

# Dicionário semântico de dados para anotação de indicadores de desempenho para gestão hospitalar

## *Semantic data dictionary for annotating performance indicators for hospital management*

Evaldo de Oliveira da Silva<sup>1</sup>, Marcello Peixoto Bax<sup>2</sup>, Frederico César Mafra Pereira<sup>3</sup>, Yuri Bento Marques<sup>4</sup>, Eduardo Cardoso Melo<sup>5</sup>

<sup>1</sup> Centro Universitário Academia, Juiz de Fora, Minas Gerais, Brasil. ORCID: <https://orcid.org/0000-0001-8754-0891>

<sup>2</sup> Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brasil. ORCID: <https://orcid.org/0000-0003-0503-3031>

<sup>3</sup> Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brasil. ORCID: <https://orcid.org/0000-0002-1971-8069>

<sup>4</sup> Instituto Federal do Norte de Minas Gerais, Teófilo Otoni, Minas Gerais, Brasil. ORCID: <https://orcid.org/0000-0003-3414-2325>

<sup>5</sup> Instituto Federal do Norte de Minas Gerais, Teófilo Otoni, Minas Gerais, Brasil. ORCID: <https://orcid.org/0000-0002-1323-5859>

**Autor para correspondência/Mail to:** Evaldo de Oliveira da Silva, [evaldo.oliveira@gmail.com](mailto:evaldo.oliveira@gmail.com)

**Recebido/Submitted:** 21 de outubro de 2022; **Aceito/Approved:** 04 de fevereiro de 2023



Copyright © 2023 Silva, Pereira, Marques, & Melo. Todo o conteúdo da Revista (incluindo-se instruções, política editorial e modelos) está sob uma licença Creative Commons Atribuição 4.0 Internacional. Ao serem publicados por esta Revista, os artigos são de livre uso em ambientes educacionais, de pesquisa e não comerciais, com atribuição de autoria obrigatória. Mais informações em <http://revistas.ufpr.br/atoz/about/submissions#copyrightNotice>.

### Resumo

**Introdução:** a gestão hospitalar é uma atividade fundamental para atender a legislações e regulações, especialmente em momentos de crise sanitária. Estratégias de gestão utilizam diferentes indicadores (KPI ou *Key Performance Indicator*) para controlar os processos dentro de um hospital. KPIs nesse segmento podem ser a taxa de ocupação (diária ou mensal), o índice de permanência de pacientes, o índice de renovação de pacientes ou a relação de pacientes internados por classificações de doenças, entre outros. A gestão dos dados gerados pelos indicadores visa suportar a tomada de decisão e melhorar os serviços de saúde prestados pela organização. Boas práticas de nomeação de dados evitam combinações de dados incompatíveis para não comprometer a tomada de decisão. **Método:** foi aplicado neste artigo um processo sistemático para preparar e integrar dados de indicadores hospitalares com base na modelagem ontológica. **Resultados:** um processo para anotação semântica dos dados suportado pela técnica SDD (*Semantic Data Dictionary*) que usa *templates* de metadados para facilitar a preparação, integração e reuso de dados na área hospitalar, especificamente, para o indicador de tempo médio de internação. **Conclusão:** o uso de ontologias na anotação semântica permite desambiguar termos, preservar a semântica dos valores extraídos dos KPIs e abre caminho para a ingestão de dados de KPIs hospitalares a partir de diferentes fontes de dados da rede hospitalar (pública e privada). A abordagem apresentada contribui com a curadoria dos dados, uma vez que a técnica SDD segue as boas práticas para a gestão de dados em diferentes áreas.

**Palavras-chave:** Modelos Dimensionais; Indicador-chave de Desempenho (ICD); Gestão Hospitalar, Dicionário de Dados; Ontologia; Anotação Semântica.

