sbbd:02

Capítulo 2

Modelagem de Bancos de Dados em Conformidade com a LGPD

Patricia Vieira da S. Barros, José Maria Monteiro, Javam C. Machado

Departamento de Computação (DC) – Universidade Federal do Ceará (UFC)

CEP 60440-900 – Fortaleza – CE – Brasil

{patricia.barros, jose.monteiro, javam.machado}@lsbd.ufc.br

Abstract

The Brazilian General Data Protection Law (Law No. 13,709/2018) establishes the rules for the processing of personal data in Brazil, applying to operations carried out by natural or legal persons in both physical and digital environments. Its objective is to protect fundamental rights, particularly freedom and privacy, by regulating the entire data lifecycle, from collection to disposal, requiring that data processing be based on specific legal grounds, such as the data subject's consent. In this context, information systems must comply with the requirements of the current legislation by incorporating compliance requirements into software development. The database becomes a central component in this process, as it is responsible for the storage, integrity, and availability of information. Therefore, this work provides a methodology to integrate the legal requirements of the LGPD into database design, facilitating system development and the execution of legal compliance audits.

Resumo

A Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018) estabelece as regras para o tratamento de dados pessoais no Brasil, aplicando-se a operações realizadas por pessoas físicas ou jurídicas, tanto no meio físico quanto digital. Seu objetivo é proteger direitos fundamentais, especialmente a liberdade e a privacidade, regulando todo o ciclo de vida dos dados, da coleta ao descarte, exigindo que o tratamento esteja fundamentado em bases legais, como o consentimento do titular. Nesse cenário, o sistema de bancos de dados torna-se um componente central, uma vez que é responsável pelo armazenamento, pela integridade e pela disponibilidade dos dados. Assim, este minicurso discute estratégias para integrar os requisitos jurídicos da LGPD ao projeto de bancos de dados, facilitando o desenvolvimento de sistemas e a realização de auditorias de conformidade com a LGPD.

2.1. Introdução e Motivação

A Lei Geral de Proteção de Dados (LGPD), Lei 13.709/18[1] regulamenta a forma pela qual as empresas podem utilizar os dados pessoais enquanto informação relacionada à pessoa natural identificada (ou identificável), além de determinar como devem ser realizados o tratamento, o armazenamento e o descarte de dados pessoais, buscando proteger os direitos fundamentais de liberdade e privacidade. Essa legislação impõe uma profunda transformação no sistema de proteção de dados no país.

A LGPD é uma legislação que envolve mudanças de processos, atualizações de documentos e contratos nas organizações e, principalmente, uma mudança de cultura no dia a dia das empresas na forma de tratar os dados pessoais. A lei brasileira inova ao tratar de questões não satisfatoriamente abordadas por outras leis setoriais de proteção de dados no Brasil. Como exemplos, tem-se uma definição mais precisa do conceito de dados pessoais, uma previsão expressa das bases legais que autorizam o tratamento de tais dados, um cuidado no processamento de dados públicos, a criação da Autoridade Nacional de Proteção de Dados (ANPD), a definição de sanções e uma maior segurança jurídica para os detentores de dados pessoais.

Por outro lado, os Sistemas de Informação (SI) estão fortemente baseados na aquisição, armazenamento e processamento de dados. Logicamente, esses sistemas necessitam estar em conformidade com a LGPD. Desta forma, a Lei proporciona um grande impacto no desenvolvimento de sistemas de informação, os quais devem agora tratar os dados pessoais da forma estipulada pela legislação, de maneira mais formal, voltando uma maior atenção para o ciclo de vida dos dados, visto que este envolve todas as operações realizadas sobre as informações obtidas por uma empresa ou instituição, desde sua coleta até a sua devida destruição. Mais formalmente, o ciclo de vida dos dados compreende todo o período no qual os dados pessoais são manipulados por uma determinada entidade (pessoa física ou jurídica).

Neste contexto, o Sistema de Bancos de Dados (SBD) passa a ser um componente ainda mais importante no desenvolvimento de *software*, uma vez que este é responsável pelo armazenamento, atualização e recuperação dos dados. Mais especificamente, um Banco de Dados (BD), que representa um conjunto de dados e seus inter-relacionamentos, deve estar aderente à LGPD. Assim, por exemplo, o resultado de um teste para o Vírus da Imunodeficiência Humana (HIV) (o qual é utilizado para diagnosticar uma infecção causada pelo vírus da imunodeficiência humana) deveria ser armazenado de forma criptografada, inviabilizando o seu acesso inclusive para os profissionais da área de banco de dados e desenvolvedores de sistemas.

Uma das alternativas para assegurar a conformidade de um banco de dados com a LGPD consiste em incorporar, já no processo de projeto de bancos de dados, os requisitos e as restrições impostos pela legislação. Desta forma, a principal contribuição deste minicurso consiste em fornecer uma estrutura metodológica, conceitual e prática que habilite profissionais, pesquisadores e auditores a projetarem esquemas de dados intrinsecamente aderentes à lei. O objetivo central é capacitar os participantes a mapearem

[1] `https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm`

e representarem formalmente os principais atores e fluxos jurídicos (como titulares, dados sensíveis, propósitos, compartilhamentos e consentimentos) diretamente na arquitetura do banco de dados, simplificando o ciclo de desenvolvimento de software e otimizando os processos subsequentes de auditoria legal, com o auxílio da ferramenta BrModeloPD[2], uma extensão da ferramenta brModelo[3] que possibilita incorporar as imposições e preceitos da LGPD ao projeto de bancos de dados, conforme ilustrado em [Barros et al. 2024]. O endereço da referida ferramenta encontra-se em [4].

2.2. Dados, Informação e Ciclo de Vida dos Dados

Inicialmente, torna-se necessário apresentar os conceitos de dados, informação, tipos de dados e ciclo de vida dos dados. Adicionalmente, esta seção discute a importância dos dados para a sociedade atual.

2.2.1. Dados X Informação

Dados são uma coleção de valores discretos que transmitem informações, descrevendo quantidade, qualidade, fatos, estatísticas, outras unidades básicas de significado. Assim, dado é o registro do atributo de um ente, objeto ou fenômeno. Na ciência da computação, dados são quaisquer sequências de um ou mais símbolos que podem ser interpretados. Um *datum* é um único símbolo em uma sequência, ou seja, um valor individual em um determinado dado. Com isso, os dados digitais são dados representados usando o sistema numérico binário que requerem interpretação para se tornarem informações. Já a informação é um conjunto de dados onde cada dado é contextualizado pela relação que ele possui com os demais dados desse conjunto.

Atualmente, a informação é o ativo mais valioso de uma organização, e por isso, está sujeito a inúmeras ameaças tanto do ambiente interno quanto externo, as quais podem explorar vulnerabilidades e, assim, comprometer as operações de negócio da empresa. Portanto, qualquer organização, pública ou privada, depende da informação para seus processos decisórios, e dificilmente poderá funcionar adequadamente sem uma quantidade significativa de informação e do conhecimento por ela proporcionado [Fontes 2012].

2.2.2. Ciclo de Vida dos Dados

É o processo que descreve o fluxo dos dados dentro de uma organização, desde o momento em que os dados são coletados até o arquivamento ou eliminação desses dados, conforme ilustrado na Figura 2.1, podendo ser subdividido em [SANTANA 2016]: origem, retenção, uso, publicação, eliminação e retificação. A etapa denominada **Origem** dos dados refere-se às diferentes formas de produção ou de recepção dos dados, seja em formato físico ou eletrônico. A etapa de **Retenção** inclui o armazenamento de dados em diversos meios (arquivo, banco de dados, documento físico ou digital). A etapa denominada **Uso** envolve todas as ações e manipulações que podem ser realizadas sobre os dados armazenados, como classificação, reprodução, processamento, avaliação ou controle, bem como possíveis modificações, produzindo informações para as tarefas que a empresa precisa

[2] https://github.com/brmodeloweb/brmodelo-app
[3] https://www.brmodeloweb.com/
[4] http://200.129.44.243:9000/

executar e gerenciar. A etapa de **Publicação** dos dados envolve as operações de transmissão, distribuição, comunicação, transferência e difusão de dados pessoais para um local fora da empresa, bem como o uso compartilhado desses dados. Esta etapa pode ou não fazer parte do ciclo de vida de uma determinada organização. A etapa de **Eliminação** consiste na exclusão de dados arquivados. Algumas vezes, isso é feito para que as empresas liberem espaço de armazenamento. Por fim, a etapa de **Retificação** dos dados refere-se ao conjunto de procedimentos destinados à correção, atualização ou complementação de dados pessoais inexatos, incompletos ou desatualizados [Amaral 2016].

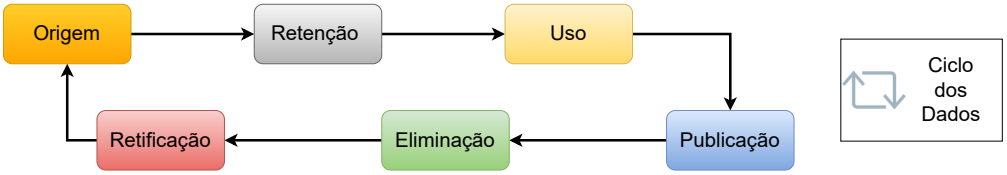


Figura 2.1: Ciclo de Vida dos Dados.

Fonte: A autora.

A LGPD estabelece diretrizes para o tratamento de dados pessoais, desde a coleta até o armazenamento, o processamento e a exclusão. Portanto, a LGPD traz impactos diretos à todas as etapas do ciclo de vida dos dados [5].

