

# Uma Proposta de Modelagem de Dados no Domínio de Fraudes em Licitações Públicas

Ronaldo dos Santos Mello<sup>1</sup>, Luis Gustavo Bornia<sup>1</sup>, Hudson Afonso Batista da Silva<sup>1</sup>, Felipe Macacari Pierotti<sup>1</sup>, Marina Benvenuti Cardeal<sup>1</sup>

<sup>1</sup>Departamento de Informática e Estatística (INE) –  
Universidade Federal de Santa Catarina (UFSC)

r.mello@ufsc.br

luis.gb@posgrad.ufsc.br, hudson.afonso@posgrad.ufsc.br

f.pierotti@grad.ufsc.br, marina.benvenuti@grad.ufsc.br

**Abstract.** *The fight against corruption is a tireless task carried out by various government bodies, such as Public Ministries and Audit Courts, in order to value the morality of their processes. The discovery of frauds in bidding processes is one of these tasks, and automated systems have been proposed to support them. However, for such systems to be successful in their analysis, the data sources accessed by them must be rich in information about the domain entities and their various relationships. It requires a proper modeling of the domain data. This work presents a data modeling proposal for the domain of bids and frauds associated with them. Our intention is that this modeling serves as a guide for data repository projects in the domain. Unlike related works, this proposal presents a comprehensive data domain modeling and a polyglot persistence proposal.*

**Resumo.** *O combate à corrupção é uma incansável tarefa realizada por diversos órgãos governamentais, como Ministérios Públicos e Tribunais de Contas, no sentido de prezar pela moralidade dos seus processos. A descoberta de fraudes em processos licitatórios é uma dessas tarefas e sistemas automatizados vêm sendo propostos para apoiá-las. Entretanto, para que tais sistemas tenham êxito em suas análises é necessário que as fontes de dados acessadas por eles sejam ricas em informações a respeito das entidades do domínio e seus diversos relacionamentos. Isso requer uma modelagem adequada dos dados do domínio. Este trabalho apresenta uma proposta de modelagem de dados para o domínio de licitações e fraudes associadas a elas. A intenção é que esta modelagem sirva como guia para projetos de repositórios de dados no domínio. Diferente de trabalhos relacionados, esta proposta apresenta uma modelagem de dados abrangente no domínio e uma proposta de persistência poliglota.*

## 1. Introdução

A criação de licitações para aquisição ou alienação de bens, assim como para contratação de serviços por parte de órgãos públicos é uma exigência da legislação brasileira [Planalto.gov.br 2024]. Uma das finalidades desta legislação é regulamentar um processo licitatório pautado principalmente na impessoalidade, moralidade, publicidade e eficiência, além de facilitar a redação dos editais públicos vinculados às licitações.

Entretanto, diversas fraudes infelizmente ocorrem em tais processos principalmente para obtenção de vantagens financeiras por parte do contratante e/ou do contratado. Conluio e superfaturamento são exemplos de tipos de fraudes [Santos and Souza 2024].

Nesse contexto, algumas soluções automatizadas vêm sendo propostas para oferecer suporte ao combate desse tipo de corrupção [Elshaar and Sadaoui 2020, Abidi et al. 2021, Júnior et al. 2023, Brandão et al. 2024]. Muitas delas utilizam técnicas de aprendizado de máquina para desenvolver sistemas inteligentes que possam alertar especialistas, como promotores e auditores fiscais, sobre possíveis irregularidades em um processo licitatório. No entanto, para que tais sistemas tenham êxito em suas análises é necessário que as fontes de dados acessadas por eles sejam ricas em informações a respeito das entidades do domínio e seus diversos relacionamentos [Abidi et al. 2021, Brandão et al. 2024]. Para tanto, uma modelagem abrangente de dados, incluindo não apenas licitações, editais, empresas e produtos, mas também denúncias, fraudes e pessoas envolvidas, é fundamental para o projeto de um Banco de Dados (BD) robusto e útil.

Este trabalho apresenta uma proposta de modelagem conceitual de dados para o domínio de licitações e fraudes associadas a elas. Esta modelagem está sendo concebida e sistematicamente enriquecida no contexto do *Projeto Céos*, uma parceria entre a UFSC e o Ministério Público de Santa Catarina (MPSC) visando o desenvolvimento de um sistema inteligente baseado em alertas para a indicação de possíveis suspeitas de fraudes em processos licitatórios<sup>1</sup>. Diferente de trabalhos relacionados, esta proposta apresenta uma modelagem detalhada e uma proposta de persistência poliglota.

Este artigo possui mais três seções. A Seção 2 discute iniciativas relacionadas. A Seção 3 detalha a modelagem proposta e a Seção 4 é dedicada à conclusão.

