

Incorporando os Requisitos e as Restrições da LGPD ao Projeto de Banco de Dados

Patricia Vieira da S. Barros¹, José Maria Monteiro¹,
Angelo Brayner¹, Javam C. Machado¹

Departamento de Computação (DC) – Universidade Federal do Ceará (UFC)
CEP 60440-900 – Fortaleza – CE – Brazil

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

Abstract. *The Brazilian General Data Protection Law (LGPD) specifies how the processing, storage, and disposal of personal data should be conducted, conditioning it to the prior authorization of the data subject. On the other hand, current information systems are heavily reliant on the use of personal data and therefore need to comply with the LGPD. In this context, the database system becomes an even more critical component in software development, as it is responsible for storing, updating, and retrieving data. However, the methodologies and tools used for database design do not incorporate the requirements and constraints of the LGPD, making it difficult to ensure compliance between databases and current legislation. This article presents a methodology, called LGPDbyD, to incorporate the impositions and principles of the LGPD into the design of databases. To achieve this, we extend the ER model, the Relational model, and the CREATE TABLE command. Additionally, we extend the brModelo tool to provide support for the requirements and constraints of the LGPD. LGPDbyD aims to facilitate the processes of database design and auditing in compliance with the LGPD.*

Resumo. *A Lei Geral de Proteção de Dados Pessoais (LGPD) tem por finalidade determinar como deve ser realizado o tratamento, o armazenamento e o descarte de dados pessoais, inclusive nos meios digitais, sempre com autorização prévia do concedente. Por outro lado, os sistemas de informação atuais estão fortemente baseados na utilização de dados pessoais e, por conseguinte, precisam estar em conformidade com a LGPD. Neste contexto, o sistema de bancos de dados 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. Contudo, as metodologias e ferramentas utilizadas para o projeto de bancos de dados não incorporam os requisitos e as restrições da LGPD, dificultando assim a conformidade entre os bancos de dados e a legislação vigente. Este artigo apresenta uma estratégia, denominada LGPDbyD, para incorporar as imposições e preceitos da LGPD ao projeto de bancos de dados. Para isso, adaptamos o modelo ER, o modelo Relacional e o comando CREATE TABLE. Adicionalmente, estendemos a ferramenta brModelo a fim de fornecer suporte aos requisitos e restrições da LGPD. A LGPDbyD busca facilitar os processos de projeto de bancos de dados e auditoria em conformidade com a LGPD.*

1. Introdução

A Lei Geral de Proteção de Dados – LGPD, Lei 13.709/18¹ 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 deve ser realizado o tratamento, o armazenamento e o descarte de dados pessoais, buscando proteger os direitos fundamentais de liberdade e de privacidade. Essa legislação impõe uma profunda transformação no sistema de proteção de dados no país. A LGPD é uma legislação que envolve uma mudança de processos, uma atualização 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 trazer questões não satisfatoriamente mencionadas por outras leis setoriais de proteção de dados existentes no Brasil, como, por exemplo, ela traz uma definição mais exata sobre o 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 ANPD (Autoridade Nacional de Proteção de Dados), a definição de sanções, proporcionando assim uma maior segurança jurídica aos detentores de dados pessoais.

Por outro lado, os Sistemas de Informação (SIs) atuais estão fortemente baseados na aquisição, armazenamento e processamento de dados pessoais. Logicamente, esses sistemas necessitam estar em conformidade com a LGPD. Desta forma, a LGPD 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 de 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 buscar assegurar a conformidade de um banco de dados em relação à LGPD consiste em tentar incorporar os requisitos e as restrições impostos pela legislação ao processo de projeto de bancos de dados. Contudo, o projeto de bancos de dados é uma atividade complexa, que envolve quatro fases distintas: i) levantamento e análise de requisitos, ii) projeto conceitual, iii) projeto lógico e iv) projeto físico. Além disso, essas etapas produzem diversos artefatos, utilizando notações distintas e, frequentemente, suportados por diferentes ferramentas de *software*. Além disso, as metodologias existentes para projeto de bancos de dados não incorporam as imposições e preceitos trazidos pela LGPD.

¹https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm

Este artigo apresenta uma estratégia, denominada LGPDbyD, para incorporar os requisitos e as restrições impostas pela LGPD ao projeto de bancos de dados. Para isso, adaptamos o modelo ER, o modelo Relacional e o comando CREATE TABLE. Adicionalmente, estendemos a ferramenta brModelo a fim de fornecer suporte à metodologia proposta neste trabalho. A LGPDbyD busca facilitar os processos de projeto de bancos de dados e auditoria em conformidade com a LGPD.

