GUIA DE DIRETRIZES E POLÍTICAS PARA GESTÃO DE DADOS

Volume 1: Dados de Referência

UFRJ

UNIVERSIDADE FEDERAL
DO RIO DE JANEIRO

Rio de Janeiro, 2025

GUIA DE DIRETRIZES E POLÍTICAS PARA GESTÃO DE DADOS, Volume 1: Dados de Referência

Versão 1.0

Autoria: Júlio Burlamaqui
Supervisão: Maria Luiza Campos

Data de Publicação: 19/08/2025

Universidade Federal do Rio de Janeiro

Sumário

Glossário

Tecnologia legada: Tecnologia desatualizada, ou não em conformidade com a atual.
Dado legado: Dado que era correto, mas agora encontra-se desatualizado.
Integração: Processo de combinar diferentes fontes em um único banco de dados.
Interoperabilidade: Capacidade de diferentes sistemas trocarem informações consistentes entre si.
Silo de informação: Sistemas isolados que não possuem integração e interoperabilidade.
Repositório: Local centralizado no qual informações são armazenadas e gerenciadas.
Histórico de versões: Registro de alterações feitas num arquivo específico.
Catálogo de dados: Um inventário detalhado especificando onde encontrar os dados nos repositórios, fornecendo-lhes metadados que os contextualizem.

Introdução

A Universidade Federal do Rio de Janeiro (UFRJ), como uma das maiores e mais importantes instituições de ensino superior do Brasil, enfrenta desafios significativos na gestão de seus ativos de informação. Este guia foi desenvolvido como produto final de uma monografia de conclusão do curso de Ciências da Computação.

Este documento inicia a série guias que possuem como objetivo a melhora da gestão de dados como um todo na UFRJ, sendo cada guia referente a uma parte específica da gestão de dados.

Este aborda a gestão de dados de referência, que é crucial para a padronização e consistência de informações em toda a universidade. Problemas como a ausência de sincronização, gestão manual, falta de clareza na propriedade dos dados e as inconsistências comprometem a qualidade dos dados mestres e, consequentemente, a tomada de decisões estratégicas e operacionais.

O presente documento propõe um modelo de governança de dados robusto, alinhado às melhores práticas do setor público brasileiro e às necessidades específicas de uma universidade federal, bem como às teorias acadêmicas do estado da arte revisadas no artigo irmão deste guia.

As recomendações centrais incluem a criação de um repositório centralizado e padronizado para dados de referência, a automação de processos por meio de interfaces de programação de aplicativos (APIs), o estabelecimento claro de papéis e responsabilidades de curadoria e propriedade de dados, e a integração sistêmica com plataformas nacionais. A implementação dessas diretrizes visa a não apenas resolver os problemas atuais, mas também a aprimorar a qualidade dos

dados, fortalecendo a governança institucional e otimizando a capacidade da UFRJ de cumprir sua missão acadêmica e social com maior eficiência e transparência.

A gestão eficaz de dados é um pilar fundamental para a governança e a eficiência operacional de qualquer grande organização, e as universidades federais brasileiras não são exceção. No contexto da UFRJ, a capacidade de gerenciar dados de referência de forma coesa e precisa é essencial para garantir a integridade de todas as informações institucionais, desde registros acadêmicos até dados financeiros e de pesquisa.

Este guia tem como propósito fornecer diretrizes e boas práticas para a gestão de dados de referência na UFRJ. Seu escopo abrange a abordagem de problemas identificados, como a ausência de sincronização, a gestão manual, a falta de clareza na propriedade e as inconsistências. Além disso, o documento servirá como um roteiro prático para a implementação de um programa de gestão de dados de referência na UFRJ.

A ideia também é a constante atualização das versões, conforme os problemas forem resolvidos, detalhando os níveis de sucesso, bem como novos problemas surgirem, promovendo soluções para estes. Além disso, a atualização se dará ao passo que a literatura avançará, desenvolvendo novos conhecimentos e técnicas eficazes na promoção de uma gestão de dados de referência de qualidade.

Conceituação

Para estabelecer uma gestão de dados eficiente na UFRJ, é fundamental compreender a natureza dos dados de referência e sua interconexão com os demais tipos de dados.