## 2. Trabalhos Relacionados

Conforme salientado anteriormente, grande parte da literatura relacionada a fraudes e licitações aplicam técnicas de aprendizado de máquina para apoiar a descoberta de irregularidades em processos licitatórios, porém não apresentam detalhes sobre seus repositórios de dados e modelagens associadas. Exceções são 2 trabalhos que propõem esquemas de dados do domínio utilizando BDs de grafos [Erven et al. 2017, da Silva 2019].

O foco de um dos trabalhos é a investigação de relações suspeitas entre prefeitos, licitantes e políticos. Assim sendo, sua modelagem de dados contempla *licitações, unidades gestoras, municípios, políticos, partidos políticos e licitantes*. Já o segundo trabalho avalia a eficiência da utilização de um banco de dados de grafos na consulta a dados do domínio. Sua modelagem considera *licitações* e seus *itens, participantes, PJs, sócios, PFs, ministérios* e seus *departamentos*. Ambos os trabalhos propõem suas modelagens a partir da análise de dados coletados de órgãos públicos: TCE-PB, TSE e CGU.

Diferente desses trabalhos relacionados, a modelagem de dados aqui apresentada é bem mais abrangente em termos de fatos considerados do domínio e propõe um projeto lógico de persistência poliglota em 2 modelos de BD.

---

<sup>1</sup><https://ceos.ufsc.br/>

### 3. Modelagem de Dados Proposta

Conforme já mencionado, esta proposta está inserida no contexto do *Projeto Céos*, uma parceria entre UFSC e MPSC para o desenvolvimento de um sistema inteligente de apoio à detecção de fraudes em licitações públicas. Este sistema conta com um BD poliglota, isto é, um BD composto de repositórios baseados em diferentes modelos de dados, como o modelo relacional, orientado a documento, orientado a grafos, dentre outros. Independente do(s) modelo(s) de dados utilizado(s) para a persistência dos dados do domínio, uma metodologia clássica de projeto de BD foi adotada, composta de 3 fases: (i) *modelagem conceitual*; (ii) *modelagem lógica* e; (iii) *modelagem física ou implementação* [Heuser 2009]. O foco deste artigo é na modelagem conceitual de dados.

#### 3.1. Modelagem Preliminar

A definição da modelagem conceitual teve início com o levantamento de requisitos de dados através do estudo do domínio (coleta de documentação jurídico/técnica e material bibliográfico) e reuniões com a equipe de Tecnologia da Informação do MPSC. Como resultado, uma primeira versão da modelagem foi produzida. Ela é apresentada na Figura 1 e apenas os principais atributos são mostrados para não prejudicar a legibilidade.

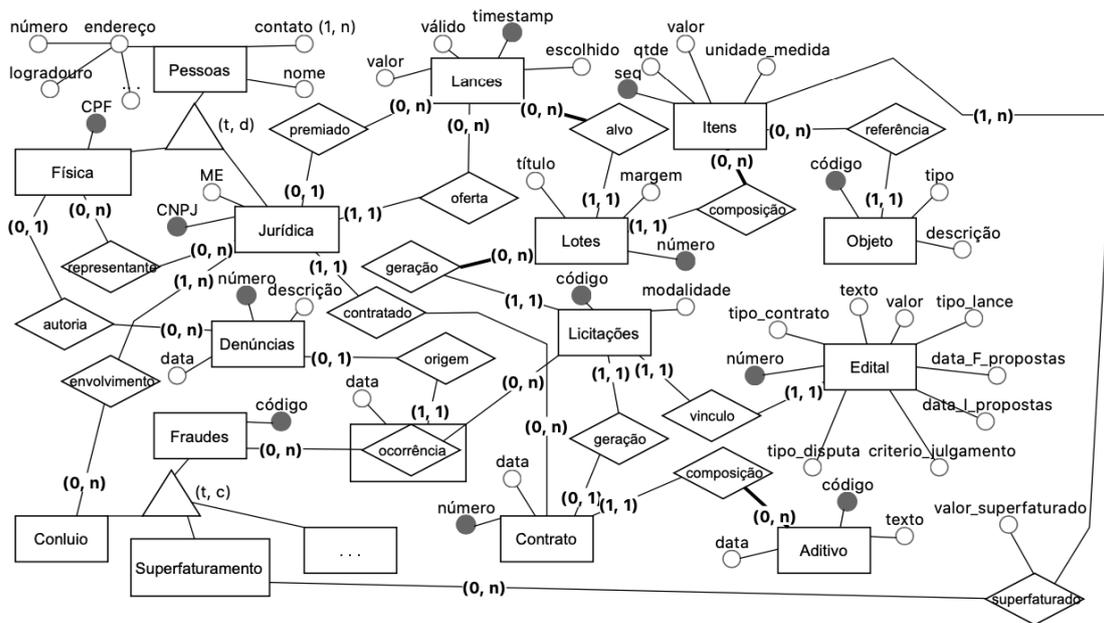


Figure 1. Modelagem conceitual preliminar.

