brModeloPD: Uma Extensão da Ferramenta BrModelo para Incorporar os Requisitos e as Restrições da LGPD ao Projeto de Banco de Dados

Patricia Vieira da S. Barros[1], **José Caique Militão França**[2], **José Maria da S. M. Filho**[1], **Javam C. Machado**[1]

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

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

caiqueeng@alu.ufc.br,

Abstract. *The General Personal Data Protection Law (LGPD) establishes how personal data should be processed, stored, and discarded, including in digital environments, always with prior authorization from the grantor. On the other hand, current information systems are heavily based on the use of personal data and, therefore, need to comply with the LGPD. In this context, the database system becomes an even more important component in software development, as it is responsible for storing, updating, and retrieving data. However, the tools used for database design do not incorporate the requirements and restrictions of the LGPD, making it difficult to ensure compliance between databases and current legislation. To address this, we extended the brModelo tool to provide support for the requirements and restrictions of the LGPD.*

Resumo. *A Lei Geral de Proteção de Dados Pessoais (LGPD) determina 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 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 extensão da ferramenta brModelo, denominada brModeloPD, que possibilita incorporar as imposições e preceitos da LGPD ao projeto de bancos de dados. A ferramenta bRModeloPD fornece suporte aos projetos conceitual, lógico e físico.*

1. Introdução

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 deve ser realizado o tratamento, o armazenamento e o descarte de dados pessoais, buscando proteger os direitos fundamentais de liberdade e de privacidade. A LGPD é uma legislação que envolve uma mudança de processos, além de mudar a cultura no dia a dia das empresas, na forma de tratar os dados pessoais.

Por outro lado, os Sistemas de Informação (SIs) atuais estão fortemente baseados na aquisição, armazenamento e processamento de dados pessoais. 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. 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. 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.

Este artigo apresenta uma extensão da ferramenta brModelo, denominada brModeloPD[2], que possibilita incorporar os requisitos e as restrições impostas pela LGPD ao projeto de bancos de dados. A ferramenta brModeloPD fornece suporte aos projetos conceitual, lógico e físico.

2. Trabalhos Relacionados

Em [Carvalho et al. 2023], ou autores propõem o modelo conceitual ER+, o qual fornece uma estrutura mais adequada para a modelagem de sistemas distribuídos com múltiplas camadas. 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. Uma extensão para linguagens de consulta que permite especificar políticas de exclusão de dados é discutida em [Sarkar and Athanassoulis 2022]. 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. 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.

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

[2] `http://200.129.44.243:9000/`

3. A Ferramenta brModeloPD

Este artigo apresenta uma extensão da ferramenta brModelo [3], denominada brModeloPD, que torna possível adicionar os requisitos da LGPD ao projeto de bancos de dados. A ferramenta brModeloPD oferece suporte tanto ao projeto conceitual, quanto lógico e físico de bancos de dados.

3.1. Exemplo de Execução

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 armazena-se: cod-paciente, cpf, nome, data-nascimento, endereço, cor, religião, sexo e gênero. Todos esses atributos referem-se a dados pessoais. O "Controlador de Dados" da referida clínica definiu que os atributos cod-paciente e cpf devem ser criptografados, além disso determinou que os atributos nome e data-nascimento devem ser anonimizados. Para o Controlador, 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 solicitou a criação dos atributos: descrição-consentimento, início-consentimento e fim-consentimento. Já para endereçar o conceito de finalidade, foi criado o atributo descrição-finalidade. Dos exames deseja-se armazenar: cod-exame, descrição e valor. Observe que nenhum desses atributos refere-se a dados pessoais. Além disso, o relacionamento entre paciente e exame possui dois atributos: data-exame e resultado. O Controlador 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. Observe na Figura 2 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.

3.2. Projeto Conceitual

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 5°, refere-se ao sujeito a quem a Lei pretende proteger. Portanto, 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, que devem ser particularmente protegidos. A notação utilizada para representar esse tipo específico de conjunto entidade, denominado "Titular", é um **retângulo com linhas tracejadas**.

[3]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. Uma das formas de tratar os dados sensíveis é a anonimização, conforme a LGPD e a criptografia não é mencionada em nenhum momento na Lei, 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: "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.

Em relação ao 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. O Artigo 14 e seus parágrafos estabelece regras específicas para o tratamento de dados relativos a **Crianças e Adolescentes**. Desta forma, tem-se o novo 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. O novo atributo proposto pelo modelo ER-PD é denominado "Compartilhado".