Metadados são informações técnicas a respeito dos dados, de quaisquer tipos que sejam, inclusive outros metadados. Assim, ele é a base de toda a gestão de base, por contextualizar cada categoria de dado.

São exemplos de informações contidas nos metadados regras de negócio, estrutura lógica e física dos bancos de dados, conexões entre atributos, características intrínsecas do surgimento do dado - como local ou data de criação -, explicação dos conceitos e usos, entre outros.

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. Esses dados são tipicamente de longo prazo, embora possam mudar ocasionalmente.

Exemplos comuns de dados de referência, no contexto específico da UFRJ, incluem códigos de cursos e DRE, tipos de vínculos, informações sobre unidades acadêmicas e administrativas, tipos de documentos acadêmicos, status do aluno, tipos de fomento, classificação de pesquisa (áreas de conhecimento), informações sobre os laboratórios, etc.

Por outro lado, os dados mestres representam entidades reais com a qual a instituição se relaciona. Possui informações precisas e essenciais para as transações comerciais críticas e o funcionamento de uma organização. Incluem informações permanentes ou semi-permanentes sobre entidades chave.

Como exemplo, podemos citar aluno (matrícula, nome completo, CPF, histórico acadêmico), servidores (SIAPE, cargo,

função, dados pessoais), fornecedores (CPF, CPNJ, informações, bancárias), projetos de pesquisa (título, coordenador, orçamento), dados de patrimônio (identificação de bens, localização).

Dados transacionais são as ações que a entidade realiza, ativa ou passivamente. Eles descrevem um evento ou transação interna ou externa que ocorre enquanto uma organização conduz seus negócios normalmente.

São exemplos de dados transacionais a edição de portarias, um aluno se matricular numa matéria, ou professor lançar nota, um servidor ser contratado, um funcionário bater o ponto, um centro promover um evento, o reitor emitir uma nota, um curso ser disponibilizado, etc.

Problemas identificados na UFRJ

A seguir, listamos alguns dos principais problemas encontrados em nossa instituição:

Replicação Manual

Talvez o mais básico dos problemas, e que mais tem consequências que se ramificam noutros problemas é a replicação manual de informações. Além de ser propensa a uma pletora de erros humanos, como erros de digitação, duplicatas, interpretações equivocadas, omissão de informações, etc. A replicação manual denota uma clara ineficiência operacional que se traduz em custos de manutenção elevados, tempo desperdiçado, e sobrecarga aos funcionários.

Outra consequência é a lentidão dos processos internos e externos, impedindo que, não só a universidade perca a capacidade de se avaliar em tempo real, mas também causa gargalo nas demandas discentes e docentes. Além disso, a manutenção manual das listas dificulta a escalabilidade dos sistemas, limitando o volume de dados.

Ausência de sincronização

Este é um processo crucial para a manutenção da consistência das informações entre os diferentes sistemas da UFRJ. Muitos desses sistemas armazenam informações que outros já armazenam, não só fazendo trabalho dobrado, como também, potencialmente, fomentando inconsistências.

Não há problema inerente em haver múltiplos sistemas, porém ele torna-se problemático quando estes não se comunicam, sendo chamados de silos. Nos silos, há a fragmentação da informação, pois cada unidade pode manter

sua própria versão dos dados de referência, com atualizações independentes e sem coordenação central.

Falta de clareza na competência dos dados

A competência dos dados se refere a quem cabe a responsabilidade da gestão, controle, manutenção, acesso, edição e distribuição dos dados dentro de uma instituição. É recorrente, na UFRJ, algum documento ou banco de dados não ter claramente um dono. Por vezes, essa ausência de competência resulta em documentos sem donos, outras vezes, um banco de dados com sobrecarga de competências.

As consequências desse problema são muitas, em várias áreas. A ambiguidade sobre a responsabilidade pode acarretar na falta de manutenção, corroendo a qualidade dos dados. No outro extremo, há mudanças constantes no que deveria ser um banco de dados relativamente estático. Não só há a necessidade de atribuir competências aos documentos e bancos de dados a setores e sistemas, como também a criação ou nomeação de uma autoridade que arbitre sobre essas disputas.