Esta modelagem basicamente representa dados de: (i) licitações com seus editais e eventuais contratos; (ii) uma tipologia de fraudes associadas a licitações, com a possível indicação da denúncia que a originou; (iii) lotes de objetos (do tipo *produto* ou *serviço*) comprados ou alienados, e lances dados para eles; (iv) pessoas físicas (PFs) e jurídicas (PJs) envolvidas em licitações (participante e/ou vencedora) ou potenciais fraudes. Duas especializações de fraudes são exemplificadas (*Conluio* e *Superfaturamento*), com a indicação das PJs envolvidas e dos itens superfaturados, respectivamente. Edital é um dos mais importantes documentos associados a uma licitação, apresentando vários dados relevantes para fins de análise, como valor, tipos de contrato, de disputa e



etc), unidade gestora (empresa contratante) e ente (município ou estado) são interessantes quando se deseja filtrar licitações por local ou categoria, por exemplo.

Ao lado esquerdo da Figura 2 percebe-se também a modelagem de pessoas físicas e jurídicas cujas associações podem levantar suspeitas. Um exemplo é um indivíduo, sócio participante de um convênio associado a uma licitação, que também seja político, podendo este político ter uma idoneidade duvidosa ou ser representante do mesmo ente ao qual a licitação está vinculada. Isso sem contar agentes públicos participantes da comissão gestora de um processo licitatório que possam ter, por exemplo, algum parentesco com sócios e/ou políticos. Por fim, a modelagem contempla ainda o acesso ao acervo de documentos (atas, pareceres, orçamentos, etc) associado a uma licitação.

### 3.3. Persistência Poliglota

As modelagens conceituais apresentadas nas seções anteriores serviram de base para a construção da primeira versão de um BD poliglota para o projeto, denominado *BD Céos*. O projeto lógico deste BD considerou 2 modelos de dados e metodologias de projeto de dados específicas para esses modelos [Heuser 2009, Frozza et al. 2022]: *relacional* e *orientado a documento*. O primeiro modelo foi utilizado para a persistência dos dados como um todo, sendo útil para buscas a dados relacionados, enquanto que o modelo de documentos foi escolhido para a representação das notas fiscais e seus itens quando se deseja um acesso rápido a toda estrutura aninhada de uma nota fiscal, uma vez que o modelo representa naturalmente dados agregados. A Figura 3 (a) ilustra parte da modelagem lógica relacional e a Figura 3 (b) a modelagem lógica orientada a documentos (b) definidas para o BD Céos. A modelagem relacional completa não é aqui apresentada devido a limitações de espaço do artigo, entretanto, detalhamos a seguir as estratégias de mapeamento aplicadas à modelagem conceitual visando facilitar a reprodutibilidade da modelagem relacional.

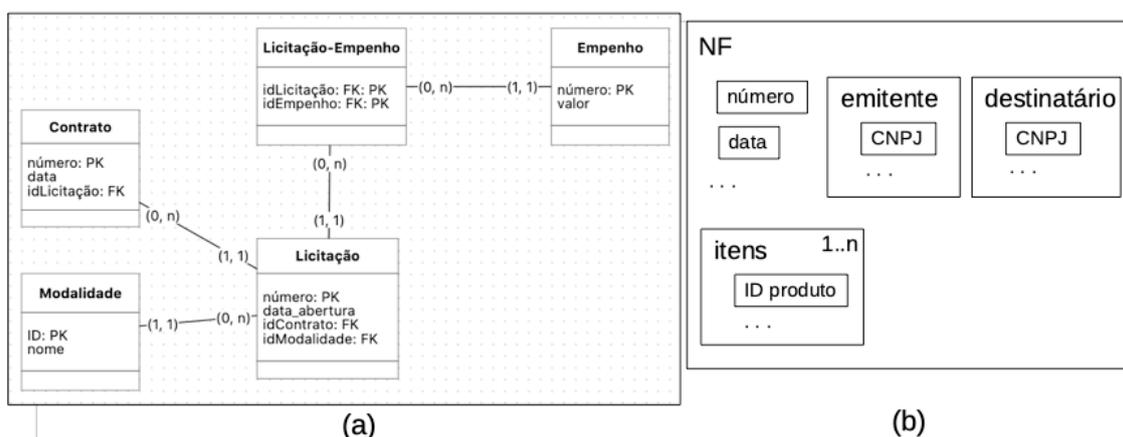


Figure 3. Modelagem lógica poliglota do BD Céos.