### Abstract

**Introduction:** hospital management is a fundamental activity to comply with laws and regulations, especially in times of health crisis. Management strategies use different indicators (KPI or *Key Performance Indicator*) to control the processes within a hospital. KPIs in this segment can be the occupancy rate (daily or monthly), patient permanence index, patient renewal index or the list of hospitalized patients for disease classifications, among others. The management of the data generated by the indicators aims to support decision making and improve health services provided by the organization. Good data naming practices avoid incompatible data combinations so as not to compromise decision making. **Method:** it was applied in this article aims to put in practice a systematic process to prepare and integrate data from hospital indicators based on ontological modeling. **Results:** a process for semantic annotation of data supported by SDD (*Semantic Data Dictionary*) technique, which uses metadata templates to facilitate the preparation, integration and reuse of data in the hospital area, specifically for the average length of stay indicator. **Conclusions:** the use of ontologies in semantic annotation allows disambiguation of terms and preserves the semantics of values extracted from KPIs and opens the way for ingesting hospital KPI data from different data sources in the hospital network (public and private). Finally, the presented approach contributes to data curation, since the SDD technique follows good practices for data management in different areas.

**Keywords:** Dimensional Models; Key Performance Indicator (KPI); Hospital Management; Data Dictionary; Ontology; Semantic Annotation.

## INTRODUÇÃO

Indicadores de performance ou KPI (*Key Performance Indicator*) fornecem informações com valores mensuráveis que permitem avaliar o alcance de metas e rever processos para melhoria contínua das atividades, criando uma base analítica para a tomada de decisões que priorizam as ações apreciadas (empiricamente) como as mais relevantes. De forma geral, os KPIs possuem métricas que permitem a análise, por exemplo, de receitas, lucros, preços e custos, atividades, qualidade ou satisfação. As métricas fornecidas por KPIs são interpretadas por gestores que podem tomar decisões com base empírica ou científica. Exemplo comum de mensuração é o percentual de aderência da realização de atividades ao que foi planejado anteriormente. No âmbito da gestão hospitalar, KPIs são vistos como mecanismos de medição de performance organizacional (Parmenter, 2015).

A Agência Nacional de Saúde Suplementar (ANS) orienta a elaboração de diferentes KPIs que visam auxiliar na avaliação do desempenho das instituições hospitalares. Essa agência divulgou fichas completas definindo indicadores gerais de qualidade hospitalar que, juntamente com os indicadores específicos por linhas de cuidado, compõem o painel de indicadores de qualidade hospitalar. Para cada um dos indicadores, as fichas documentam a sua fórmula de cálculo, critérios de processamento, definição dos termos, unidade de medida e referência de meta (Agência Nacional de Saúde Suplementar, 2020).

A modelagem de KPIs deve ser condicionada em boas práticas de nomeação de dados como, por exemplo, em padrões de nomenclaturas (*naming conventions*) ou dicionários de dados (Kimball & Ross, 2013). No entanto é possível que somente a nomeação das coleções de dados e seus atributos não provenham uma compreensão correta para gerar os cálculos de um KPI. Com isso, os KPIs acabam resultando em combinações de dados incompatíveis, comprometendo os valores e prejudicando a tomada de decisão. Tem-se, portanto, clareza quanto à importância da garantia de nomeação dos dados para a geração de KPIs adequados.

No contexto da área da Ciência da Informação, a Curadoria Digital pode contribuir com a modelagem de KPIs. Técnicas de descrição de dados com metadados favorecem a qualidade e a preservação e facilitam a descoberta de novas informações e conhecimentos pelo reuso de dados (Medeiros, 2018; Pereira, 2021). Outras iniciativas, como padrões de dados abertos vinculados (Linked Open Data/LOD), visam melhorar a qualidade dos sistemas de informações nas mais diversas áreas, incluindo sistemas da área médica (Zaidan & Bax, 2013).

Apresenta-se neste artigo um “caso de uso” de anotação de dados de um modelo dimensional para cálculos de KPIs na área hospitalar, baseado na abordagem SDD (Rashid et al., 2017). Trata-se do relato de um experimento que procurou examinar o potencial de uso de SDDs no contexto da geração de indicadores na área hospitalar. A abordagem emprega ontologias para oferecer uma descrição detalhada de *datasets*, a ponto de ser automatizada e tratada por computador. Permite também o enfoque na semântica dos dados, incluindo informações que podem prontamente ser processadas. Ao considerar essas características dos SDDs, argumenta-se que é alcançada uma representação legível e padronizada por máquina para o registro de metadados com base em *datasets* e em metodologias que usam linguagens de mapeamento, a partir da simplificação dos requisitos de conhecimento de programação, separando os elementos de anotação daqueles relacionados com a abordagem do componente de software. De acordo com Bax e Silva (2020), a aplicação de SDDs em modelos dimensionais permite alinhar e harmonizar interpretações de conceitos, diferentes escalas e unidades de medida que descrevem os dados, facilitando a integração semântica dos dados de diferentes origens ou de organizações hospitalares diferentes.