Conceitos definidos extraoficialmente

A definição formal, vinda de posições elevadas da gestão, é importante para viabilizar a comunicação e compreensão entre sistemas, fomentando a integração e interoperabilidade. A inexistência ou inobservância dela cria léxicos paralelos, isolando cada sistema em seu próprio silo. Isso torna o compartilhamento de informações difícil porque funcionários de um sistema podem não entender as informações que lhes foi passada de outro sistema.

Medidas para melhoria da Gestão dos Dados de Referência

A Cartilha de Governança de Dados do Poder Executivo Federal estabelece um *framework* de governança de dados com regras e processos sobre como as instituições da Administração Pública devem gerenciar seus dados. Partimos dele como base, e adaptamo-nos às nossas necessidades. As seguintes medidas visam a institucionalização de políticas que irão primariamente mitigar problemas relacionados à má gestão dos dados de referência, mas também, que no médio e longo prazo, impactarão a gestão de dados holisticamente dentro da Universidade Federal do Rio de Janeiro.

Criação do Comitê de Dados

Precisamos formalizar um comitê ou grupo de trabalho interdisciplinar para traduzir suas diretrizes abstratas e genéricas em políticas internas concretas e acionáveis. Para compor esse comitê, poderíamos nos voltar ao grupo de trabalho interdisciplinar já existente, montado para a realização do Plano de Dados Abertos da Universidade Federal do Rio de Janeiro.

Apesar desta medida não se relacionar diretamente com dados de referência, é imperativa para iniciar uma gestão de dados no geral corretamente. Sem ela, qualquer medida implementada seria apenas uma solução provisória sem plano de gestão continuada.

A este comitê, caberia 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.

Instituição de Cargos de Dados

Essa atribuição bem definida de funções não é uma mera formalidade, já que será incumbido a eles o poder-dever de garantir conformidade às regulações, bem como de zelar pelos dados e promover a cultura dos dados. Ao comitê, cabe a escolha dos funcionários para ocupar os cargos, mas advogamos para a existência da seguinte hierarquia:

- **Executivo de Dados:** É a maior liderança no que tange aos dados. Ele deve estar antenado às inovações tecnológicas e burocráticas, garantindo modernidade e eficiência à universidade. Nossa recomendação para essa posição é a TIC, ou reitoria, ou o comitê mencionado anteriormente.
- **Proprietário dos Dados:** É o responsável pela gestão, controle e manutenção de um ativo de dados 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. Nossa recomendação para essa posição são as pró-reitorias, ou os diretores dos centros ou institutos, ou os chefes de departamento.
- **Curador de Dados:** São responsáveis pela manutenção da cultura de boas práticas de gestão de dados de referência na UFRJ. 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 de referência estejam limpos, consistentes, precisos,

atualizados e compartilháveis. Nossa recomendação para essa posição são os integrantes do NCE ou SIGA.

- **Guardião dos Dados:** São responsáveis 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. Nossa recomendação para essa posição é ter um guardião por sistema da UFRJ.

Centralização, integração e interoperabilidade:

Além do comitê e de funções especializadas em dados, precisamos realizar algumas mudanças estruturais e culturais sobre como lidamos com dados na nossa universidade. Portanto, proponho que adotemos um estilo de gestão de dados consolidado. Ele oferece diversos benefícios, como uma visão unificada, foco analítico e a possibilidade de cuidar da qualidade dos dados. Tudo isso sem o impacto disruptivo que estilos como coexistência ou centralizado causam.

Para que seja possível realizar tal mudança, precisamos definir e distinguir, com a direção do nosso comitê - em particular dos guardiões de dados -, quais são os dados mestre e de referência. Para tanto, precisamos dar continuidade aos nossos esforços de mapeamento de sistemas e fontes de dados, bem como das competências. Finalmente, devemos decidir como estruturar nosso modelo de dados, definindo as regras de negócio e hierarquia de dados.