2. Trabalhos Relacionados

Por questões didáticas, esta seção está organizada em duas partes. Inicialmente, apresentamos os trabalhos que buscam, de alguma forma, assegurar a conformidade entre sistemas de informação e a LGPD. Em seguida, discutimos as pesquisas que tiveram por finalidade estender ou adaptar o projeto de bancos de dados para contextos específicos.

2.1. Conformidade entre Sistemas de Informação e a LGPD

Em [Araújo et al. 2021], os autores apresentam um mapeamento sistemático envolvendo tanto a GDPR (*General Data Protection Regulation*) quanto a LGPD. Adicionalmente, os autores propõem um método, denominado LGPD4BP (LGPD for *Business Process*), para obter a conformidade dos processos de negócios em relação a LGPD. O LGPD4BP é composto por um questionário de avaliação e um método de modelagem com um catálogo de padrões de modelagem. O método foi aplicado em um estudo de caso do colégio de aplicação da Universidade Federal de Pernambuco (UFPE) e validado por uma turma de pós-graduação que aplicou o método e respondeu um questionário sobre facilidade de uso e completude do método. Os resultados das avaliações dos alunos demonstraram que a parte mais difícil foi a modelagem do processo de negócio e não os componentes do método proposto. O método LGPD4BP orienta os analistas a avaliar a conformidade dos processos de negócio com a LGPD e guia os analistas a modelarem processos de acordo com esta legislação. No entanto, o método LGPD4BP não especifica quem são os atores, os agentes de tratamento e a autoridade certificadora, ou seja, não detalha como a modelagem foi realizada.

Em [Canedo et al. 2021], os autores discutem o processo executado para a implementação da LGPD em uma Agência da Administração Pública Federal (ADPF), usando a notação BPMN. Eles destacaram a necessidade de definir novos papéis e responsabilidades dentro das agências brasileiras da administração pública, a nível federal, uma vez que, com a vigência da nova legislação de proteção de dados pessoais, todo e qualquer órgão público deve estar aderente a essa nova Lei.

Em [Carauta Ribeiro and Dias Canedo 2020], os autores descreveram critérios e medidas que orientam a necessidade de cumprimento da LGPD nos processos de TIC na Universidade de Brasília (UnB). A pesquisa foi aplicada aos sistemas de *software* da UnB [Lachaud 2020]. Inicialmente, os autores utilizaram o método *Analytical Hierarchy Process* (AHP) para realizar uma análise de requisitos. Em seguida, eles aplicaram o Método de Classificação de Preferência para Avaliação de Enriquecimento (PROMETHEE) e o processo de Análise de Decisão Multicritério - MCDA para efetivação das normas conforme a LGPD. Neste trabalho, o nível de proteção de dados, o risco de segurança, a gravidade do evento e o risco de proteção de dados foram definidos como principais requisitos para segurança de dados pessoais na UnB. Por fim, o critério de risco de proteção de dados foi estabelecido como foco da implementação da LGPD da UnB.

Em [Shastri et al. 2019], os autores buscam entender como os SBDs podem ser afetados pela GDPR e fazem uma análise detalhada das operações típicas de SBDs à luz dos requisitos da GDPR. Além disso, o artigo propõe uma série de métricas para avaliar o desempenho dos SBDs em relação ao cumprimento da GDPR, tais como: o tempo de execução de consultas, o consumo de recursos e a eficácia das técnicas de anonimização e pseudonimização. Para validar a abordagem proposta, os autores conduziram uma série de experimentos usando diferentes tipos de bancos de dados e cargas de trabalho.

2.2. Adaptações no Projeto de Bancos de Dados

[Kamble 2008] propõe um modelo conceitual para lidar com dados que possuem múltiplas dimensões, oferecendo uma estrutura que permite a representação, organização e análise desses dados, além de compreender como diferentes dimensões relacionam-se entre si e como extrair informações dessas relações. O trabalho proposto em [Dani and Getta 2005] apresenta uma metodologia e uma simbologia para modelagem conceitual voltada para *Data Streams*, abordando os desafios da modelagem em computação com fluxo contínuo de dados, visando melhorar a eficiência e a eficácia do processamento de *Data Streams* em tempo real.

Uma extensão para o modelo Entidade Relacionamento (ER) é discutida em [Carvalho et al. 2023]. Os autores apresentam o modelo conceitual ER+, o qual fornece uma estrutura mais adequada para a modelagem de sistemas distribuídos com múltiplas camadas. Adicionalmente, os autores discutem como essa nova extensão lida com questões de escalabilidade e desempenho em sistemas distribuídos, com a inclusão de técnicas para distribuir a carga de trabalho entre os diferentes nós do sistema, com a otimização da comunicação entre as camadas e com a consistência dos dados em um ambiente distribuído.