A modelagem relacional basicamente converte entidades e relacionamentos muitos-para-muitos para tabelas, como é o caso das tabelas *Licitação* e *Licitação-Empenho*, respectivamente. Os demais tipos de relacionamentos tornam-se em geral chaves estrangeiras (FKs), como é o caso do relacionamento entre licitação e modalidade, que gerou uma FK denominada *idModalidade* na tabela *Licitação*. Já a modelagem

orientada a documentos gera comumente dados aninhados, como é o caso de um objeto *nota fiscal (NF)* que possui um objeto aninhado representando os *itens* desta nota. As modelagens relacional e orientada a documentos foram fisicamente implementadas nos sistemas de BD *PostgreSQL* e *MongoDB*, respectivamente.

#### 4. Conclusão

Este trabalho apresenta uma modelagem de dados abrangente para o domínio de fraudes em licitações públicas. A principal contribuição desta proposta é fornecer um guia para projetos de repositórios de dados no contexto de sistemas informatizados voltados à integração e análise de dados neste domínio, ou mesmo a evolução/expansão dos esquemas de BDs presentes em sistemas voltados a este domínio.

Um exemplo de aplicação desta modelagem, conforme salientado na Seção 3.3, é o *BD Céos*, projetado no âmbito do *Projeto Céos* para integrar várias fontes de dados de órgãos governamentais brasileiros e servir de infraestrutura para um sistema inteligente que apoia a descoberta de fraudes em processos licitatórios. Um estudo de caso em andamento no âmbito do projeto é considerar características temporais e espaciais que influenciam no preço para avaliação de ocorrência de sobrepreço, definido pela Lei nº 14.133 como "*preço orçado para licitação ou contratado em valor expressivamente superior aos preços referenciais de mercado*" [Planalto.gov.br 2024].

A modelagem de dados proposta avança a literatura no sentido de propor uma modelagem mais ampla do domínio baseada fortemente em fontes de dados reais de órgãos governamentais em âmbito federal e estadual. Mesmo assim, pretende-se, como trabalhos futuros, a consideração de mais fontes de dados e a continuidade do enriquecimento da modelagem proposta visando fornecer mais subsídios para a análise de dados do domínio. Para tanto, métodos específicos de integração de dados focados no entendimento dos dados do domínio necessitarão ser projetados e desenvolvidos para possibilitar uma correta conexão de novos fatos identificados nas fontes com a modelagem de dados existente.

#### References

- Abidi, W. U. H. et al. (2021). Real-Time Shill Bidding Fraud Detection Empowered With Fussed Machine Learning. *IEEE Access*, 9:113612–113621.
- Brandão, M. A. et al. (2024). PLUS: A Semi-automated Pipeline for Fraud Detection in Public Bids. *Digit. Gov. Res. Pract.*, 5(1):5:1–5:16.
- da Silva, K. R. P. (2019). Análise de Relações Suspeitas em Licitações na Paraíba baseado em Teoria de Grafos. Trabalho de Conclusão de Curso de Sistemas de Informação, UFPB - <https://repositorio.ufpb.br/jspui/handle/123456789/17172>. Último acesso: 15 de julho de 2024.
- Dados.gov.br (2024). Portal de Dados Abertos. <https://dados.gov.br/>. Último acesso: 15 de julho de 2024.
- Elshaar, S. and Sadaoui, S. (2020). Detecting Bidding Fraud using a Few Labeled Data. In *XII International Conference on Agents and Artificial Intelligence, ICAART*, pages 17–25. SCITEPRESS.

- Erven, G. C. G. V., Holanda, M., and Carvalho, R. N. (2017). Detecting Evidence of Fraud in the Brazilian Government Using Graph Databases. In *Recent Advances in Information Systems and Technologies - Volume 2*, volume 570 of *Advances in Intelligent Systems and Computing*, pages 464–473. Springer.
- Frozza, A. A., Schreiner, G. A., and dos Santos Mello, R. (2022). Projeto de Bancos de Dados NoSQL. In *Tópicos em Gerenciamento de Dados e Informações: Minicursos do SBBDD 2022*, pages 26–52. Sociedade Brasileira de Computação.
- Heuser, C. A. (2009). *Projeto de Banco de Dados*. Sagra-Luzzato.
- Júnior, D. G. et al. (2023). Classificação de Fraudes em Licitações Públicas através do Agrupamento de Empresas em Conluíus. In *XI Workshop de Computação Aplicada em Governo Eletrônico, WCGE*, pages 13–24. Sociedade Brasileira de Computação.
- Planalto.gov.br (2024). Lei de Licitações e Contratos Administrativos. [https://www.planalto.gov.br/ccivil\\_03/\\_Ato2019-2022/2021/Lei/L14133.htm](https://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2021/Lei/L14133.htm). Último acesso: 15 de julho de 2024.
- Santos, F. B. and Souza, K. R. (2024). *Como Combater a Corrupção em Licitações: Detecção e Prevenção de Fraudes*. Fórum, 4 edition.