Homogeneização dos termos:

Além de adotar um estilo consolidado de gerir os dados, devemos também desenvolver um dicionário de termos, ou glossário de negócios, unificado entre os sistemas. Podemos nos inspirar nos já estabelecidos em nível nacional, por exemplo o CENSUP do INEP, Sucupira do CAPES ou a plataforma Lattes do CNPq, e fora da área da educação, podemos citar também, o Vocabulário Controlado Básico (VCB) do Senado Federal. Essa medida padroniza os conceitos de forma clara e una entre os sistemas, auxiliando a integração e compartilhamento de informações, fomentando interoperabilidade e comunicação.

Ele é a ferramenta que dá suporte à governança de dados. Ele serve como um documento oficial que define a semântica de cada dado importante na universidade, como "Aluno Ativo" ou "Curso de Graduação". Ao ter um significado unificado para cada termo, você elimina a ambiguidade e garante que todos, dos desenvolvedores de TI aos gestores, entendam e usem os dados da mesma forma.

Regras de Nomeação:

As regras de nomeação garantem que todos os dados de referência sejam identificados de forma consistente. Isso evita a confusão e o retrabalho. Se relaciona diretamente com o exposto acima. É preciso definir padrões claros para os nomes das categorias e dos valores dentro delas. Por exemplo, deve-se decidir se os nomes devem ser todo em maiúsculas, ou minúsculas, ou camelCase, ou snake_case, etc. Também precisa-se criar convenções para siglas e abreviaturas, se haverá pontos para separar as letras, ou não (como em SP vs S.P. para São Paulo).

Podemos pegar a dica passada pela gestão de dados do Operador Nacional de Serviços Elétricos ao adotar convenções

de mnemônicos definidos globalmente na organização. Com eles, pode-se entender rapidamente qualquer informação de qualquer setor. Por exemplo, tb_nomeDaTabela para tabelas, ou vw_nomeDaView para views, etc.

Gerir as Categorias de Dados

Em vez de simplesmente organizar e estruturar os dados de referências em listagens de valores, deveríamos agrupá-los em estruturas de categorias lógicas. Por exemplo, no lugar de ter uma longa listagem de todos os cursos da UFRJ, seria interessante dividi-los em famílias, como “Cursos de Graduação”, “Cursos de Pós-Graduação”, “Cursos de Especialização”, etc. Essa Abordagem pode ser utilizada noutros contextos.

Além do agrupamento de dados comuns, poderíamos criar hierarquias, imbuindo conhecimento prévio nos dados. Por exemplo, os dados de referências geográficas podem ser hierarquizados da seguinte maneira: País -> Estado -> Cidade -> Bairro -> Logradouro. Isso torna a navegação e compreensão mais acessível.

Automatização:

Este busca mitigar diretamente o problema de inserção manual dos dados, garantindo a coleta e atualização automatizada das fontes de dados internos e externos. Isso poderá ser obtido através de APIs (Application Programming Interface) que integrem em tempo real nossos sistemas com bancos de dados que nos forneçam dados de referência, como os dos Correios, IBGE, MEC, etc.

Os catálogos de APIs governamentais do Portal do Governo

Federal (https://www.gov.br/conecta/catalogo/) e do Tesouro Transparente (https://www.gov.br/tesouronacional/pt-br/central-de-conteudo/apis) nos fornecem um caminho de como documentar e disponibilizar as APIs públicas. Além de bem documentadas, elas devem ser seguras, confiáveis e capazes de lidar com grandes volumes de dados.

Elas permitem que os sistemas se comuniquem diretamente, garantindo que os dados de referência estejam corretos de forma rápida, segura e confiável, sem intervenção humana, fomentando assim, a eficiência operacional.

Sincronização:

Essa se relaciona diretamente com a utilização de APIs, pois as fontes de dados devem ser coletadas e atualizadas periodicamente. A falta de sincronização acarreta em conflitos de informações, retrabalho e interpretações desatualizadas. Devemos priorizar a sincronização e atualização periódica em bases de dados críticas.