[Khan et al. 2004] aborda a importância de integrar os requisitos e as restrições do negócio nos modelos conceituais de banco de dados, com destaque para a importância da comunicação eficaz entre os desenvolvedores de banco de dados e os *stakeholders* do negócio, garantindo que os requisitos do negócio sejam compreendidos e traduzidos corretamente para o modelo de banco de dados. Desta forma, as organizações podem desenvolver sistemas mais alinhados com suas necessidades específicas, resultando em maior eficiência, flexibilidade e satisfação dos usuários.

Em [Sarkar and Athanassoulis 2022], os autores apresentam uma extensão para linguagens de consulta que permite especificar políticas de exclusão de dados. Assim, os desenvolvedores podem definir regras para a exclusão automática de dados diretamente em um comando SQL (em geral na cláusula INSERT) com base em critérios como, por exemplo, o período de tempo em que o dado permanece armazenado. Desta forma, ao inserir uma determinada “tupla” é possível definir que esta deverá ser removida automaticamente após 5 anos, por exemplo. Essa estratégia é de fundamental importância em contextos onde a privacidade e a conformidade com diferentes legislações são preocupações relevantes. Já em [de Abreu et al. 2021], os autores discutem como garantir que o processamento de consultas em bancos de dados respeite os consentimentos dos usuários. A solução proposta baseia-se no desenvolvimento de extensões da linguagem SQL que possuem como finalidade incorporar considerações de consentimento durante o processamento das consultas, permitindo que os desenvolvedores expressem explicitamente nos comandos SQL as condições sob as quais os dados podem ser acessados e utilizados.

3. A Estratégia Proposta: LGPDbyD

Este artigo apresenta uma estratégia, denominada LGPDbyD, que tem por finalidade incorporar os requisitos e as restrições da LGPD no projeto de bancos de dados. Para isso, a estratégia proposta adiciona pequenas adaptações aos conceitos e notações comumente utilizadas nos projetos conceitual, lógico e físico. A ideia central consiste em alterar o mínimo possível os modelos já existentes.

Vale destacar que, em geral, o projeto de bancos de dados é suportada por ferramentas CASE (*Computer-Aided Software Engineering*)². As vantagens fornecidas pelas ferramentas CASE são inúmeras, entre as quais destacam-se: a capacidade de *feedback* ao usuário em termos de mudanças de requisitos, o aumento de produtividade no desenvolvimento de produtos de *software*, a diminuição do tempo gasto no desenvolvimento de software, a qualidade do sistema, a padronização, a capacidade de substituir pessoas em projetos e a amplitude para resolver problemas grandes e complicados [Favero 2019].

Existem uma gama de ferramentas comerciais que auxiliam o projeto de banco de dados. Entre as mais conhecidas tem-se ERWin e DBDesign. Porém, essas ferramentas não fornecem suporte ao projeto conceitual, somente aos projetos lógico e físico. Por outro lado, a ferramenta brModelo³ oferece suporte tanto ao projeto conceitual, quanto lógico e físico. Ela é amplamente utilizada em ambientes educacionais para ensinar e praticar conceitos relacionados ao projeto de banco de dados. O brModelo oferece uma interface gráfica amigável o qual simplifica o processo de criação e edição de modelos, tornando-a uma ferramenta valiosa para estudantes, professores e profissionais de tecnologia da informação [dos Santos Mello et al. 2021].

Neste contexto, uma segunda contribuição deste artigo consiste em uma extensão da ferramenta brModelo, denominada brModeloPD, que incorpora as notações da metodologia LGPDbyD, incluindo os projetos conceitual, lógico e físico. A ferramenta brModeloPD pode ser acessada por meio da *Web*⁴ e o seu código fonte está disponível de forma *online*⁵.

3.1. Projeto Conceitual

Inicialmente, propomos uma adaptação do modelo Entidade Relacionamento (ER), denominada ER-PD, com o objetivo de possibilitar o projeto conceitual de bancos de dados em conformidade com a LGPD, a qual permite representar os principais conceitos presentes na LGPD, tais como: dados pessoais, consentimento, tipo de tratamento a ser executado sobre os dados e titular de dados pessoais.

O **Titular dos Dados Pessoais**, conforme a própria LGPD especifica em seu Artigo 5º, 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 1).

