ são usados? E externos à UFRJ?
- e. Com que frequência essas fontes são checadas por atualizações? Como é mantido essa consistência?
- f. Como os dados atualizados conseguem se relacionar com dados antigos? Existe alguma gestão de histórico ou tabela de alteração dos registros?
- g. Quais mecanismos formais são utilizados pelo seu sistema para garantir qualidades e consistência entre os dados mestre?
- h. Para quais sistemas ou entidades seus dados são externados?

2. Interoperabilidade e integração

- a. Quais outros sistemas – internos e externos à UFRJ – se relacionam com o seu? Como é feita essa integração?
- b. Como seu sistema identifica e lida com dados duplicados ou conflitantes internamente? E em relação a outros sistemas?

- c. Existem padrões de interoperabilidade para troca de dados?
- d. Há documentação formal sobre o fluxo de dados entre sistemas?
- e. Como a UFRJ lida com sistemas legados em relação a padrões de integração?

3. Qualidade dos dados

- a. Existe algum mecanismo de aferição de qualidade? Se sim, que tipo de ferramenta e método é usado?

4. Exemplos práticos

- a. Quais os desafios técnicos para implementar novas ferramentas e procedimentos em seu sistema? E os burocráticos?
- b. Cite um exemplo recente em que inconsistências nos dados impactou negativamente as operações do seu sistema. De que maneira foi resolvido?
- c. Quais são as prioridades a curto prazo para melhorar a gestão de dados mestre no seu sistema?

Perguntas específicas:

1. SIGA

- a. Você poderia me fornecer exemplos de dados do SIGA que coexistam com outros sistemas de forma que haja consistência entre eles?
- b. O SIGA é considerado a “melhor versão da verdade” (best version of truth/system of records) para os dados acadêmicos na UFRJ? Se não, qual compete com ele? Em quais âmbitos?
- c. Como são resolvidos os casos em que outros sistemas diferem do SIGA?

2. PR2

- a. Há replicação manual dos dados entre sistemas? Quais riscos isso pode trazer?
- b. Existem processos para automatizar a atualização dos dados dos orientadores ou bolsas em múltiplos sistemas?
- c. Como é feita a comunicação em tempo real entre sistemas da PR2 e SIGA? Há latência?

Perguntas finais:

1. Quais documentos ou manuais você sugeriria que eu consultasse para entender melhor a gestão atual de dados mestre e de referências da UFRJ?
2. Na sua opinião, qual o maior gargalo para melhorar a gestão de dados mestre?