2.3. Lei Geral de Proteção de Dados - LGPD

A Lei n° 13.709 de 14 de agosto de 2018, denominada Lei Geral de Proteção de Dados Pessoais - LGPD, foi idealizada com o objetivo de criar um âmbito legal para a proteção dos dados pessoais da população brasileira. A partir do momento em que outros países iniciaram suas legislações específicas sobre a proteção de dados, tal como a *General Data Protection Regulation* (GDPR) na União Europeia, verificou-se a necessidade do Brasil também elaborar uma legislação acerca da proteção de dados, visto que sua inexistência poderia resultar em uma perda importante de competitividade internacional[6].

Recentemente, observou-se um aumento exponencial no fluxo de transações envolvendo dados pessoais, o que acabou por criar ramos de negócios inteiramente novos e aumentar a eficiência de diversos setores da economia. Nesse mesmo sentido, a LGPD impacta todas as áreas da sociedade e sua promulgação proporciona maior estabilidade e segurança jurídica aos diversos ramos de negócios existentes e aos que deverão surgir posteriormente, decorrentes da transformação digital sem precedentes que estamos vivenciando. A LGPD é uma legislação que tem um imenso impacto econômico, social e regulatório, e cuja implementação nas empresas e nos órgãos públicos não é uma tarefa simples, considerando a novidade que este tema representa para a sociedade brasileira [Pinheiro 2020].

[5] https://www.lgpdbrasil.com.br/lgpd-e-o-ciclo-de-vida-dos-dados-pessoais

[6] http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm

O Art. 2° da LGPD define como fundamentos: respeito à privacidade; autodeterminação informativa; liberdade de expressão, de informação, de comunicação e de opinião; inviolabilidade da intimidade, da honra e da imagem; desenvolvimento econômico, tecnológico e inovação; livre iniciativa, livre concorrência e a defesa do consumidor; direitos humanos, o livre desenvolvimento da personalidade, a dignidade e o exercício da cidadania pelas pessoas naturais. Já o Art. 6° da LGPD estabelece um conjunto de dez princípios: Finalidade, Adequação, Necessidade, Livre Acesso, Qualidade dos Dados, Transparência, Segurança, Prevenção, Não-Discriminação e Responsabilização e Prestação de Contas.

Adicionalmente, vale destacar que para a LGPD, **Tratamento** é qualquer operação realizada com dados pessoais (Art. 5°, X). A LGPD também estabelece **punições** para infrações que envolvam incidentes de segurança de dados (Arts. 52 a 54).

2.3.1. Principais Atores da LGPD

O **Titular dos Dados Pessoais** é o *ator principal*, visto que são estabelecidos os princípios e garantias da LGPD em seu benefício. Segundo o Artigo 5º, V, o titular dos dados é *"a pessoa natural a quem se referem os dados pessoais que são objeto de tratamento"*. Com isso, a proteção conferida pela LGPD é voltada às pessoas físicas, não alcançando os dados das pessoas jurídicas, e compreende os dados que possam identificar ou tornar identificável uma pessoa.

A LGPD especifica como *Agentes de Tratamento*: controlador, operador e encarregado de dados pessoais. O **Controlador** é uma pessoa natural ou jurídica, de direito público ou privado, a quem ***competem as decisões*** referentes ao tratamento de dados pessoais, conforme estabelece o Art. 5°, VI da LGPD. É ele quem decide sobre o tratamento de dados pois possui autonomia, sendo uma figura obrigatória. O **Operador** é a pessoa natural ou jurídica, de direito público ou privado, ***que realiza o tratamento*** de dados pessoais em nome do controlador, como especifica o Art. 5°, VII da LGPD. É quem executa o tratamento de dados a pedido do controlador, sob ordens lícitas, sem autonomia. O **Encarregado**, como definido no Art. 5°, VIII da LGPD é a pessoa indicada pelo controlador e operador para *atuar como canal de comunicação* entre o controlador, os titulares dos dados e a **ANPD**, que é o *órgão da administração pública responsável por zelar, implementar e fiscalizar o cumprimento desta Lei em todo o território nacional* (Art. 5º, XIX).

2.3.2. Tipos de Dados

No seu Art. 5°, incisos I, II e III, a LGPD diferencia 3 (três) tipos de dados: Dado Pessoal, Dado Sensível e Dado Anonimizado. **Dado Pessoal** são aqueles relativos à pessoa física identificada ou que possa ser identificada com o cruzamento de duas ou mais informações. **Dado Sensível** é um dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural. **Dado Anonimizado** é o dado relativo ao titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento. A *anonimização* é uma técnica de processamento de dados

que remove ou modifica informações que possam identificar a pessoa, garantindo sua desvinculação. Nestes casos, a LGPD não se aplicará ao dado. Mas, ressalta-se que o dado somente é considerado anonimizado se não permitir que, por meios técnicos ou outros, seja reconstruído o caminho para revelar quem é o titular do dado. Se a identificação ocorrer, não se tratará de dado anonimizado, mas sim de dado pseudonimizado, e estará sujeito à LGPD, como descrito no Art. 12, da LGPD.

2.4. Lacunas entre a LGPD e os Sistemas de Banco de Dados

Apesar da ampla difusão da LGPD no contexto organizacional brasileiro, observa-se um hiato significativo entre os requisitos legais estabelecidos e sua efetiva incorporação ao processo de projeto de bancos de dados. Enquanto a LGPD define princípios, direitos e obrigações relacionados ao tratamento de dados pessoais, as metodologias tradicionais de modelagem de bancos de dados concentram-se em aspectos funcionais, estruturais e de desempenho, sem oferecer procedimentos explícitos para representar requisitos de privacidade, consentimento, finalidade e retenção de dados. Com isso, a adequação à LGPD, em geral, ocorre após o projeto, a construção e a disponibilização dos sistemas de bancos de dados, elevando os custos de implementação e reduzindo a eficácia das medidas de proteção de dados. Desse modo, evidencia-se a necessidade de metodologias capazes de integrar requisitos legais e técnicos, desde as fases iniciais do projeto de bancos de dados, promovendo uma abordagem alinhada aos princípios de Privacidade por Projeto (*Privacy by Design*) [Cavoukian et al. 2009] e à conformidade regulatória.

2.5. Integrando a LGPD aos Sistemas de Banco de Dados

Esta seção tem por objetivo apresentar as estratégias existentes que possibilitam incorporar, de alguma forma, os requisitos e as restrições da LGPD aos sistemas de bancos de dados. Uma das estratégias existentes é denominada Privacidade por Projeto (*Privacy by Design*), que busca incorporar mecanismos de privacidade desde as fases iniciais do desenvolvimento de sistemas, processos e tecnologias, garantindo a proteção de dados de forma preventiva e integrada. Desta forma, esta estratégia busca garantir a privacidade dos dados desde a fase de projeto dos bancos de dados. Esse conceito estabelece que a privacidade deve ser considerada um requisito fundamental de projeto, orientando a definição de processos, arquiteturas e tecnologias que garantam a proteção dos dados pessoais ao longo de todo o ciclo de vida dos dados [Cavoukian et al. 2009].

Em [Sarkar and Athanassoulis 2022], os autores propõem uma extensão às linguagens de consulta de banco de dados que permite especificar políticas de exclusão de dados diretamente nas consultas. Assim, os desenvolvedores podem definir regras para a exclusão automática de dados com base em critérios, como o tempo de existência dos dados ou outros atributos relevantes. A exclusão automática de dados é fundamental no contexto da LGPD, ma vez que os dados pessoais não podem ser armazenados por um tempo indefinido. Em [de Abreu et al. 2021], os autores introduzem extensões à linguagem SQL para incorporar considerações de consentimento no processamento de consultas, permitindo que os desenvolvedores expressem explicitamente, nos comandos SQL, as condições sob as quais os dados podem ser acessados e utilizados.

Já [Vieira et al. 2026] aborda a associação entre os requisitos da LGPD e o processo de projeto de bancos de dados. Os autores mostram a existência de um lapso entre as exigências legais de proteção de dados e as metodologias tradicionais de modelagem, que normalmente não contemplam aspectos como a finalidade, o consentimento, a minimização e a retenção de dados. Para reduzir essa lacuna, propõe-se uma abordagem que visa incorporar requisitos e restrições da LGPD desde as fases iniciais do desenvolvimento, promover a conformidade regulatória, aumentar a transparência no tratamento de dados pessoais e apoiar a construção de sistemas mais seguros.

2.6. A Metodologia LGPDbyD

Essa seção apresenta a metodologia **LGPDbyD**, que tem por objetivo incorporar os requisitos e as restrições da LGPD ao projeto de bancos de dados. Para isso, a metodologia estende os conceitos e as notações comumente utilizadas nos projetos conceitual, lógico e físico, com a ideia central de alterar os modelos existentes o mínimo possível. Além disso, a LGPDbyD facilita os processos de desenvolvimento de sistemas e de auditoria de conformidade com a LGPD, uma vez que os principais conceitos da legislação são tratados no nível do sistema de banco de dados.

2.6.1. Projeto Conceitual

A metodologia LGPDbyD inicialmente propõe uma adaptação do modelo Entidade-Relacionamento (ER), denominada ER-PD, com o objetivo de viabilizar o projeto conceitual de bancos de dados em conformidade com a LGPD. Esta extensão permite representar os conceitos-chave definidos pela LGPD, tais como: dados pessoais, tipos de operações de processamento de dados e o titular dos dados.

O **Titular dos Dados Pessoais**, conforme a própria LGPD especifica em seu Artigo 5°, refere-se ao sujeito a quem a Lei pretende proteger. Portanto, o **Titular dos Dados Pessoais** é um conceito central no modelo ER-PD, sendo representado como um tipo particular de conjunto entidade, o qual indica a presença de atributos pessoais, os quais devem ser particularmente protegidos. A notação utilizada para representar esse tipo específico de conjunto entidade, denominado "Titular", é um **retângulo com linhas tracejadas** (Figura 2.2).