²<https://www.iso.org/standard/43189.html>

³<https://docs.brmodeloweb.com/#/logical-model/README>

⁴[https://app.brmodeloweb.com/!](https://app.brmodeloweb.com/)

⁵<https://github.com/jmmfilho/lgpdbyd>

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 Criptografia não é mencionada em nenhum momento na LGPD, mas é uma das alternativas comumente utilizadas para assegurar a anonimização de dados. 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 11 novos tipos de atributos (Figura 1), cujos principais são: “atributo pessoal” (P), atributo “sensível” (S), atributo “anonimizado” (A) e atributo “criptografado” (C). Adicionalmente, o modelo ER-PD adicionou também outros dois novos tipos de atributos: atributo “identificador” (I) e atributo “Semi-Identificador” (SI), a fim de representar conceitos comumente utilizados na área de privacidade de dados. Um atributo **Identificador** é aquele que podem identificar indivíduos de forma única, como, por exemplo: CPF, nome, email. Já um atributo **Semi-Identificador** é aquele que não identifica explicitamente um indivíduo, mas que, quando combinado com outros atributos possibilita a identificação de um indivíduo [Brito and Machado 2017]. Data de nascimento e CEP são exemplos de atributos semi-identificadores. Vale ressaltar que, caso deseje-se tratar a privacidade de atributos “pessoais” ou “sensíveis”, em geral, aplica-se um método de anonimização ou criptografia.

De acordo com a LGPD, o consentimento é uma declaração explícita do titular que concorda com o uso dos seus dados para uma determinada finalidade (Art. 5º, XII). Ademais, o titular dos dados pessoais pode indicar por quanto tempo seus dados podem ser utilizados. A fim de representar o conceito de consentimento, previsto na LGPD, o modelo ER-PD propõe dois novos tipos de atributos: atributos relacionados ao **Período de Consentimento** e atributos relacionados à **Finalidade**, ou seja, ao propósito de utilização dos dados (Figura 1).

A LGPD estabelece regras específicas para o tratamento de dados relativos a **Crianças e Adolescentes**, Artigo 14 e seus parágrafos. Para representar essa característica, o modelo ER-PD adiciona um novo tipo de atributo denominado “Criança e Adolescente”. Por fim, o titular dos dados pode autorizar que os seus dados, ou parte deles, possam ser compartilhados com terceiros. Para representar esse requisito, o modelo ER-PD propõe um novo tipo de atributo denominado atributo “Compartilhado”. A Figura 1 ilustra a notação adicionada pelo modelo ER-PD.













Símbolo	Representação	Símbolo	Representação
	Entidade Titular	PC 	Atributo Período de Consentimento
P 	Atributo Pessoal	F 	Atributo Finalidade
S 	Atributo Sensível	CP 	Atributo Compartilhado
A 	Atributo Anonimizado	CAD 	Atributo de Criança e Adolescente
C 	Atributo Criptografado	I 	Atributo Identificador
CS 	Atributo de Consentimento	SI 	Atributo Semi-Identificador

Figura 1. Notação do Modelo ER-PD.

3.1.1. Exemplo de Execução

A seguir, iremos ilustrar a utilização do modelo ER-PD no projeto conceitual de bancos de dados aderentes à LGPD. Inicialmente, considere que uma clínica de exames médicos deseja projetar um banco de dados para armazenar as informações de pacientes e exames. Assuma que um paciente pode realizar zero ou mais exames e que um exame pode ser realizado por zero ou mais pacientes.

Dos pacientes deseja-se armazenar: cod-paciente, cpf, nome, data-nascimento, endereço, cor, religião, sexo e gênero. Observe que todos esses atributos se referem a dados pessoais. Considere ainda que o “Controlador de Dados” da referida clínica definiu que os atributos cod-paciente e cpf devem ser criptografados, além disso ele determinou que os atributos nome e data-nascimento devem ser anonimizados. Para o “Controlador de Dados”, os atributos cor, religião, sexo e gênero são atributos pessoais sensíveis, ou seja, requerem um cuidado especial. Porém, nenhum tipo de tratamento foi especificado para esses atributos. Ademais, sabe-se que o atributo endereço refere-se a um dado pessoal, mas também constitui um semi-identificador. Contudo, nenhum tratamento especial foi definido para o atributo endereço. Com a finalidade de modelar aspectos relacionados ao consentimento, o “Controlador de Dados” solicitou a criação dos atributos: descrição-consentimento, início-consentimento e fim-consentimento. Já para endereçar o conceito de finalidade, o “Controlador de Dados” solicitou a criação do atributo descrição-finalidade. Dos exames deseja-se armazenar: cod-exame, descrição e valor. Observe que nenhum desses atributos refere-se a dados pessoais. Note ainda que o relacionamento entre paciente e exame possui dois atributos: data-exame e resultado. Considere que o “Controlador de Dados” definiu que o resultado do exame, que é um dado pessoal, deve ser criptografado, e que o atributo data-exame é um semi-identificador, cujo tratamento não foi especificado.