O texto está organizado como segue: a Seção 2 traz o referencial teórico contemplando os conceitos de modelagem dimensional de dados com KPIs, ontologias, anotação de dados por meio de ontologias e trabalhos correlatos. A Seção 3 descreve o processo de anotação usando o SDD. A Seção 4 aplica o SDD, relatando a anotação necessária para criar KPIs voltados para gestão de indicadores na área hospitalar e apresenta os resultados. A Seção 5 traz as conclusões e considerações finais, além de sugerir trabalhos futuros.

## REFERENCIAL TEÓRICO E TRABALHOS CORRELATOS

### Modelagem de KPIs, Anotação Semântica de Dados e Ontologias

De acordo com Roldán-García, García-Niéto, Maté, e Aldana-Montes (2019), KPIs permitem que especialistas modelem objetivos empresariais por meio de variáveis quantitativas com faixas numéricas e limites claros. Os autores mostram que a utilidade dos KPIs foi comprovada em vários domínios, como Educação, Saúde e Agricultura. A *web* semântica surge como uma tecnologia poderosa para representação de conhecimento e modelagem de dados por meio de formatos de representação explícita e padrões como RDF (*Resource Description Framework*) e OWL. Os autores destacam que o uso dessas tecnologias e a anotação semântica de KPI enriquece a estratégia empresarial. Nesse sentido, Wetzstein, Ma, e Leymann (2008) propõem que os KPIs sejam modelados por analistas que demonstrem habilidades para explorar anotações semânticas de processos de negócios.

No entendimento de Kimball e Ross (2013), um modelo de dados dimensional agrupa facetas ao redor de dados numéricos, organizados em tabelas de fatos. Para Bax e Silva (2019, nov. 27-29), a análise dos fatos usa essas facetas para combinar filtros que atendam às necessidades do usuário e apoiem suas demandas para tomadas de decisão. O modelo dimensional da Figura 1, organizado como uma estrela (fatos no centro), permite armazenar dados sobre o indicador de “Tempo Médio de Internação (TMI)” (Agência Nacional de Saúde Suplementar, 2020).



Figura 1. Modelo dimensional de dados para o indicador TMI

Com base nos elementos desse modelo, o TMI pode ser consultado a partir da combinação das seguintes dimensões: CID (Classificação Internacional de Doenças), Faixa Etária, Tempo e Município de Origem. A tabela fato (FTempoMedioInternacao) possui chaves (estrangeiras) oriundas das outras tabelas e os atributos tempo médio de internação ( $TP_{MEDIO\_INTERNACAO}$ ), pacientes admitidos por dia no período ( $NU_{PCNT\_ADMTD\_PERIODO}$ ) e altas de pacientes no período ( $NU_{ALTA\_PCNT\_PERIODO}$ ), com os quais extrai-se o TMI a partir de funções de agregação e fórmulas. De acordo com a ANS, o tempo médio de internação é a mensuração do tempo médio, em dias, de permanência dos pacientes admitidos na instituição em determinado período. A fórmula de cálculo é a seguinte:

$$TMI = \frac{\sum \text{Pacientes Admitidos por Dia no período}}{\sum \text{Altas de pacientes no período}}$$

Em que “Pacientes admitidos por dia no período” é a medida da assistência prestada a um paciente internado durante o período de um dia hospitalar, ou seja, é o volume de pacientes que estão pernoitando no hospital em cada dia, independentemente do horário de admissão e desconsiderando-se o dia de saída. “Altas de pacientes no período” considera aquelas saídas da instituição que se dão por alta (cura, melhora, estado inalterado, evasão, desistência do tratamento, transferência externa) ou por óbito.

Ao anotar dados de forma semântica, é necessário que o domínio representado por um modelo conceitual seja compreendido. Selecionam-se dados para serem anotados pelos termos existentes e definidos por conceitos do domínio, sendo que a semântica dos dados é explicitada e formalizada por meio do uso de ontologias. A anotação permite gerar fragmentos de conhecimento do domínio que podem ser representados por grafos de conhecimento (Hogan, Blomqvist, Cochez, & Zimmermann, 2021).