Segundo a LGPD, dentre os dados pessoais alguns são considerados "sensíveis", os quais devem possuir um tratamento específico, como destacado em seu Artigo 11. Ainda segundo a LGPD, uma das formas de tratar os dados sensíveis é a anonimização. A técnica de criptografia não é mencionada em nenhum momento na LGPD, mas é uma das alternativas comumente utilizadas para assegurar a pseudo-anonimização de dados. A criptografia transforma dados legíveis em dados cifrados, protegidos por uma chave. Enquanto existir a possibilidade de reverter o processo com essa chave, os dados continuam sendo, em princípio, reidentificáveis. Portanto, eles não estão verdadeiramente anonimizados (em sentido estrito); estão protegidos ou pseudonimizados. No contexto da LGPD, dado anonimizado é aquele que não permite a identificação do titular, considerando os meios técnicos razoáveis disponíveis. Já a criptografia, por ser reversível quando há chave, normalmente não elimina definitivamente o vínculo com o titular. Portanto, a criptografia contribui para a proteção e a pseudonimização de dados, mas não caracteriza,

por si só, anonimização quando houver possibilidade de reversão ou reidentificação.

Com a finalidade de representar o fato de um atributo armazenar um dado pessoal, bem como o tratamento que deve ser realizado sobre esse dado, o modelo ER-PD propõe a utilização de nove novos tipos de atributos (Figura 2.2), sendo os principais: atributo Pessoal" (P), atributo Sensível" (S), atributo Anonimizado" (A) e atributo Criptografado" (C). Além disso, o modelo ER-PD introduz dois tipos adicionais de atributos: atributo Identificador" (I) e atributo "Semi-Identificador" (SI), para representar conceitos comumente utilizados no domínio de privacidade de dados. Um atributo **Identificador** é aquele que pode identificar unicamente um indivíduo—como Cadastro de Pessoas Físicas (CPF), nome ou endereço eletrônico. Um atributo **Semi-Identificador**, por sua vez, não identifica explicitamente um indivíduo isoladamente, mas pode fazê-lo quando combinado com outros atributos [Brito and Machado 2017]. Exemplos incluem data de nascimento e Código de Endereço Postal (CEP). É importante notar que, em geral, quando a privacidade de atributos pessoais ou sensíveis precisa ser preservada, técnicas de anonimização ou criptografia são aplicadas.

A LGPD estabelece regras específicas para o processamento de dados relacionados a **Crianças e Adolescentes**, no Artigo 14 e seus parágrafos. Para representar esta característica, o modelo ER-PD adiciona um novo tipo de atributo denominado "Criança e Adolescente" (CAD). Por fim, o titular dos dados pode autorizar o compartilhamento de seus dados, ou partes deles, com terceiros. Para representar este requisito, o modelo ER-PD propõe um novo tipo de atributo, o atributo "Compartilhado" (CP). A Figura 2.2 ilustra as notações introduzidas pelo modelo ER-PD.

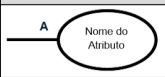
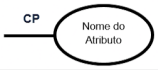
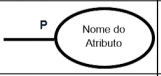
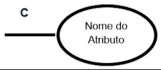
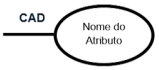
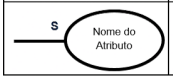
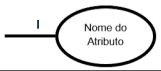
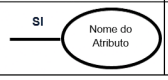
Símbolo	Representação	Símbolo	Representação	Símbolo	Representação
	Titular dos Dados Pessoais		Atributo Anonimizado		Atributo Compartilhado
	Atributo Pessoal		Atributo Criptografado		Atributo de Criança e Adolescente
	Atributo Sensível		Atributo Identificador		Atributo Semi-Identificador

Figura 2.2: Notação do Modelo ER-PD.

Fonte: A autora.

2.6.1.1. Exemplo de Execução

A seguir, será ilustrado um exemplo de aplicação do modelo ER-PD em todas as fases no projeto de bancos de dados em conformidade com a LGPD. Inicialmente, será mostrado o uso do modelo ER-PD no projeto conceitual de bancos de dados em conformidade com a LGPD. Desta forma, considere que uma clínica médica deseja projetar um banco de dados utilizado para armazenar informações sobre pacientes e exames médicos. Assuma que um paciente pode realizar zero ou mais exames e que um exame pode ser realizado por zero ou mais pacientes.

Para os pacientes, devem ser armazenados os seguintes dados: cod-paciente, CPF, nome, data de nascimento, endereço, etnia, religião, sexo e gênero. É importante notar que todos esses atributos são classificados como dados pessoais. Além disso, assuma que o "Controlador de Dados" da clínica definiu que os atributos cod-paciente e CPF devem ser criptografados, e que o nome e a data de nascimento devem ser anonimizados. De acordo com o "Controlador", os atributos etnia, religião, sexo e gênero são considerados dados pessoais sensíveis e, portanto, requerem atenção especial; no entanto, nenhuma técnica específica de processamento de dados foi definida para eles. Adicionalmente, o atributo endereço é considerado tanto um dado pessoal quanto um semi-identificador, embora nenhum processamento especial tenha sido especificado para este caso também. Para os exames médicos, os atributos a serem armazenados incluem: cod-exame, descrição e custo. Nenhum desses atributos é classificado como dado pessoal. Por fim, o relacionamento entre paciente e exame inclui dois atributos: data-exame e resultado. O "Controlador de Dados" especificou que o atributo resultado, que se qualifica como dado pessoal, deve ser criptografado, e que o atributo data-exame é um semi-identificador para o qual nenhum tratamento específico foi definido.

A Figura 2.3 ilustra o esquema conceitual produzido durante a fase de projeto conceitual, utilizando o modelo ER-PD, conforme modelado por meio da ferramenta br-ModeloPD. Observe que o conjunto entidade Paciente é representado por um retângulo com linhas tracejadas, indicando que ele corresponde a um Titular de Dados Pessoais. Note também que, ao lado do nome de cada atributo, o modelo indica — entre colchetes — o tipo de dado e o tipo de tratamento exigido, quando aplicável. Por exemplo, o atributo cod-paciente deve ser criptografado (o que é representado pelo símbolo [C]), enquanto o atributo nome deve ser anonimizado ([A]). Vale destacar que o método a ser utilizado para anonimizar o atributo nome não é especificado. Adicionalmente, os atributos data-nascimento e cor representam, respectivamente, um dado pessoal ([P]) e uma informação sensível ([S]). Todavia, nenhuma forma de tratamento foi especificada para esses atributos. Por fim, observe que o atributo valor do conjunto entidade Exame não representa dados pessoais.

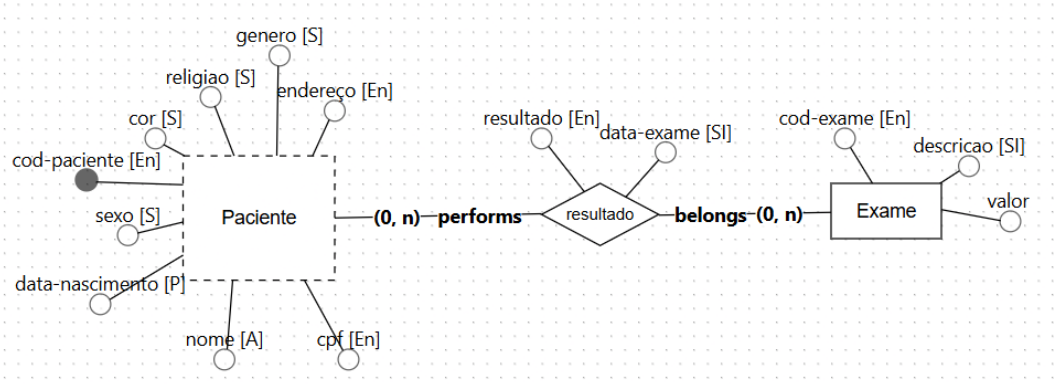


Figura 2.3: Esquema Conceitual no Modelo ER-PD.

Fonte: A autora.

2.6.2. Projeto Lógico

A próxima etapa do processo de projeto de banco de dados é denominada projeto lógico, cujo propósito é produzir um esquema lógico para o banco de dados, utilizando como entrada o esquema conceitual desenvolvido durante o projeto conceitual. O esquema lógico é expresso por meio de um modelo de dados, suportado por um sistema de gerenciamento de banco de dados, referido genericamente como modelo lógico. O modelo lógico mais comumente utilizado é o modelo relacional, e o esquema é tipicamente representado por um **diagrama relacional (DR)**.

A metodologia LGPDbyD propõe uma adaptação do modelo Relacional, denominada R-PD, para viabilizar o projeto lógico de bancos de dados em conformidade com a LGPD. O modelo R-PD permite representar os conceitos-chave definidos pela LGPD, tais como dados pessoais, tipos de operações de tratamento de dados e o titular dos dados. Adicionalmente, o modelo R-PD suporta a representação de atributos relativos a crianças e adolescentes.

A Figura 2.4 ilustra a notação adotada no modelo R-PD. Observe que uma relação (tabela) que representa um titular de dados pessoais é simbolizada por um retângulo com linhas pontilhadas. Além disso, ao lado do nome de cada atributo, entre colchetes, o modelo destaca tanto o tipo de dado quanto o método de tratamento a ser aplicado.

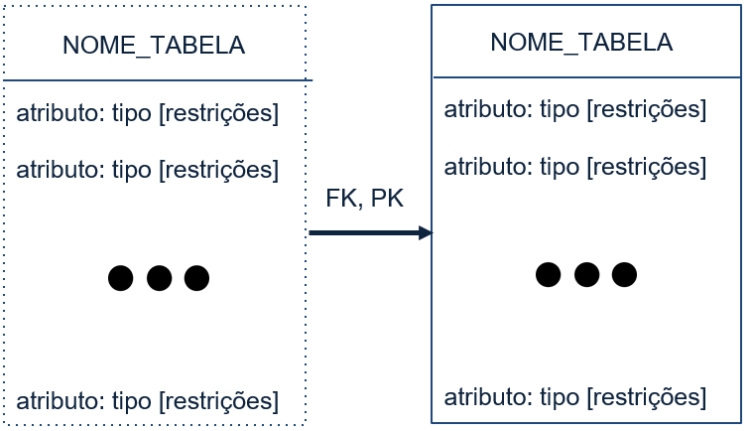


Figura 2.4: Notação do Modelo R-PD.