A Figura 2 ilustra o esquema conceitual, gerado na fase de projeto conceitual, utilizando o modelo ER-PD, modelado com a utilização da ferramenta brModeloPD. Observe que o conjunto entidade “Paciente” é representado por um retângulo com linha pontilhada, indicando que este representa um “Titular de Dados Pessoais”. Note também que ao lado do nome de cada atributo, entre colchetes, o modelo destaca o tipo do dado e o tratamento que deve realizado sobre ele.

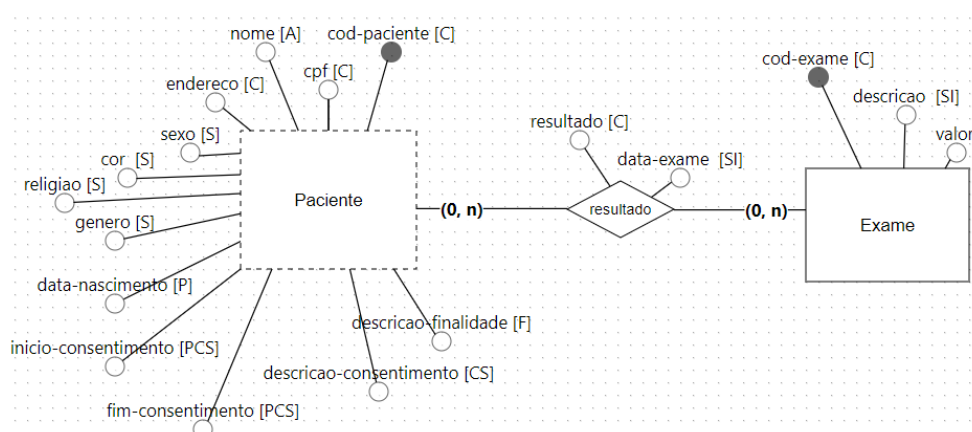


Figura 2. Esquema Conceitual Utilizando a Ferramenta brModeloPD.

3.2. Projeto Lógico

A próxima etapa no projeto de banco de dados é denominada projeto lógico e tem por finalidade produzir um esquema lógico para o banco de dados, utilizando como entrada o esquema conceitual desenvolvido durante o projeto conceitual. O esquema lógico é representado por meio de um modelo de dados utilizado por um sistema de banco de dados comercial, o qual é chamado genericamente de modelo lógico. O modelo lógico mais frequentemente usado é o modelo relacional e o esquema é usualmente representado por meio do **diagrama relacional (DR)**.

Neste trabalho, propomos uma adaptação do modelo Relacional, denominado R-PD, com o objetivo de possibilitar o projeto lógico de bancos de dados em conformidade com a LGPD. O modelo R-PD permite representar os principais conceitos presentes na LGPD, tais como: dados pessoais, tipo de tratamento executado sobre os dados e titular de dados pessoais. O modelo R-PD possibilita ainda representar atributos relacionados ao consentimento, finalidade, bem como a dados de crianças e adolescentes. A Figura 3 ilustra a notação utilizada 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 o tipo do dado e o tratamento que deve realizado sobre ele.

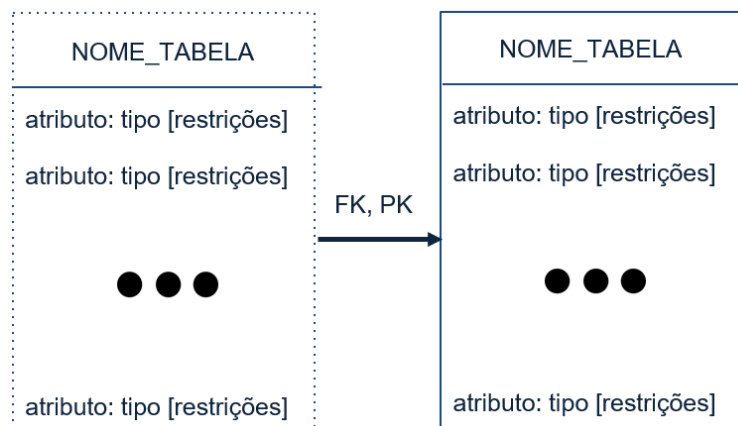


Figura 3. Notação do Modelo R-PD.

3.2.1. Exemplo de Execução

A seguir, iremos ilustrar a utilização do modelo R-PD no projeto lógico 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.