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

Figura 1. Notação do Modelo ER-PD.

A Figura 1 ilustra a notação adicionada pelo modelo ER-PD. Já a Figura 2 ilustra o esquema conceitual, gerado na fase de projeto conceitual, utilizando a ferramenta br-ModeloPD. Observe a adição de um novo tipo de conjunto entidade e dos novos tipos de atributos.

3.3. Projeto Lógico

O esquema lógico é representado por meio de um modelo de dados utilizado por um sistema de banco de dados comercial, chamado 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.

Para exemplificar o projeto lógico de bancos de dados aderentes à LGPD, vamos considerar o mesmo contexto descrito na seção anterior, envolvendo uma clínica de exames médicos que deseja armazenar informações de pacientes e exames.

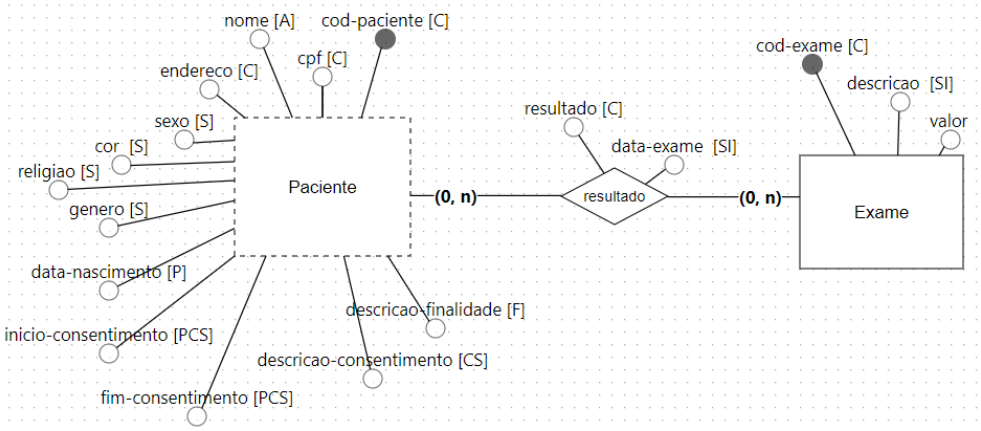

Figura 2. Esquema Conceitual Utilizando a Ferramenta brModeloPD.

A Figura 3 exibe 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".

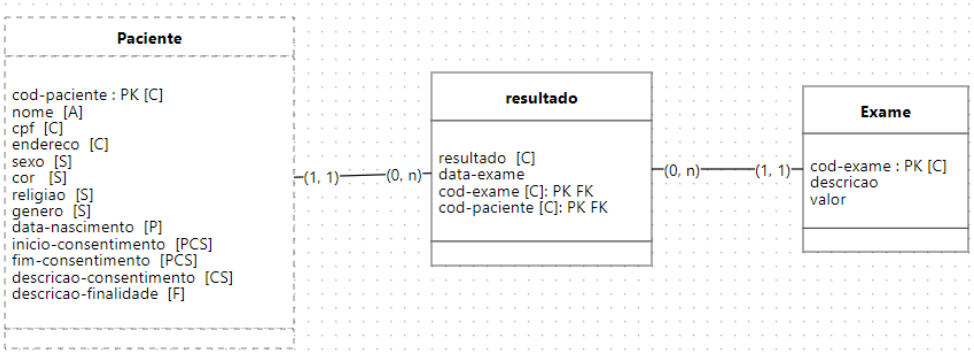

Figura 3. Esquema Lógico Utilizando a Ferramenta brModeloPD.

3.4. 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. 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), que poderão ser utilizados para auditorias de conformidade com a LGPD.

A Listagem 1 ilustra o comando CREATE TABLE gerado pela ferramenta brModeloPD para a tabela "Paciente". As restrições advindas da LGPD são inseridas por meio de "comentários" (neste exemplo, seguindo a sintaxe do PostgreSQL). Por motivo de falta de espaço, não iremos exibir os comandos para criar as tabelas "Exame" e "Resultado". Todavia, eles podem ser facilmente concebidos a partir da Listagem 1.

Listagem 1. 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 */
)
```

4. Conclusões e Trabalhos Futuros

Este trabalho propõe uma ferramenta denominada brModeloPD cuja finalidade consiste em fornecer suporte 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. 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

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

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.

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.