Fonte: A autora.

2.6.2.1. Exemplo de Execução

Consideremos o mesmo contexto descrito na seção anterior, envolvendo uma clínica de exames médicos que deseja armazenar informações sobre pacientes e exames. Inicialmente, o esquema conceitual é mapeado para o esquema lógico, visando a sua implementação técnica. Nesse processo, cada conjunto de entidades presente no esquema conceitual será desenhado para uma relação (tabela) no esquema lógico. Em seguida, os atributos são mapeados mediante a definição de seus tipos de dados e das restrições de domínio correspondentes (ex: *NOT NULL*, *UNIQUE*, *Primary Key - PK* e *Foreign Key - FK*). Posteriormente, aplicam-se as definições específicas da LGPD a cada um dos atributos

mapeados, conforme apresentando na Figura 2.5, que expõe o esquema lógico completo para o exemplo de execução previamente descrito, utilizando a ferramenta brModeloPD.

O segundo passo na tradução do esquema conceitual para o esquema lógico consiste no mapeamento dos "Conjuntos Relacionamentos". Sabe-se que todo "Conjunto Relacionamento" de cardinalidade N:M é representado no modelo relacional por uma nova relação. Neste sentido, uma nova relação denominada "Resultado" será criada para representar o "Conjunto Relacionamento" Resultado.

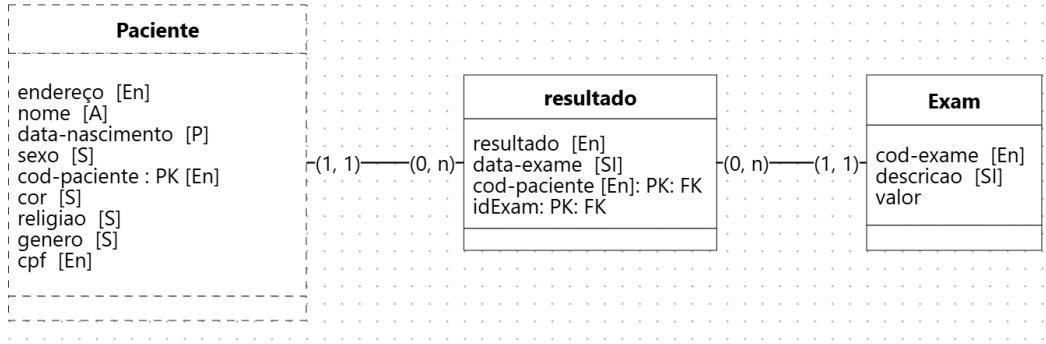


Figura 2.5: Esquema Lógico no Modelo ER-PD.

Fonte: A autora.

2.6.3. Projeto Físico

A próxima e última fase do projeto de banco de dados é denominada projeto físico. Esta fase recebe como entrada o esquema lógico, descrito no modelo R-PD, e produz como saída o esquema físico do banco de dados, o qual especifica as formas de organização de arquivos e as estruturas de armazenamento internas que serão utilizadas. O esquema físico será representado por meio de uma coleção de comandos DDL.

A metodologia LGPDbyD propõe uma extensão do comando SQL CREATE TABLE, denominada SQL-PD, com o objetivo de possibilitar o projeto físico de bancos de dados em conformidade com a LGPD. O SQL-PD permite representar os principais conceitos presentes na LGPD, a partir de metadados (comentários em um comando SQL). Esses metadados poderão ser utilizados para auditorias de conformidade com a LGPD.

A listagem 2.1 ilustra a gramática padrão da linguagem SQL para o comando CREATE TABLE. Observe que esta gramática permite definir o nome da tabela, os atributos, com seus respectivos tipos de dados, chaves primárias e/ou estrangeiras, além das restrições de integridade (ou consistência).

Listagem 2.1: Gramática Padrão do Comando SQL CREATE TABLE

```
CREATE TABLE nome_tabela
  (
  nome_colina tipo_de_dados [NOT NULL],
  [nome_coluna tipo_de_dados [NOT NULL]      ],
        [CONSTRAINT nome_restricao]
UNIQUE nome_coluna
          |PRIMARY KEY (nome_coluna {, nome_coluna})
 |FOREIGN KEY (nome_coluna {, nome_coluna})
                        REFERENCES nome_tabela
                            [ON DELETE CASCADE|
                                  SET NULL|NO ACTION],
                        [ON UPDATE CASCADE],
          |CHECK (predicado)
)
```

Baseada na listagem anterior, a Listagem 2.2 exibe a gramática estendida do comando CREATE TABLE (SQL-PD). Observe que esta gramática permite definir se a tabela corresponde ou não a um titular de dados pessoais, quais atributos são pessoais ou sensíveis, quais atributos devem ser anonimizados, criptografados, entre outros.

Listagem 2.2: Gramática Padrão do Comando SQL CREATE TABLE

```
CREATE TABLE nome_tabela
  (
  nome_colina tipo_de_dados [NOT NULL],
  [nome_coluna tipo_de_dados [NOT NULL]      ],
        [CONSTRAINT nome_restricao]
              UNIQUE nome_coluna
                    | PRIMARY KEY (nome_coluna {, nome_coluna})
                    | FOREIGN KEY (nome_coluna {, nome_coluna})
              REFERENCES nome_tabela
                     [ON DELETE CASCADE |
                                SET NULL | NO ACTION ],
                     [ON UPDATE CASCADE],
                    | CHECK (predicado)
                      | PESSOAL nome_coluna
                      | SENSIVEL nome_coluna
                      | ANONIMIZADO nome_coluna
                      | CRIPTOGRAFADO nome_coluna
                  | COMPARTILHADO nome_coluna
                      | IDENTIFICADOR nome_coluna
                      | SEMI IDENTIFICADOR nome_coluna
)
```

2.6.3.1. Exemplo de Execução

A seguir, iremos ilustrar a utilização da extensão SQL-PD no projeto físico de bancos de dados aderentes à LGPD. Para isso, vamos considerar o mesmo contexto descrito na seção anterior, envolvendo uma clínica de exames médicos que deseja armazenar informações de pacientes e exames.

A Listagem 2.3 ilustra o comando de criação da tabela “Paciente” na extensão SQL-PD. Na Listagem 2.4 tem-se o comando para a definição da tabela “Exame”. Já na Listagem 2.5 observa-se a criação da tabela “Resultado”, respectivamente.

Listagem 2.3: Comando CREATE TABLE para a Tabela Paciente (SQL-PD)

```
CREATE TABLE Paciente
        (cod_paciente integer NOT NULL,
        cpf varchar NOT NULL,
        nome varchar NULL,
        endereco varchar NULL,
        data_nascimento date NULL,
        sexo char NULL,
        cor char NULL,
        religiao char NULL,
        genero varchar NULL,
        CONSTRAINT c1 PRIMARY KEY (cod_paciente)
        /* , */
        /* CONSTRAINT c2 Criptografado cod_paciente, */
        /* CONSTRAINT c3 Criptografado cpf, */
        /* CONSTRAINT c4 Sensivel  sexo, */
        /* CONSTRAINT c5 Sensivel cor,*/
        /* CONSTRAINT c6 Sensivel religiao, */
        /* CONSTRAINT c7 Sensivel genero, */
        /* CONSTRAINT c8 Sensivel data_nascimento, */
        /* CONSTRAINT c9 Anonimizado nome, */
        /* CONSTRAINT c10 Anonimizado endereco */
    )
```

Listagem 2.4: Comando CREATE TABLE para a Tabela Exame (SQL-PD)

```
CREATE TABLE Exame
      (cod_exame integer NOT NULL,
        descricao varchar,
        valor numeric,
        CONSTRAINT c1 PRIMARY KEY cod-exame
        /* , */
        /* CONSTRAINT c2 Criptografado cod-exame */
    )
```

Listagem 2.5: Comando CREATE TABLE para a Tabela Resultado (SQL-PD)

```
CREATE TABLE Resultado
      (cod_paciente integer NOT NULL,
        cod_exame integer NOT NULL,
        data_exame date,
        resultado varchar,
        CONSTRAINT c1 PRIMARY KEY cod_paciente, cod_exame, data_exame
        CONSTRAINT c2 FOREING KEY cod_paciente REFERENCES Paciente (cod_paciente),
        CONSTRAINT c3 FOREING KEY data_exame REFERENCES
        Exame (data_exame)
        /* , */
        /* CONSTRAINT c4 Criptografado cod_paciente, */
        /* CONSTRAINT c5 Criptografado resultado */
    )
```

2.7. Modelando os Conceitos de Propósito e Consentimento

A LGPD estabelece que os dados pessoais são de titularidade dos respectivos sujeitos, ordenando a observância de critérios, restrições técnicas e organizacionais para as atividades de armazenamento, processamento, disponibilização e compartilhamento dessas informações. Tratando-se da proteção de dados, um dos princípios centrais consiste na concessão ao titular do controle sobre o ciclo de vida de seus dados, incluindo as condições sob as quais tais dados podem ser acessados, processados ou transferidos entre sistemas.

Esse controle é exercido por meio do **Consentimento**, mecanismo que define que o tratamento de dados pessoais somente pode ocorrer mediante autorização prévia, explícita e associada a uma finalidade específica (Art. 5º, XII). Ademais, a legislação assegura ao titular a possibilidade de estabelecer limites temporais para o uso de seus dados, o que implica a necessidade de mecanismos computacionais capazes de gerenciar prazos de validade, revogação e conformidade ao longo do tempo. Com o objetivo de representar esse conceito previsto na Lei, a metodologia LGPDbyD adota uma extensão da linguagem SQL, de modo a incorporar mecanismos capazes de expressar e gerenciar restrições associadas ao tratamento de dados pessoais.