Inicialmente, vamos realizar o mapeamento do esquema conceitual para o esquema lógico. Neste sentido, cada conjunto entidade presente no esquema conceitual será mapeado para uma relação (tabela) no esquema lógico. Em seguida, mapeamos os atributos, definindo, para cada atributo, o seu tipo de dados e suas restrições (*NOT NULL*, *UNIQUE*, *Primary Key - PK* e *Foreign Key - FK*). Posteriormente, mapeamos, para cada atributo, os conceitos particulares da LGPD.

Atributo	Tipo	NOT NULL	UNIQUE	PK	FK	P	S	A	C	CS	PCS	F	CP	CAD	I	SI
cod-paciente	integer	S	S	S	N	S	S	S	S	N	N	N	N	N	S	N
cpf	varchar	S	S	N	N	S	S	S	S	N	N	N	N	N	S	N
nome	varchar	S	N	N	N	S	S	S	N	N	N	N	N	S	N	S
endereco	varchar	N	N	N	N	S	S	S	N	N	N	N	N	N	S	N
data-nascimento	date	N	N	N	N	S	S	N	N	N	N	N	N	N	S	S
sexo	char	N	N	N	N	S	S	N	N	N	N	N	N	N	N	N
cor	char	N	N	N	N	S	S	N	N	N	N	N	N	N	N	N
religiao	varchar	N	N	N	N	S	S	N	N	N	N	N	N	N	N	N
genero	varchar	N	N	N	N	S	S	S	N	N	N	N	N	N	N	N
inicio-consentimento	date	N	N	N	N	N	N	N	N	S	S	N	N	N	N	N
fim-consentimento	date	N	N	N	N	N	N	N	N	S	S	N	N	N	N	N
descricao-consentimento	varchar	N	N	N	N	N	N	N	N	S	N	N	N	N	N	N
descricao-finalidade	varchar	N	N	N	N	N	N	N	N	N	N	S	N	N	N	N

Tabela 1. Representação da Relação Paciente em Modo Tabular.

Atributo	Tipo	NOT NULL	UNIQUE	PK	FK	P	S	A	C	CS	PCS	F	CP	CAD	I	SI
cod-exame	integer	S	S	S	N	N	N	S	S	N	N	N	N	N	N	N
descricao	varchar	N	N	N	N	N	N	N	N	N	N	N	N	N	N	S
valor	char	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Tabela 2. Representação da Relação Exame em Modo Tabular.

A Tabela 1 ilustra os atributos da relação “Paciente” em modo tabular. Já a Tabela 2 mostra os atributos da relação “Exame”. Observe que o atributo “cod-paciente” da relação “Paciente” representa um dado pessoal e sensível, por este motivo utilizamos o valor **S** nas colunas **P** e **S**. Ademais, foi especificado que o atributo “cod-paciente” deveria ser criptografado, o que justifica a utilização do valor **S** nas colunas **A** e **C**. Como **C** é tipo mais restritivo, destacamos essa coluna em negrito e vermelho.

O próximo 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”. A Tabela 3 ilustra os atributos da relação “Resultado” em formato tabular. Note que o atributo “cod-exame” da relação “Exame” não representa um dado pessoal. Todavia, ele será criptografado, uma vez que este atributo irá compor a chave primária da relação “Resultado”, a qual contém o resultado de um certo exame realizado por um determinado paciente, o qual consiste em um atributo pessoal e sensível, que também deve ser criptografado.

Atributo	Tipo	NOT NULL	UNIQUE	PK	FK	P	S	A	C	CS	PCS	F	CP	CAD	I	SI
cod-paciente	integer	S	S	S	S	S	S	S	S	N	N	N	S	N	S	N
cod-exame	integer	S	S	S	S	N	N	S	S	N	N	N	N	N	S	N
resultado	varchar	N	N	N	N	S	S	S	S	N	N	N	N	N	N	N
data-exame	date	N	N	S	N	N	N	N	N	N	N	N	N	N	N	S

Tabela 3. Representação da Relação Resultado em Modo Tabular.

A Figura 4 ilustra o esquema lógico para o exemplo de execução previamente descrito, utilizando a ferramenta brModeloPD. Observe que a relação “Paciente” é representada por um retângulo com linhas pontilhadas, pois trata-se de um “Titular de Dados”.