Conforme explicado por Rashid et al. (2017), o SDD é uma abordagem de anotação de dados que manipula um conjunto de padrões de metadados fundamentados em ontologias que descrevem objetos, representados por dados, em classes, propriedades e relacionamentos. De acordo com Bax e Silva (2020), a anotação por SDD é um processo executado manualmente por especialistas do domínio, que associa os dados de um dataset a conceitos/classes das ontologias a fim de enriquecer os dados presentes nesse conjunto. Rashid et al. (2017) recomendam o uso da ontologia SIO (Semanticscience Integrated Ontology)<sup>1</sup>, que fornece propriedades para descrever os relacionamentos entre objetos e atributos como modelo de representação do conhecimento, facilitando a descoberta de conhecimento. A abordagem deve ser orientada por ontologistas, engenheiros do conhecimento ou cientistas da informação, que demonstrem compreensão do domínio, dos conceitos relacionados na ontologia e dos *datasets* anotados. A anotação utiliza um conjunto de documentos (*templates* de metadados), os quais serão posteriormente explicados neste artigo.

O SDD provê também a representação do conhecimento, podendo ser persistente e detectável, de forma que os dados estejam acessíveis para eventuais divulgações em plataformas abertas. O uso de vocabulários estruturados e de ontologias como melhores práticas compatíveis com RDF contribuem para que os dados mapeados sejam interoperáveis (Rashid et al., 2017).

<sup>1</sup> <https://bioportal.bioontology.org/ontologies/SIO>

Um grafo de conhecimento representa objetos de interesse para os especialistas do domínio, conjuntamente a conexões entre tais objetos. As restrições e propriedades na estrutura dos grafos são impostas por meio de ontologias. Grafos de conhecimento permitem que pessoas e diferentes aplicações reutilizem as definições nele modeladas, pavimentando o caminho para a inferência de novos fatos, enriquecendo o conhecimento e o seu compartilhamento. O padrão RDF pode ser usado para representar grafos de conhecimento a partir de declarações no formato de triplas, estas geradas como fragmentos do conhecimento resultante da anotação por SDD. A conceitualização do domínio por meio da formalização semântica de ontologias é uma das argumentações para o uso de SDDs. (Pan, 2017).

Este artigo utiliza a ontologia KPIOnto, que fornece conceitos e propriedades para descrever indicadores de desempenho por meio de recursos da *web* semântica, bem como para anotar e alinhar conceitualmente a compreensão de diferentes profissionais sobre o KPI modelado (Diamantini, Potena, & Storti, 2016). A KPIOnto constitui-se de classes como: Indicator, Dimension, AggregationFunction e Formula, sendo Indicator a principal. Ela especifica um indicador pelas propriedades: hasDimension, hasFormula e hasAggrFunction (para uso de funções de agregação) (Bax & Silva, 2019, nov. 27-29).

A Figura 2 apresenta as relações existentes na KPIOnto<sup>2</sup> com instâncias que se relacionam com os conceitos referentes ao modelo dimensional apresentado na Figura 1. As instâncias foram criadas com a ferramenta WebProtegé (Protegé, 2020). A instância KPITMI (em azul) é um tipo da classe Indicator. A relação “is-a” é representada pela linha pontilhada amarela. A instância KPITMI se relaciona com demais conceitos da KPIOnto, tais como AggregationFunction e Dimension pelas propriedades hasAggrFunction e hasDimension, respectivamente, estas representadas por linhas azuis sólidas. A instância AVGTimeOfHospitalization é um de tipo de AggregationFunction que representa o tempo médio de internação. As instâncias ICD, Time, AgeGroup e CountyOfOrigin são um tipo de Dimension.