2.7.1. Modelando o Conceito de Propósito

Para a representação do conceito de Propósito, é essencial a definição de um "identificador nominal" que o caracterize. Adicionalmente, a descrição de um propósito pode ser enriquecida por meio de um comentário, com o objetivo de ampliar sua expressividade semântica.

Além disso, assume-se a existência de relações hierárquicas entre propósitos, nas quais um determinado propósito p_2 pode ser definido como descendente ou herdar propriedades de um propósito de nível superior p_1, possibilitando a organização e o reaproveitamento de definições de finalidade em diferentes contextos de uso, como apresentado na Figura 2.6.

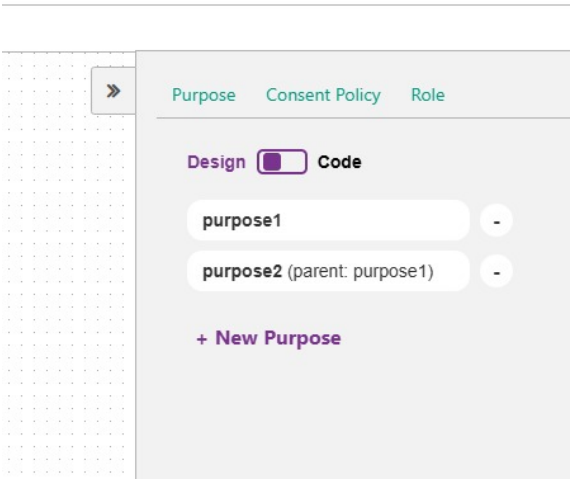


Figura 2.6: Hierarquia entre Propósitos p_1 e p_2.

Fonte: A autora.

Nesse contexto, quando um titular de dados concede consentimento para o uso de suas informações pessoais com a finalidade P_1, tais dados podem igualmente ser acessados para a finalidade P_2. Entretanto, **a relação inversa não é válida**; isto é, o consentimento concedido para a finalidade P_2 não implica, automaticamente, a autorização para o acesso aos dados com a finalidade P_1. Posto isso, com o objetivo de viabilizar a definição de propósitos, a LGPDbyD propõe a introdução de um novo comando da linguagem SQL, denominado **CREATE PURPOSE**, cuja sintaxe é apresentada na Listagem 2.6.

Listagem 2.6: Sintaxe do Comando CREATE PURPOSE

```
CREATE PURPOSE NOME [PARENT [=] NOME]
  [COMMENT [=] STRING];
```

2.7.2. Consentimento Padrão das Relações

A LGPDbyD propõe a utilização de um **Consentimento Padrão** para cada relação, o qual pode ser de dois tipos distintos: “Proibitivo” ou “Permissivo”. Caso o consentimento padrão de uma determinada relação *r* seja do tipo “proibitivo”, todo e qualquer acesso aos dados dessa relação **está por padrão proibido** e estes só poderão ser acessados caso exista um consentimento permitindo explicitamente o seu uso. Por outro lado, caso o consentimento padrão de uma determinada relação *r* seja do tipo “permissivo”, os seus dados podem, **por padrão, ser acessados livremente** e somente serão proibidos caso exista um consentimento desautorizando a sua utilização. O consentimento padrão é implementado na LGPDByD por dois novos tipos de restrição: **Consentimento Proibitivo** e **Consentimento Permissivo**.

Seguindo o exemplo de execução que estamos utilizando, considere que desejamos especificar que o consentimento padrão para a tabela *Paciente* seja do tipo *proibitivo*, enquanto o consentimento padrão da tabela *Exame* seja do tipo *permissivo*. As listagens 2.7 e 2.8 a seguir ilustram como adicionar as restrições para especificar o consentimento padrão das relações Paciente e Exame. Desta forma, os dados da tabela *Paciente* somente poderão ser acessados se houver algum consentimento permitindo explicitamente sua uti-

lização. Já os dados da tabela *Exame* poderão ser acessados livremente e somente terão seu acesso restringido caso exista um **Consentimento** explicitamente desautorizando a sua utilização.

Listagem 2.7: Consentimento Padrão da Relação Paciente

```
ALTER TABLE Patient
      ADD CONSTRAINT CD1 PROHIBITIVE CONSENT
```

Listagem 2.8: Consentimento Padrão da Relação Exame

```
ALTER TABLE Result
      ADD CONSTRAINT CD2 PERMISSIVE CONSENT
```

2.7.3. Modelando o Conceito de Consentimento

Antes de detalhar a forma como a LGPDbyD modela o conceito de Consentimento, apresenta-se um exemplo de uma instância da relação *Paciente*, conforme ilustrado na Tabela 2.1 correspondente.

Tabela 2.1: Instância ilustrativa da relação Resultado.

cod-paciente	cod-exame	data-exame	resultado
1	10	2025/01/01	Resultado 1
1	20	2025/02/02	Resultado 2
2	10	2025/03/03	Resultado 3
2	30	2025/04/04	Resultado 4
3	40	2025/05/05	Resultado 5
3	30	2025/06/06	Resultado 6
4	50	2025/07/07	Resultado 7
4	60	2025/08/08	Resultado 8

É inegável que cada paciente detém o direito de estabelecer as condições sob as quais seus dados podem ser acessados ou permanecem restritos. Com o objetivo de viabilizar a definição de propósitos específicos de uso, a LGPDbyD introduz um novo comando SQL, denominado **CREATE CONSENT**, cuja sintaxe detalhada é apresentada na Listagem 2.9.

Listagem 2.9: Sintaxe do Comando CREATE CONSENT

```
CREATE CONSENT NOME
 AS (ALLOW/DENY) PURPOSE SET
 TO TABLE_LIST [(COLUM_LIST)] [WHERE P(X)];
```

Para exemplificar a criação e o uso da abordagem proposta para a gestão de Consentimento, considere que Amy consente com o uso de todos os seus dados pessoais para os propósitos P1, P2 e P3. Nesse caso, é necessário criar um novo registro de Consentimento, onde o comando para a criação do mesmo é mostrado na Listagem 2.10.

Listagem 2.10: Criando Consentimento C1

```
CREATE CONSENT C1
 AS ALLOW PURPOSE P1, P2, P3
 TO Paciente
 WHERE cod_paciente = 1;
```

Além disso, considere que Anna consente com o uso de todos os seus dados para os propósitos P1 e P2, **mas não para o propósito P3**. Nesse caso, é necessário criar um novo registro de consentimento. O comando para a criação desse consentimento é mostrado na Listagem 2.11.

Listagem 2.11: Criando Consentimento C2

```
CREATE CONSENT C1
 AS ALLOW PURPOSE P1, P2
 TO Paciente
 WHERE cod_paciente = 2;
```

Ademais, suponha que Eva consente com o uso **apenas** dos atributos nome, data de nascimento, sexo e cor, **exclusivamente oara o propósito P2**. Nesse caso, também é necessário criar um novo registro de consentimento, o qual é mostrado na Listagem 2.12.

Listagem 2.12: Criando Consentiumento C3

```
CREATE CONSENT C3
 AS ALLOW PURPOSE P2
 TO Paciente (nome, data-nascimento, sexo, cor)
 WHERE cod_paciente = 3;
```

Existe a situação em que Ben consente com o uso dos atributos nome, data de nascimento, sexo e cor para o propósito P2. Ele também consente com o uso dos atributos nome, data de nascimento e sexo para o propósito P1. Nesse caso, **são necessários dois registros de consentimento *distintos* para representar com precisão as preferências de consentimento de Ben**. Os comandos para a criação desses consentimentos são mostrados na Listagem 2.13 e Listagem 2.14.

Listagem 2.13: Criando Consentimento C4

```
CREATE CONSENT C4
AS ALLOW PURPOSE P2
TO Patient (name, birthday, sex, color)
WHERE cod-patient = 4;
```

Listagem 2.14: Criando Cosentimento C5

```
CREATE CONSENT C5
AS ALLOW PURPOSE P1
TO Patient (name, birthday, sex)
WHERE cod-patient = 4;
```

Para finalizar, considere que Eric não forneceu nenhum consentimento explícito. Nesse caso, não é necessário criar nenhum registro de consentimento. A tabela 2.2 apresenta os metadados dos consentimentos em forma de tabela.

2.7.4. O Conceito de Role

A metodologia apresentada resulta na geração de um grande volume de registros de Consentimento. Para otimizar a gestão e o aproveitamento desses consentimentos, propõe-se o conceito de **Role (Regra)**, que consiste essencialmente em um agrupamento de Consentimentos. Um usuário que recebe autorização para utilizar uma determinada **Role** *R* passa

Tabela 2.2: Representação dos Metadados de Consentimentos.

Nome	Propósito Definido	Tabela	Colunas	Tipo	Condição
C1	P1, P2, P3	Paciente	All	Allow	WHERE cod_paciente = 1
C2	P1, P2	Paciente	All	Allow	WHERE cod_paciente = 2
C3	P2	Paciente	nome, data_nascimento, sexo, cor	Allow	WHERE cod_paciente = 3
C4	P2	Paciente	nome, data_nascimento, sexo, cor	Allow	WHERE cod_paciente = 4
C5	P1	Paciente	nome, data_nascimento, sexo	Allow	WHERE cod_paciente = 4

a ter automaticamente acesso a todos os registros de consentimento a ela associados. Assim, para possibilitar a criação de Roles, a LGPDbyD recomenda a utilização do comando **GRANT CONSENT**. Por exemplo, caso seja necessário criar uma Role denominada **R1**, esta pode agrupar os registros de Consentimento previamente definidos, como C1, C2, C3, C4 e C5, permitindo uma gestão mais eficiente e estruturada dos acessos, como ilustrado em na regra **R1**, na Listagem 2.15 e os metadados dos Consentimentos em forma tabular são visualizados na Tabela 2.3

Listagem 2.15: Criando Regra R1

```
GRANT CONSENT C1, C2, C3, C4, C5 TO ROLE R1
```

Tabela 2.3: Representação dos Metadados das Regras.