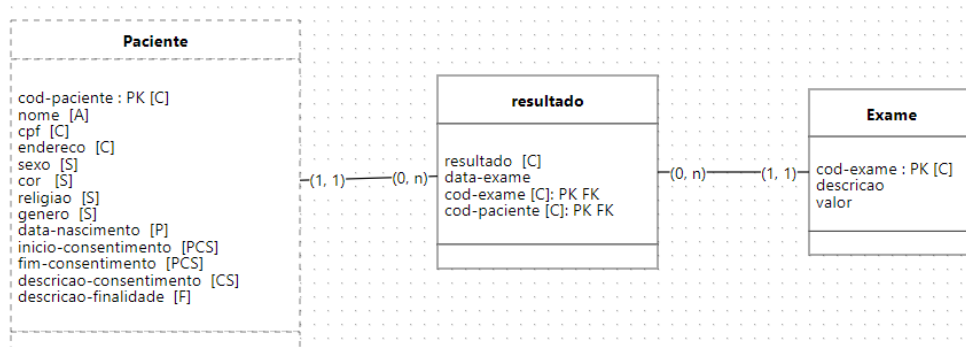


Figura 4. Esquema Lógico Utilizando a Ferramenta brModeloPD.

3.3. Projeto Físico

A 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 (*Data Definition Language*).

Neste trabalho, propomos uma adaptação no comando SQL CREATE TABLE, denominada SQL-PD, com o objetivo de possibilitar o projeto físico de bancos de dados em conformidade com a LGPD. Essa adaptação 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 gramática do comando CREATE TABLE adaptado (SQL-PD) é exibida na Listagem 1. Observe que 11 novos tipos de restrições foram adicionados: PESSOAL, SENSIVEL, ANONIMIZADO, CRIPTOGRAFADO, IDENTIFICADOR e SEMI IDENTIFICADOR, CONSENTIMENTO, PERIODO CONSENTIMENTO, FINALIDADE, COMPARTILHADO e CRIANCA E ADOLESCENTE.

Listagem 1. Comando CREATE TABLE Estendido (SQL-PD)

```
CREATE TABLE nome-tabela
(nome-coluna 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
| CONSENTIMENTO nome-coluna
| PERIODO CONSENTIMENTO nome-coluna
| FINALIDADE nome-coluna
| IDENTIFICADOR nome-coluna
| SEMI IDENTIFICADOR nome-coluna
)
```

3.3.1. 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 exemplo de execução descrito na seção anterior, envolvendo uma clínica de exames médicos que deseja armazenar informações de pacientes e exames. As Listagens 2, 3 e 4 ilustram os comandos CREATE TABLE gerado pela ferramenta brModeloPD para as tabelas “Paciente”, “Exame” e “resultado”, respectivamente. Observe que as restrições advindas da LGPD são inseridas por meio de “comentários” (neste exemplo, seguindo a sintaxe do PostgreSQL).

Listagem 2. Comando CREATE TABLE Paciente (SQL-PD)

```
CREATE TABLE Paciente
(cod-paciente integer NOT NULL,
cpf char NOT NULL,
nome varchar NULL,
endereco varchar NULL,
data-nascimento date NULL,
sexo char NULL,
cor char NULL,
religiao char NULL,
genero varchar NULL,
inicio-consentimento date NULL,
fim-consentimento date NULL,
descricao-consentimento varchar NULL,
descricao-finalidade 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, */
/* CONSTRAINT c11 Finalidade descricao-finalidade, */
/* CONSTRAINT c12 Consentimento descricao-consentimento, */
/* CONSTRAINT c13 Periodo Consentimento inicio-consentimento, */
/* CONSTRAINT c14 Periodo Consentimento fim-consentimento */
)
```