A sincronização com dados de referência de bases externas do Governo Federal facilitarão a integração com outras instituições e plataformas, agilizando e facilitando relatórios e auditorias externas e internas. Existem muitos métodos de fazer uma sincronização que leve em conta a necessidade de ter informações atualizadas com o custo computacional de conferir por atualizações. A escolha da periodicidade e dos métodos e tecnologias usadas é de critério do comitê, variando de acordo com a criticidade da informação recebida e com a expectativa de mudança dela. A listagem de cursos é central para universidade, mas não muda com muita frequência, por outro lado, taxa de câmbio muda diariamente, mas não é tão relevante à nossa

entidade. Deve ser debatida e acordada a frequência das atualizações das fontes dos dados de referência.

Criação de um repositório para dados históricos:

Com a atualização de informações, surge o novo desafio da gestão de dados legados. O que fazer com registros antigos, que usam terminologias antigas, ou contêm classificações que não existem mais, como cidades que deixaram de existir, ou situações jurídicas extintas? Devemos ter protocolos e mecanismos para lidar com esses dados, fazendo algum tipo de recepção e equivalência, modernizando-os, não-obstantemente, mantendo sua essência. Essas equivalências precisam ser muito bem documentadas nos metadados.

Dessa forma, precisamos implementar um repositório, como um data warehouse, que permita o armazenamento de dados que foram substituídos, mas que ainda sejam úteis para auditoria, contextualização histórica ou análise temporal.

Catálogo de Dados

Um catálogo de dados é uma ferramenta que funciona como um inventário de todos os dados disponíveis na universidade. Ele documenta metadados, como quem é o dono, qual a origem, qual a definição, e como os dados são usados. Em vez de depender de solicitações manuais ou de saber por experiência onde um dado está, um Catálogo de Dados permite que um usuário procure por "dados de matrícula" e encontre rapidamente os conjuntos de dados relevantes, quem são os responsáveis e como eles podem ser acessados.

Além disso, devido aos metadados contidos, o catálogo reforma a semântica, contextualizando e explicando os dados. Quando um termo como "servidor contratado" é buscado, o catálogo mostra a definição oficial do **Dicionário de Termos**, quem é o responsável por essa definição, e quais sistemas a utilizam. Isso evita interpretações erradas e garante a consistência em relatórios e análises.

Definição de Políticas de Qualidade dos Dados

Na verdade, devemos ter mecanismos e regras para resolução de conflitos no geral, não apenas com dados legados. Isso se relaciona com semântica global, porque além de defini-la, precisamos também estabelecer uma política de qualidade de dados para superar dados conflitantes. Isso inclui a definição de métricas de qualidade, regras de validação, e protocolos de correção de erros e aprovação das alterações de valores. A qualidade dos dados deve ser um objetivo mensurável e gerido ativa e continuamente.

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.

BRASIL. **Cartilha de Governança de Dados**: Volume III - Papéis e responsabilidades de Governança de Dados no Poder Executivo Federal. 2024. Disponível em: https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/governancadedados/arquivos/CartilhaGovDadosvol3.pdf . Acesso em: 25/07/2025.

BRASIL. **Manual para inclusão de Solicitação de Adesão Conecta gov.br**. 2024. Disponível em: https://docs.google.com/document/d/1a2kTNOZuNgnKAqpztm-quQdB1XO5tDtX/view?tab=t.0. Acesso em: 03/08/2025.

BURLAMAQUI, Júlio. **Gestão de Dados de Referência:** Um estudo de caso e proposta de arquitetura para a Universidade Federal do Rio de Janeiro. 2025.

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

DATAVERSE PROJECT. **Guia de Usuário.** Disponível em: https://guides.dataverse.org/en/4.20/user/. Acessado em: 27/07/2025.

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

UNIVERSIDADE ESTADUAL DE CAMPINAS. **Deliberação CONSU-A-050/2020, de 06/10/2020.** Institui a Política Institucional de Acesso Aberto à Produção Intelectual e Científica da Universidade Estadual de Campinas e estabelece os repositórios oficiais de depósito das produções. 2020.

UNIVERSIDADE ESTADUAL DE CAMPINAS. **Tutorial para o Repositório de Dados de Pesquisa da UNICAMP.** Disponível em: https://www.sbu.unicamp.br/sbu/wp-content/uploads/2025/07/Redu-Tutorial.pdf. Acesso em: 27/07/2025.

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