Nome	Conjunto de Consentimentos
R1	C1, C2, C3, C4, C5

Com as "Regras" criadas, o próximo passo é associar os usuários do sistema de banco de dados às Regras previamente definidas. Para isso, a LGPDbyD utiliza o comando **GRANT USER**, como exemplificado na Listagem 2.16, que mostra o comando para associar o usuário **USER1** à "Regra" **R1**. A Tabela 2.4 apresenta os metadados usados para associar usuários às respectivas regras.

Listagem 2.16: Associando o usuario **USER1** com a Regra **R1**

```
GRANT USER USER1 TO R1
```

2.7.5. Execução e Reescrita de Consultas na Metodologia LGPDbyD

Na metodologia LGPDbyD, ao solicitar a execução de uma consulta SQL, o usuário do banco de dados deve especificar o propósito para o qual o resultado da consulta será utilizado. Dessa forma, toda solicitação de execução de consulta SQL estará sempre associada a um ou mais propósitos.

Tabela 2.4: Metadados para Associar Usuários a Regras.

Usuario	Regras
User1	R1

Para ilustrar o processo de execução de consultas SQL segundo a metodologia LGPDbyD, suponha que o usuário USER1 tenha solicitado a execução da consulta Q1 para o Propósito P1. A consulta *Q1* está ilustrada na Listagem 2.17.

Listagem 2.17: Consulta SQL Q1

```
SELECT *
FROM Patient
```

Observe que a consulta Q1 deve ser **reescrita** pelo processador de consultas para retornar apenas os dados que foram autorizados pelos pacientes para o propósito P1. A implementação do mecanismo de reescrita da consulta está além do escopo deste trabalho. Contudo, para fins ilustrativos, a consulta reescrita é mostrada na Listagem 2.18 e o resultado retornado é apresentado na Tabela 2.5.

Listagem 2.18: Reescrita da Consulta Q1 para o Proposito P1

```
SELECT *
FROM Patient
WHERE cod-patient = 1 OR cod-patient = 2
UNION
SELECT NULL, NULL, name, NULL, birthday, sex, NUL, NULL, NULL
FROM Patient
%WHERE cod-patient = 4
```

Tabela 2.5: Resultado da Reescrita da Consulta Q1 para o Propósito P1.

cod-paciente	cpf	nome	endereço	data-nascimento	sexo	cor	religião	gênero
1	123.456.789-00	Amy	123 St	30/10/1980	F	White	Christian	Female
2	234.567.890-01	Anna	234 St	30/09/1981	F	Black	Muslim	Queer
NULL	NULL	Ben	NULL	30/07/1983	M	NULL	NULL	NULL

É importante lembrar que a política padrão de Consentimento para a relação "Resultado" é do tipo permissivo. Assim, seus dados podem, por padrão, ser acessados livremente e só serão restritos caso exista um consentimento explícito negando seu uso.

A título de exemplificação, consideremos que Amy fornece consentimento para o uso dos resultados de seus exames médicos para os propósitos *P1*, *P2* e *P3*. Nesse caso, dado que o consentimento padrão para a relação *Resultado* é do tipo **permissivo**, não é necessário um registro explícito de consentimento. Por outro lado, Anna autoriza o uso dos resultados de seus exames para os propósitos *P1* e *P2*, mas retém explicitamente o consentimento para o propósito *P3*. Portanto, deve ser criado um registro de consentimento do tipo negar, conforme mostrado na Listagem 2.19.

Listagem 2.19: Criando Consentimento C6

```
CREATE CONSENT C6
AS DENY PURPOSE P3
TO Result
WHERE cod_patient = 2;
```

Antes de mostrar exemplos adicionais de execução de consultas, considere que a **Role R1** foi modificada para incluir os propósitos C6, C7, C8 e C9. Para isso, executamos o comando mostrado na Listagem 2.20. É importante notar que o **USER1** já havia sido associado à **Role R1**, portanto, nenhuma associação adicional é necessária. Além disso, consideremos um exemplo de uma instância da relação *Resultado*, conforme ilustrado na Tabela.

Listagem 2.20: Atualizando Regra R1

```
GRANT CONSENT C1, C2, C3, C4, C5, C6, C7, C9 TO ROLE R1
```

Tabela 2.6: Instância ilustrativa da relação Resultado.

cod-paciente	cod-exame	data-exame	resultado
1	10	2025/01/01	Resultado 1
1	20	2025/02/02	Resultado 2
2	10	2025/03/03	Resultado 3
2	30	2025/04/04	Resultado 4
3	40	2025/05/05	Resultado 5
3	30	2025/06/06	Resultado 6
4	50	2025/07/07	Resultado 7
4	60	2025/08/08	Resultado 8

Para ilustrar o processo de execução de consultas SQL segundo a metodologia LGPDbyD, suponha que o usuário **USER1** tenha solicitado a execução da consulta *Q2* para o Propósito **P1**. A consulta *Q2* está ilustrada na Listagem 2.21.

Listagem 2.21: Consulta SQL Q2

```
SELECT *
FROM Result
```

Observe que a consulta Q2 deve ser **reescrita** pelo processador de consultas para retornar apenas os dados que foram autorizados pelos pacientes para o propósito P1. A consulta reescrita é mostrada na Listagem 2.22 e o resultado retornado é apresentado na Tabela 2.7.

Listagem 2.22: Reescrita da Consulta Q2 para o Proposito P1

```
SELECT *
FROM Result
EXCEPT
SELECT *
FROM result
WHERE cod_paciente = 4 OR cod_paciente = 3
UNION
SELECT cod_paciente, cod_exame, data_exame, NULL
FROM result
WHERE cod_paciente = 3
```

Tabela 2.7: Saída da Reescrita da Consulta *Q2* para o Propósito **P1**.

cod-paciente	cod-exame	data-exame	resultado
1	10	2025/01/01	Resultado 1
1	20	2025/02/02	Resultado 2
2	10	2025/03/03	Resultado 3
2	30	2025/04/04	Resultado 4
3	40	2025/05/05	NULL
3	30	2025/06/06	NULL

2.7.6. Modelando a Duração do Armazenamento de Dados Pessoais

A LGPD não estabelece um prazo fixo e universal para o armazenamento de dados pessoais. Em vez disso, define princípios orientadores e condições para a retenção de dados, particularmente em relação ao propósito, necessidade e adequação. Contudo, de acordo com a Lei, os dados pessoais devem ser mantidos apenas pelo período necessário para cumprir a finalidade para a qual foram coletados. A base legal para essa exigência está prevista no Artigo 15 da LGPD, que afirma: "O término do tratamento de dados pessoais ocorrerá nas seguintes hipóteses: I – verificação de que a finalidade foi alcançada ou de que os dados deixaram de ser necessários ou pertinentes ao alcance da finalidade específica; II – fim do período de tratamento; III – comunicação do titular, inclusive no exercício de seu direito de revogação do consentimento; IV – determinação da autoridade nacional, quando houver violação à LGPD."

Nesse contexto, torna-se necessário modelar o período pelo qual os dados pessoais devem, em princípio, ser armazenados. Para viabilizar isso, a metodologia LGPDbyD introduz um novo tipo de restrição, denominado **DURATION**, que define uma lista de possíveis durações de armazenamento para uma determinada relação. A sintaxe da restrição **DURATION** é mostrada na Listagem 2.23.

Listagem 2.23: Sintaxe da Restrição Duração

```
ALTER TABLE TABLE_NAME
ADD CONSTRAINT NAME_CONSTRAINT DURATION (d1<desc1>, d2<desc2>,...);
```

Hipoteticamente, considere que os dados dos pacientes podem ser armazenados por 30 dias ou por 60 dias. Assim, existem duas possíveis durações de armazenamento para as tuplas na relação Paciente, genericamente denominadas ***t1*** e ***t2***. A Listagem 2.24 mostra o comando **ALTER TABLE** usado para adicionar a restrição **DURATION**.

Listagem 2.24: Adicionando a Restrição de Duração na Relação Paciente

```
ALTER TABLE Patient
ADD CONSTRAINT D1
DURATION (t1 = '30 days', t2 = '60 days');
```

Ao inserir uma nova tupla na relação Paciente, torna-se necessário também especificar qual das alternativas de duração de armazenamento pré-definidas será associada a essa tupla em particular. Considere, por exemplo, a inserção da paciente Amy na tabela Paciente. Nesse caso, o comando **INSERT** deve indicar qual das opções de duração de armazenamento previamente definidas para a Tabela Paciente será atribuída à tupla que armazena os dados de Amy.

Suponha que Amy tenha autorizado o armazenamento de seus dados por apenas 30 dias. Nesse caso, a Listagem 2.25 mostra o comando **INSERT** para adicionar os dados de Amy. Outrossim, suponha que os demais pacientes tenham optado por autorizar o armazenamento de seus dados por um período de 60 dias. A Tabela 2.8 ilustra uma instância da relação Paciente com os metadados necessários para gerenciar a retenção dos dados.