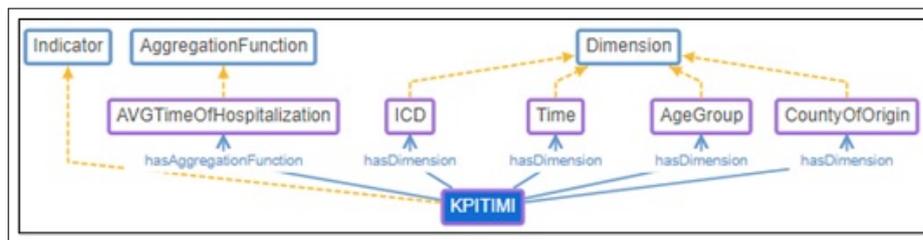


Figura 2. Ontologia KPIOnto com instâncias do indicador KPITMI  
Fonte: Diamantini et al. (2016).

## Trabalhos Correlatos

Tsai et al. (2015) coletaram dados de pesquisas em hospitais nos Estados Unidos e na Inglaterra para examinar as relações entre os conselhos de hospitais, as práticas dos gestores e a qualidade do atendimento prestado. Constataram que os hospitais com práticas de gestão mais eficazes forneciam cuidados de melhor qualidade e que hospitais com conselhos administrativos que usaram métricas de qualidade clínica de forma mais eficaz tiveram melhor desempenho da equipe de gestão do hospital no estabelecimento de metas e operações. Os autores mostram a importância da qualidade das variáveis das métricas de gestão dos hospitais para aumentar a compreensão nos processos de tomada de decisão e auxiliar as equipes gerenciais da linha de frente a incrementarem a qualidade do atendimento.

N4PCC<sup>3</sup> recrutou quatro hospitais que representam 285 leitos de internação na África e elaborou KPIs para permitir a medição de indicadores-chave de desempenho para cirurgia e avaliação de resultados perioperatórios<sup>4</sup> em conformidade com a OMS (Organização Mundial da Saúde). O desenvolvimento do conjunto de dados dos indicadores levantados representou um passo inicial para a construção de um registro nacional perioperatório e um banco de dados de resultados cirúrgicos, uma prioridade de consenso para a agenda de pesquisa dessa área, além de uma recomendação do roteiro para procedimentos cirúrgicos. De acordo com o N4PCC, o principal aprendizado com a implementação do registro dos indicadores incluiu: o uso de um conjunto de dados estreito projetado pelas partes interessadas que era consistente com suas prioridades, reduzindo assim a carga de captura de dados e minimizando o desperdício nessa atividade; uma plataforma baseada em aplicativos móveis de fácil utilização que superou a necessidade de treinamento extensivo do usuário final ou instalação de *software*; a capacidade de usuários previamente especificados fazerem download de seus próprios dados. Essa infraestrutura garantiu tanto a acessibilidade quanto a propriedade dos dados e promoveu os princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) na melhoria dos serviços de saúde.

<sup>2</sup>Especificação da KPIOnto, <https://kdmg.dii.univpm.it/kpionto/specification/>

<sup>3</sup><https://associationofanaesthetists-publications.onlinelibrary.wiley.com/doi/full/10.1111/anae.15394>

<sup>4</sup>Período dividido em duas subetapas: admissão hospitalar e controle de estudos pré-operatórios (Davrioux, Palermo, Serra, & Giménez, 2019).

Rashid e McGuinness (2020) avaliaram a aderência da técnica SDD (*Semantic Data Dictionary*) aos princípios FAIR, comparando-os a outras abordagens de anotação de dados. Na comparação, o SDD recebeu a maior nota (conforme visto na Seção 3). Geralmente acompanhando *datasets*, os dicionários de dados convencionais (não semânticos) facilitam o gerenciamento de dados. Para formalizar a descrição dos dados com os metadados dos dicionários, ontologias podem ser aplicadas, pois elas proveem enriquecimento semântico e formalização lógica do significado dos dados, evitando interpretações discrepantes.

Dang, Hedayati, Hampel, e Toklu (2008) utilizam ontologias como um modelo formal de representação declarativa do conhecimento para tornar os sistemas de informação da área da saúde adaptáveis e inteligentes, e, assim, servirem melhor os pacientes. Os autores apresentam uma estrutura de conhecimento ontológico que cobre os domínios da saúde de um hospital, desde as tarefas médicas ou administrativas, até ativos hospitalares, registros de pacientes e dados sobre medicamentos.