Listagem 3. Comando CREATE TABLE 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 4. Comando CREATE TABLE 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 FOREIGN KEY cod-paciente REFERENCES Paciente (cod-paciente),
CONSTRAINT c3 FOREIGN KEY cod-exame REFERENCES Exame (cod-exame)
/* , */
/* CONSTRAINT c4 Criptografado cod-paciente, */
/* CONSTRAINT c5 Criptografado resultado */
)
```

4. Ferramentas de Suporte à Conformidade com a LGPD

A partir dos metadados gerados no projeto físico, é possível identificar que atributos são pessoais ou não, que atributos são sensíveis, que atributos devem ser anonimizados ou criptografados etc. Desta forma, pode-se pensar em desenvolver ferramentas que façam verificações automáticas nas próprias tabelas do banco de dados, a fim de averiguar se o que está armazenado no banco é, de fato o que foi definido no projeto físico. Ademais, seria possível também desenvolver ferramentas para guiar atividades de auditoria de conformidade com a LGPD.

4.1. Ferramenta de Checagem Automática/Auditoria

A ideia dessa ferramenta consiste em utilizar os metadados adicionados no projeto físico (em cada comando CREATE TABLE) para verificar, por exemplo, se os atributos definidos como anonimizados ou criptografados de fato estão armazenados no banco de dados com o tratamento previamente definido. Após estas verificações, a ferramenta poderia gerar um relatório de conformidade. Como o Relatório em mãos, o Controlador ou DPO poderá solicitar as mudanças necessárias para que a empresa ou organização esteja em conformidade com a LGPD.

4.2. Criação de Pacotes/APIs para Desenvolvedores

A proposta dessa ferramenta consiste em disponibilizar pacotes (*packages*), ou APIs, independentes de SGBDs, contendo a implementação de diferentes métodos de anonimização e criptografia. De posse desses pacotes, o programador/desenvolvedor pode criar *Triggers* e *Procedures* para implementar, no próprio sistema de banco de dados, o tratamento adequado, e 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, a partir do pacote disponibilizado, a função mais adequada e cria *Triggers* que serão executados durante a inserção ou atualização desse atributo. Desta forma, a ferramenta proposta pode contribuir para diminuir o tempo de desenvolvimento e de adequação à LGPD.

5. Conclusões e Trabalhos Futuros

Neste trabalho, apresentamos uma estratégia, denominada LGPDbyD, para incorporar os requisitos e as restrições da LGPD no projeto de bancos de dados. Para isso, adicionamos pequenas adaptações no modelo ER, no modelo Relacional e no comando CREATE TABLE. Adicionalmente, estendemos a ferramenta brModelo a fim de fornecer suporte à metodologia LGPDbyD. A metodologia proposta busca facilitar os processos de projeto de bancos de dados e auditoria em conformidade com a LGPD, além de englobar as especificações dispostas nos Capítulos I, II, III, VI, VII e X da lei, haja vista serem dispositivos que toda e qualquer empresa, seja pública ou privada, deve atender para estar em *compliance* com a LGPD.

Como trabalhos futuros pretendemos desenvolver ferramentas que auxiliem o projeto, a implementação e auditoria de bancos de dados em conformidade com a LGPD. Adicionalmente, buscaremos fornecer suporte ao conceito de finalidade (ou propósito), a fim de assegurar que um determinado dado somente seja utilizado para a finalidade especificada pelo Titular. Por fim, planejamos fornecer suporte ao conceito de período de consentimento, possibilitando, por exemplo, remover dados automaticamente sempre que o período de utilização especificado pelo Titular dos dados seja ultrapassado.

Referências

- Araújo, E., Vilela, J., Silva, C., and Alves, C. (2021). Are my business process models compliant with lgpd? the lgpd4bp method to evaluate and to model lgpd aware business processes. In *XVII Brazilian Symposium on Information Systems*, pages 1–9. Sociedade Brasileira de Computação.
- 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.
- Canedo, E. D., Cerqueira, A. J., Gravina, R. M., Ribeiro, V. C., Camoes, R., dos Reis, V. E., de Mendonça, F. L. L., and de Sousa Jr, R. T. (2021). Proposal of an implementation process for the brazilian general data protection law (lgpd). In *ICEIS (1)*, pages 19–30. Sociedade Brasileira de Computação.
- Carauta Ribeiro, R. and Dias Canedo, E. (2020). Using mcda for selecting criteria of lgpd compliant personal data security. In *The 21st Annual International Conference on Digital Government Research*, pages 175–184.
- Carvalho, G., Bernardino, J., Pereira, V., and Cabral, B. (2023). Er+: A conceptual model for distributed multilayer systems. *IEEE Access*, 11:62744–62757.
- Dani, A. and Getta, J. (2005). Conceptual modelling of computations on data streams. *Proceedings of the 2nd Asia-Pacific Conference on Conceptual Modelling*, 43.
- 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.
- dos Santos Mello, R., Cândido, C. H., and Neto, M. B. S. (2021). brmodelo: An initiative for aiding database design. volume 12.
- Favero, E. S. (2019). Um protótipo de referência para ferramentas case de modelagem em ambiente web. *Universidade Federal do Pampa; (2019); 105*.
- Kamble, A. S. (2008). A conceptual model for multidimensional data. In *APCCM*, volume 8, pages 29–38.
- 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.
- Lachaud, E. (2020). Iso/iec 27701 standard: Threats and opportunities for gdpr certification. *Eur. Data Prot. L. Rev.*, 6:194.
- 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.
- Shastri, S., Banakar, V., Wasserman, M., Kumar, A., and Chidambaram, V. (2019). Understanding and benchmarking the impact of gdpr on database systems. *arXiv preprint arXiv:1910.00728*.