Listagem 2.25: Inserção de dados com duração associada na relação Paciente

```
INSERT INTO Patient (cod_paciente, cpf, nome, endereco, data_nascimento,
                     sexo, cor, religiao, genero)
VALUES (1, '123', 'Amy', '123 St', '1980-10-30',
        'F', 'White', 'Christian', 'Female')
WITH DURATION t1;
```

Tabela 2.8: Instância Ilustrativa da Relação *Paciente* com Metadados de Duração.

cod-patient	cpf	name	address	birthday	sex	...	insert-date	duration
1	123.456.789-00	Amy	123 St	1980-10-30	F	...	2013-05-25	t1
2	234.567.890-01	Anna	234 St	1981-09-30	F	...	2014-05-25	t2
3	456.789.012-34	Eva	456 St	1982-08-30	F	...	2015-05-25	t2
4	678.901.234-56	Ben	678 St	1983-07-30	M	...	2016-05-25	t2
5	891.234.567-89	Eric	891 St	1984-06-30	M	...	2017-05-25	t2

Com o objetivo de preservar a compatibilidade com sistemas legados que já utilizam o comando 'INSERT' sem a cláusula 'WITH DURATION', torna-se viável a implementação de um mecanismo automático de tratamento dessas inserções, como, por exemplo, o uso de *triggers*. Nesse contexto, sempre que uma tupla for inserida sem a especificação explícita da duração, o sistema pode adotar automaticamente a política mais restritiva, correspondente ao menor período de retenção de dados. Como alternativa, podem ser consideradas políticas menos restritivas ou ainda uma opção padrão previamente definida, conforme as diretrizes estabelecidas.

A definição da política a ser aplicada nesses casos pode ser configurada pelo DPO, de modo a alinhar o comportamento do sistema às exigências legais e organizacionais. Em qualquer cenário, a imposição de uma restrição de duração implica a necessidade de armazenamento da data de inserção das tuplas, de forma a viabilizar o controle do tempo de retenção. Ademais, faz-se necessária a definição de uma estratégia para lidar com situações em que a duração associada a uma determinada tupla é atingida ou expirada. Contudo, o detalhamento do projeto e da implementação de tal estratégia extrapola o escopo deste Capítulo.

2.8. Ferramentas de Suporte

A utilização de ferramentas CASE (*Computer-Aided Software Engineering*) no processo de projeto de bancos de dados representa um avanço significativo em termos de produtividade, padronização e qualidade dos artefatos produzidos. Ferramentas como BrModelo, ERWin, MySQL Workbench e Oracle SQL Developer Data Modeler oferecem ambientes integrados que suportam as diferentes fases do projeto de bancos de dados, permitindo que o projetista construa diagramas de forma visual e intuitiva, reduzindo a ocorrência de erros estruturais e mantendo a rastreabilidade entre os diferentes níveis de abstração

do projeto. Além disso, essas ferramentas automatizam tarefas repetitivas e propensas a falhas, como a geração de scripts DDL (*Data Definition Language*), a verificação de restrições de integridade e a engenharia reversa de esquemas já existentes, contribuindo para a redução do tempo de desenvolvimento e para a consistência entre o modelo documentado e a implementação efetiva do banco de dados. O uso de ferramentas CASE também favorece a comunicação entre os membros da equipe de desenvolvimento, uma vez que os modelos produzidos seguem notações padronizadas, o que facilita a leitura, a revisão e a manutenção dos esquemas ao longo do ciclo de vida do banco de dados. Assim, a adoção dessas ferramentas não apenas agiliza o processo de projeto, mas também eleva o rigor metodológico e a documentação técnica, aspectos fundamentais para a qualidade e a longevidade dos sistemas de bancos de dados.

2.8.1. A Ferramenta BrModeloPD

A escolha da brModelo como base para implementação da BrModeloPD foi motivada por esta ser uma ferramenta de código aberto. Essa particularidade permitiu que o desenvolvimento da BrModeloPD se beneficiasse da estrutura existente, evitando a necessidade de desenvolver uma solução completa do zero. O foco principal, portanto, foi direcionado à integração das funcionalidades da LGPD sobre a base já estabelecida do brModelo.

A decisão também baseou-se em sua arquitetura extensível, que permite a implementação de novos módulos de forma desacoplada do núcleo *core* do sistema. Esse enfoque, análoga ao funcionamento de um *plugin*, confirma a separação de interesses entre as camadas da plataforma original e a camada desenvolvida nesta ferramenta. Tal autonomia estrutural é estratégica, pois permite evoluções e modificações personalizadas sem a necessidade de alteração no código-fonte do brModelo. Além disso, essa estrutura favorece a acessibilidade e a portabilidade, uma vez que a funcionalidade desenvolvida coexiste com o sistema base sem estabelecer uma dependência rígida ou um acoplamento forte.

De posse dessas informações, primeiramente estabeleceu-se em compreender como o brModelo armazena e trata os dados dentro do banco de dados em geral e, a partir desse entendimento passou-se a integrar a LGPD no banco. Verificou-se que existe uma modelagem para o banco conceitual e uma modelagem em paralelo para o lógico, como também há, no próprio código do brModelo, um sistema de conversão para converter a modelagem conceitual em lógica, além de existir um serviço que transforma uma modelagem lógica em um texto equivalente a uma saída SQL.

Com isso, a Lei foi implementada como um vetor de atributos completamente nulo, nos quais os elementos referentes a cada propriedade da LGPD podem ser **verdadeiros** ou **falsos**. Essa abordagem permitiu que cada atributo de cada classe na modelagem de banco de dados fossem preenchidos manualmente pelo usuário, garantindo a granularidade necessária para a conformidade com a LGPD. Além da implementação no *backend*, foi desenvolvida uma interação visual que permita ao usuário realizar essas mudanças nos atributos, alterando a forma como o elemento era posicionado e seu título na tela, refletindo as categorias da Legislação.

Como etapa final dessa integração, houve a conversão das informações da LGPD para o código SQL, optou-se por inserir que as informações da LGPD sejam declaradas como comentários nas tabelas e colunas do SQL gerado. Essa metodologia garante que as normas referentes à Lei estejam facilmente visualizáveis e acessíveis, sem comprometer a funcionalidade da *query* SQL, incorporando, dessa forma, a Lei como informação contextual, mantendo a funcionalidade do banco de dados.

Um ponto significativo de divergência e inovação entre a integração da LGPD no SQL foi a **introdução de uma nova Lógica de Consentimento atribuídas aos Propósitos**. Esta lógica, que se baseia no uso de Propósitos e Regras de Consentimento, foi implementada do zero, uma vez que era completamente nova e não se alinhava com a estrutura existente da aplicação. Embora utilizada na interface do brModelo, essa lógica de variáveis não estava diretamente ligada aos atributos de uma classe ou à modelagem do banco de dados. Ao contrário, era intrinsecamente conectada à execução das *queries* do banco, conforme a Política de Consentimento. A escolha de como lidar com esses Propósitos e as Regras de Consentimento foi um desafio, mas optou-se por integrá-las ao modelo lógico do banco de dados, onde as definições já estavam mais estabelecidas, e ao serviço SQL, onde o texto de Consentimento passou a ser incluído como parte do resultado da *query* SQL.

2.9. Desafios e Pesquisa Futura

No transcorrer desse Capítulo, demonstra-se que a inclusão de requisitos jurídicos na modelagem de bancos de dados não deve ser vista como uma atividade posterior ao desenvolvimento, mas sim como um componente inerente e fundamental de todo o ciclo de criação de sistemas de informação. A proposição da metodologia **LGPDbyD** e a implementação da ferramenta **BrModeloPD** revelou-se eficaz ao incorporar os preceitos e obrigações da LGPD logo nas etapas preliminares de um projeto.

Isso abre a possibilidade, como trabalhos futuros, para o desenvolvimento de ferramentas capazes de realizar verificações automatizadas diretamente nas tabelas do banco de dados, a fim de verificar se os dados armazenados estão, de fato, consistentes com o que foi definido durante o projeto do banco de dados. Além disso, tais ferramentas também poderiam ser desenvolvidas para apoiar atividades de auditoria de conformidade com a LGPD.

2.9.1. Ferramentas de Suporte à Conformidade com a LGPD

Com base nos metadados gerados durante a aplicação da estratégia LGPDbyD, é possível identificar quais atributos são pessoais ou não pessoais, quais são classificados como sensíveis e quais devem ser anonimizados ou criptografados, entre outras classificações. Isto abre a possibilidade para o desenvolvimento de ferramentas capazes de realizar verificações automatizadas diretamente nas tabelas de banco de dados, a fim de verificar se os dados armazenados são, de fato, consistentes com o que foi definido durante o projeto do banco de dados. Além disso, tais ferramentas poderiam também ser desenvolvidas para apoiar atividades de auditoria de conformidade com a LGPD.

2.9.1.1. Ferramenta Automatizada de Verificação e Auditoria

O propósito desta ferramenta é aproveitar os metadados adicionados durante o projeto do banco de dados para verificar, por exemplo, se os atributos definidos como anonimizados ou criptografados estão, de fato, armazenados no banco de dados com o tratamento de dados apropriado aplicado conforme previamente especificado. Após realizar essas verificações, a ferramenta poderia gerar um relatório de conformidade. Com este relatório em mãos, o Controlador de Dados ou o Encarregado de Proteção de Dados (DPO) seria capaz de solicitar as mudanças necessárias para garantir que a organização esteja totalmente em conformidade com a LGPD. A Figura 2.7 ilustra uma possível arquitetura para uma ferramenta de auditoria automatizada.

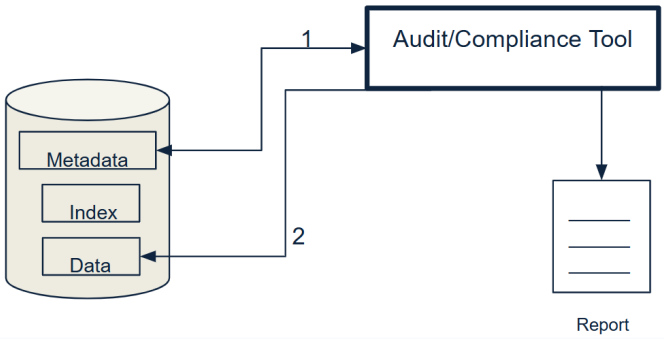


Figura 2.7: Uma Arquitetura Proposta para um Sistema de Auditoria Automatizado.

Fonte: A autora.