A utilização do SDD para anotar os dados de KPIs com base em uma modelagem ontológica é uma abordagem que diferencia este artigo dos trabalhos mencionados. A anotação é feita manualmente por especialistas que conhecem o domínio modelado para, em seguida, proceder-se com a geração de grafos de conhecimento de uma ontologia e de *templates* de metadados, sendo úteis para inferir novos conhecimentos que podem melhorar a tomada de decisão dentro da organização.

## MÉTODOS E TÉCNICAS

O `sdd2rdf`<sup>5</sup> é um *script* de *software* que interpreta o SDD a partir de um conjunto de *templates* (*InfoSheet*, *Dictionary Mapping*, *CodeBook*, *Code Mapping*, *TimeLine*, *Properties Table*) e converte os dados do *dataset* descrito pelo SDD em um grafo de conhecimento expresso no formato RDF e de acordo com a ontologia especificada, possibilitando a interoperabilidade dos dados. Para demonstrar as possibilidades de acesso aos dados anotados no grafo, o `sdd2rdf` cria alguns exemplos de consultas SPARQL<sup>6</sup>. A anotação formaliza um vocabulário sobre os dados e abre caminho para a sua interoperabilidade e integração por diferentes fontes. De acordo com Rashid et al. (2017), após escolher quais dados do *dataset* anotar, os artefatos a seguir devem ser utilizados:

- Ontologia de domínio. A ontologia formaliza os conceitos do problema de pesquisa. Deve-se buscar reutilizar ontologias consolidadas no domínio do problema.
- *Dictionary Mapping* (DM). Anota a semântica das colunas do *dataset*. Cada linha do DM mapeia uma coluna do dataset, formalizando-a conceitualmente e explicitando suas relações com os outros dados do mesmo *dataset*, bem como a sua proveniência<sup>7</sup>.
- *CodeBook*. Um *codeBook* estrutura os dados categóricos<sup>8</sup> de um *dataset*, mapeando-os para conceitos correspondentes na ontologia. Dessa forma, o cientista da informação se preocupa com o tratamento dos dados, criando categorias e estabelecendo um código para cada uma. O *codeBook* possui os seguintes campos para anotação: Coluna (entidade a ser anotada), Código, Descrição e a Classe da Ontologia.
- *Infosheet*. Organiza os metadados de descrição do SDD; importante principalmente para o seu compartilhamento em redes, conforme o princípio FAIR de “encontrabilidade”.
- Grafo de Conhecimento (RDF). Resulta da interpretação da dupla "SDD (*templates* de metadados) + Dados" pelo *script* `sdd2rdf`, gerando o grafo RDF. Caso seja necessário persistir os dados, o usuário pode armazenar o grafo em um *triplestore*<sup>9</sup> para consulta posterior.

Para iniciar o uso do SDD, os dados mapeados para as ontologias são as colunas do *dataset*. Os objetos caracterizados nos *datasets* podem estar implicitamente representados. Os objetos implícitos serão explicitados no SDD e formalizados no grafo final gerado. A explicitação dos objetos implícitos favorece a integração semântica dos dados nos níveis conceituais mais abstratos do projeto, permitindo alinhar, homogeneizar e harmonizar interpretações de conceitos que descrevem aqueles dados a serem integrados.

<sup>5</sup><https://github.com/tetherless-world/SemanticDataDictionary>

<sup>6</sup>SPARQL Protocol and RDF Query Language – Linguagem de consulta elaborada pelo W3C para acesso a dados em formato RDF

<sup>7</sup>Como forma de anotar a proveniência do dado, o DM mapeia as entidades preexistentes que são relevantes na anotação dos dados por meio do campo “wasDerivedFrom”. Já o campo “wasGeneratedBy” descreve a atividade de geração associada à anotação de dados no DM (Rashid & McGuinness, 2020).

<sup>8</sup>Dados categóricos são dados agrupados. Podem derivar de observações feitas de dados qualitativos, ou de observações de dados quantitativos agrupados em determinados intervalos (Agresti, 2003).

<sup>9</sup>O armazenamento em *triplestore* é um banco de dados com o propósito de armazenar e recuperar triplas por meio de consultas semânticas (DBPEDIA, 2020).

Collum	Entity	Relation	sio:InRelationTo
??kpitmi	kpiOnto:Indicator	kpiOnto:isUsedBy	??hospital
??hospital	sio:Organazation	kpiOnto:hasKpi	??kpitmi

Tabela 3. Especificação do DM para objetos implícitos

### Anotação semântica de dados para gerar KPIs

Nesta seção, descreve-se o exemplo de anotação de dados para geração de KPIs no contexto da gestão hospitalar. O modelo da Figura 1 foi utilizado como esquema de dados para recuperação dos *datasets* e dados fictícios foram inseridos nas suas tabelas: FTempoMedioInternacao, DCID, DFaixaEtaria, Dtempo, DMunicipioOrigem, onde o prefixo “D” é a designação de tabela de dimensão e o prefixo “F” refere-se à tabela fato para análise de indicadores. O Sistema Gerenciador de Banco de Dados PostgreSQL foi usado para persistir os *datasets*.

A geração do dataset a ser anotado requereu o relacionamento das tabelas de dados, permitindo a criação de uma visão de dados contendo como resultado o dataset representado na Tabela 1. Nota-se a existência das colunas e dos dados a serem utilizados na anotação.

ID_KPI	CID	Faixa Etária	Município de Origem	Tempo por Mês	Tempo por Ano	Tempo Médio de Internação
1	F60-F69	ADULTO	BELO HORIZONTE	MAIO	2021	21,45
2	F30-F39	IDOSO	RIO DE JANEIRO	JUNHO	2021	20,38

Tabela 1. Dataset a ser anotado.

Além da recuperação do *dataset*, é importante ressaltar que os elementos *Dictionary Mapping*, *CodeBook* e *Infosheet* (utilizados em um SDD) são definidos por meio de arquivos em formato CSV (*Comma Separated Values*). A seguir é descrita a execução do processo de anotação:

- Ontologia de Domínio. As ontologias utilizadas foram a KPIOnto e a SIO. A KPIOnto possui conceitos consensuados que descrevem KPIs, enquanto a SIO é a ontologia padrão utilizada nos SDDs.
- Dictionary Mapping (DM). O DM (Tabelas 2 e tab:tabela3) mapeia para as ontologias de domínio (SIO e KPIOnto) as seguintes propriedades dos KPIs: dimensões TempoPorMes (tempo na dimensão mês), TempoPorAno (tempo na dimensão ano), CID (grupo de CID), FaixaEtaria e MunicipioDeOrigem; função de agregação de TempoMedioDeInternacao. Na Tabela 2, é possível identificar conceitos implícitos sobre o domínio, em que os dados mapeados são de um “hospital”. Os dados implícitos são anotados no artefato DM do SDD para que sejam também enriquecidos semanticamente. Com isso, novos dados anotados aparecem, servindo de pontes para representar mais amplamente o conhecimento. É uma preparação para considerar novos dados na análise explicitando relacionamentos que até o momento estavam implícitos.

Collum	Attribute	sio:AttributeOf	rdfs:Label
Id_Kpi	sio:Identifier	??kpitmi	Identificador do KPI
TempoPorMes	kpiOnto:hasDimension	??kpitmi	Descrição do Mês
TempoPorAno	kpiOnto:hasDimension	??kpitmi	Número do Ano
CID	kpiOnto:hasDimension	??kpitmi	Classificação Internacional de Doenças
FaixaEtaria	kpiOnto:hasDimension	??kpitmi	Faixa Etária
MunicipioDeOrigem	kpiOnto:hasDimension	??kpitmi	Município de Origem do Paciente
TempoMedioDeInternacao	kpiOnto:hasAggFunction	??kpitmi	Tempo Médio de Internação

Tabela 2. Especificação do DM para objetos explícitos.

- Infosheet. A Tabela 5 possui os metadados e seus vocabulários que descrevem o SDD a fim de melhorar a localização dos dados na web, considerado um princípio FAIR:

Collum	Code	Label	Class
DFaixaEtaria	1	IDOSO	kpionto:faixaetaria
DFaixaEtaria	2	ADULTO	kpionto:faixaetaria
DFaixaEtaria	3	ADOLESCENTE	kpionto:faixaetaria
DCID	F00-F09	F00-F09	kpionto:cid
DCID	F30-F39	F30-F39	kpionto:cid
DTempoMes	1	Janeiro	kpionto:tempomes
DTempoMes	2	Fevereiro	kpionto:tempomes
DTempoAno	2021	2021	kpionto:tempoano
DTempoAno	2022	2022	kpionto:tempoano