2.9.1.2. Criação de Pacotes/APIs para Desenvolvedores

O propósito desta ferramenta é fornecer pacotes (em inglês, packages), ou APIs, independentes de SGBDs, contendo a implementação de diferentes métodos de anonimização e criptografia. Com estes pacotes, o programador/desenvolvedor pode criar Triggers e Procedures para implementar, no próprio sistema de banco de dados, o tratamento apropriado, definido no projeto do banco de dados, para os atributos que representam dados pessoais. Por exemplo, se um determinado atributo precisa ser armazenado de forma criptografada, o desenvolvedor seleciona, do pacote fornecido, a função mais apropriada e cria Triggers que serão executados durante a inserção ou atualização deste atributo. Desta forma, a ferramenta proposta pode contribuir para reduzir o tempo de desenvolvimento e a conformidade com a LGPD. A Figura 2.8 apresenta um possível modelo arquitetônico para uma API destinada a auxiliar desenvolvedores ou profissionais de banco de dados.

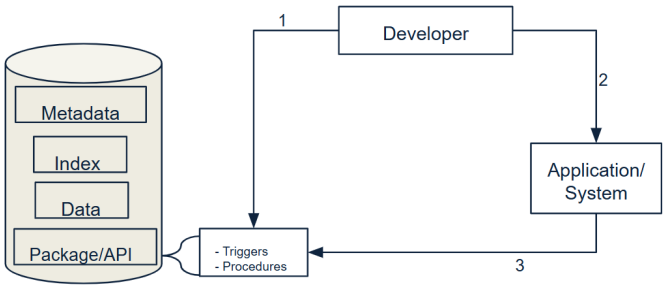


Figura 2.8: Uma Possível Arquitetura para uma *API* de Suporte ao Desenvolvedor ou Profissional de Banco de Dados.

Fonte: A autora.

2.10. Conclusões

Este minicurso apresentou os principais conceitos relacionados à Lei Geral de Proteção de Dados Pessoais (LGPD) e discutiu os desafios de incorporar seus requisitos ao projeto de bancos de dados. Inicialmente, foram abordados aspectos fundamentais da legislação, incluindo seus princípios, atores, tipos de dados e implicações para o tratamento de informações pessoais. Em seguida, foram revisados os fundamentos do projeto de bancos de dados, evidenciando a lacuna existente entre as metodologias tradicionais de modelagem e as exigências legais impostas pela LGPD. A partir dessa análise, foi apresentada a metodologia **LGPDbyD**, cuja proposta consiste em integrar requisitos de privacidade e proteção de dados desde as fases iniciais do projeto, abrangendo os níveis conceitual, lógico e físico.

Além da metodologia proposta, foi apresentada a ferramenta **BrModeloPD**, desenvolvida para apoiar a aplicação prática dos conceitos da LGPD durante a modelagem de bancos de dados. Os resultados que foram obtidos demonstram que a incorporação dos requisitos legais diretamente nos artefatos de projeto favorece a conformidade regulatória, amplia a transparência no tratamento de dados e contribui para o desenvolvimento de sistemas mais seguros e alinhados aos princípios de *Privacy by Design*. Espera-se que as estratégias discutidas neste minicurso auxiliem pesquisadores, estudantes e profissionais no desenvolvimento de soluções capazes de atender simultaneamente às necessidades técnicas dos sistemas de informação e às exigências legais de proteção de dados.

Referências

[Amaral 2016] Amaral, F. (2016). *Introdução à ciência de dados: mineração de dados e big data*. Alta Books Editora.

[Araújo 2008] Araújo, M. (2008). Modelagem de dados–teoria e prática. *Revista Saber Digital*, 1(01):27–64.

[Barros et al. 2024] Barros, P. V. d. S., França, J. C. M., da SM Filho, J. M., and Machado, J. C. (2024). brmodelopd: Uma extensão da ferramenta brmodelo para incorporar os requisitos e as restrições da lgpd ao projeto de banco de dados. In *Simpósio Brasileiro de Banco de Dados (SBBD)*, pages 101–106 SBC.

[Brito and Machado 2017] Brito, F. T. and Machado, J. C. (2017). Preservação de privacidade de dados: Fundamentos, técnicas e aplicações. *Jornadas de atualização em Informática*, pages 91–130.

[Carvalho et al. 2023] Carvalho, G., Bernardino, J., Pereira, V., and Cabral, B. (2023). Er+: A conceptual model for distributed multilayer systems. *IEEE Access*.

[Cavoukian et al. 2009] Cavoukian, A. et al. (2009). Privacy by design: The 7 foundational principles. *Information and privacy commissioner of Ontario, Canada*, 5(2009):12.

[Chen 1976] Chen, P. P.-S. (1976). The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems*, 1(1):9–36.

[Codd 1970] Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387.

[Dani and Getta 2005] Dani, A. and Getta, J. (2005). Conceptual modelling of computations on data streams.

[de Abreu et al. 2021] de Abreu, C., Praciano, F. D., Amora, P. R., and Machado, J. C. (2021). Consql: Consentimentos em sql para o processamento de consultas orientado a propósitos. In *Anais Estendidos do XXXVI Simpósio Brasileiro de Bancos de Dados*, pages 8–14 SBC.

[de Oliveira 2024] de Oliveira, J. P. M. (2024). *Modelagem Conceitual e Ontologia: Uma introdução rigorosa sobre a Modelagem Conceitual dos primórdios da ciência até a ontologia computacional*. Independently Published.

[Elmasri and Navathe 2022] Elmasri, R. and Navathe, S. B. (2022). *Fundamentals of Database Systems*. Pearson, 7th edition.

[Fontes 2012] Fontes, E. (2012). *Políticas e Normas para a Segurança da Informação*. Brasport.

[Gruber 1993] Gruber, T. R. (1993). A translation approach to portable ontology specifications.

[Guarino 1998] Guarino, N. (1998). *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*, volume 46. IOS press.

[Heuser 2009] Heuser, C. A. (2009). *Projeto de Banco de Dados*. Bookman, Porto Alegre, 6 edition.

[Kamble 2008] Kamble, A. S. (2008). A conceptual model for multidimensional data. In *APCCM*, volume 8, pages 29–38.

[Khan et al. 2004] Khan, K. M., Kapurubandara, M., and Chadha, U. (2004). Incorporating business requirements and constraints in database conceptual models. In *Proceedings of the first Asian-Pacific conference on Conceptual modelling-Volume 31*, pages 59–64.

[Li et al. 2009] Li, Z., Yang, M. C., and Ramani, K. (2009). A methodology for engineering ontology acquisition and validation. *AI EDAM*, 23(1):37–51.

[Pinheiro 2020] Pinheiro, P. P. (2020). *Proteção de Dados Pessoais: Comentários à Lei n. 13.709/2018-LGPD*. Saraiva Educação SA.

[Pressman and Maxim 2020] Pressman, R. S. and Maxim, B. R. (2020). *Software Engineering: A Practitioner's Approach*. McGraw-Hill, 9th edition.

[SANTANA 2016] SANTANA, R. C. G. (2016). Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. *Informação & Informação; v. 21, n. 2 (2016); 116–142*, 24(2).

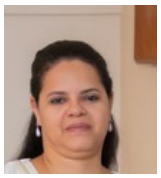
[Sarkar and Athanassoulis 2022] Sarkar, S. and Athanassoulis, M. (2022). Query language support for timely data deletion. In *Proceedings of the 25th International Conference on Extending Database Technology*, volume 2.

[Sommerville 2015] Sommerville, I. (2015). *Software Engineering*. Pearson, 10th edition.

[Vieira et al. 2026] Vieira, P., Monteiro, J. M., Machado, J., and Brayner, A. (2026). Incorporating lgpd requirements and restrictions into database design. *Journal of Information and Data Management*, 17(1):133–145.

Sobre os autores

Patricia Vieira da Silva Barros



Possui Mestrado em Ciência da Computação pela Universidade Federal do Pernambuco – UFPE (2015). Atualmente é Doutoranda em Ciência da Computação pela Universidade Federal do Ceará – UFC. Especialista em Tecnologias para Web, pela Universidade Federal do Piauí (2003) e Graduação em Direito(2006) e Computação (2002). Trabalha como Professora Assistente II no Curso de Bacharelado em Sistemas de Informação no Campus Senador Helvídio Nunes de Barros/Picos, campus da Universidade Federal do Piauí - UFPI. Tem interesse nas áreas de Banco de Dados atuando principalmente nos seguintes temas: Ontologias, Representação do Conhecimento, Lógica, IA e Direito e Computação. Participa de grupos de pesquisa SWORD e PAAD. http://lattes.cnpq.br/5516550975920922.

José Maria da Silva Monteiro Filho

Possui graduação em Bacharelado em Computação pela Universidade Federal do Ceará - UFC (1998), Mestrado em Ciência da Computação pela UFC (2001) e Doutorado em Informática pela Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio (2008). Atualmente é professor na UFC. Tem experiência na área de Ciência da Computação, com ênfase em Banco de Dados e Engenharia de Software, atuando principalmente nos seguintes temas: sintonia automática de bancos de dados, bancos de dados em nuvem, dados ligados na Web e qualidade de software. Detalhes em: http://lattes.cnpq.br/9790693300026949.

Javam de Castro Machado

É professor titular do Departamento de Computação da UFC, onde fundou e coordena, há mais de 15 anos, o Laboratório de Sistemas e Bancos de Dados (LSBD). Javam fez doutorado na Université Joseph Fourier - Grenoble I - França (1995) e foi diretor de tecnologia da informação e coordenador de inovação tecnológica da Pro-Reitoria de Pesquisa também da UFC. Foi igualmente coordenador da Comissão Especial de Banco de Dados da SBC (2017) e pesquisador visitante na Telecom Sud Paris (2001) e no AT&T Labs-Research (2018 e 2020). Membro da SBC, ACM e IEEE, o professor Javam se interessa cientificamente pelas áreas de privacidade de dados e de justiça algorítmica. Detalhes em: http://lattes.cnpq.br/9884980518986225.