**Tabela 4.** Codebook - dimensões DTempo, DFaixaEtaria e DCID e DMunicipioDeOrigem

- *Codebook.* A Tabela 4 traz o *Codebook*, que descreve os dados categóricos do *dataset*, mapeando-os para a ontologia KPIOnto. São mapeadas as dimensões DCID, DFaixaEtaria, Dtempo, DMunicipioOrigem e a função de agregação TempoMedioDeInternacao.
- dct:creator: responsável pelo preenchimento;
- dct:contributor: contribuidores na criação do Infosheet e execução do processo;
- dct:created: data de criação;
- dct:description: propósito do SDD;
- owl:imports: endereço das ontologias utilizadas no SDD;
- schema:keywords: palavras-chave;
- dct:publisher: responsável por publicar;
- dct:title: título do SDD.

Atributo	Valor
dct:creator	José da Silva e Maria Silveira
dct:contributor	Maria Silveira
dct:created	30/09/2022
dct:description	Anotação semântica do dicionário de dados para geração do KPI na área hospitalar
owl:imports	http://semanticscience.org/ontology/sio-subset-labels.owl
schema:keywords	KPI, Gestão hospitalar
dct:publisher	José da Silva
dct:title	Geração de KPIs com base na anotação semântica de modelos de dados dimensionais

**Tabela 5.** Especificação do Infosheet

- Grafo de Conhecimento. O grafo RDF, representando o conhecimento sobre os KPIs, é persistido no Virtuoso<sup>10</sup>, onde é possível manipulá-lo com a linguagem SPARQL (Erling & Mikhailov, 2009). Na sequência, é apresentado um trecho do grafo RDF em sintaxe TTL<sup>11</sup>, representando a anotação e integração semântica dos dados das linhas da tabela de dados da Tabela 1. A primeira linha do *dataset* tem seus dados anotados pelos metadados do DM usando ontologias (Tabelas 2 e 3).

```

Id_Kpi a owl:Class ;
  rdfs:label "Identificador do KPI"^^xsd:string ;
  rdfs:subClassOf [ a owl:Restriction ;
    owl:allValuesFrom <http://hadatac.org/ont/kpionto#kpitmi> ;
    owl:onProperty sio:isAttributeOf ],
  sio:Identifier ;
  prov:generatedAtTime "2021-05-22T03:12:39+00:00"^^xsd:dateTime .

CID a kpiOnto:hasDimension, kpionto#CID ;
  rdfs:label "Classificacao Internacional de Doencas"^^xsd:string ;

```

<sup>10</sup><https://virtuoso.openlinksw.com/>

<sup>11</sup><https://www.w3.org/TR/turtle/>

```

sio:hasValue "F60-F69"^^xsd:string ;
sio:isAttributeOf <http://hadatac.org/ont/kpionto#kpitmi> ;
prov:generatedAtTime "2021-05-22T03:12:40+00:00"^^xsd:dateTime .

FaixaEtaria a kpiOnto:hasDimension, kpionto#FaixaEtaria ;
rdfs:label "Faixa Etaria"^^xsd:string ;
sio:hasValue "ADULTO"^^xsd:string ;
sio:isAttributeOf <http://hadatac.org/ont/kpionto#kpitmi> ;
prov:generatedAtTime "2021-05-22T03:12:40+00:00"^^xsd:dateTime .

MunicipioDeOrigem a kpiOnto:hasDimension, kpionto#MunicipioDeOrigem ;
rdfs:label "Municipio De Origem"^^xsd:string ;
sio:hasValue "BELO HORIZONTE"^^xsd:string ;
sio:isAttributeOf <http://hadatac.org/ont/kpionto#kpitmi> ;
prov:generatedAtTime "2021-05-22T03:12:40+00:00"^^xsd:dateTime .

TempoPorMes a kpiOnto:hasDimension, kpionto#TempoPorMes ;
rdfs:label "Descricao do Mes"^^xsd:string ;
sio:hasValue "MAIO"^^xsd:string ;
sio:isAttributeOf <http://hadatac.org/ont/kpionto#kpitmi> ;
prov:generatedAtTime "2021-05-22T03:12:39+00:00"^^xsd:dateTime .

TempoPorAno a kpiOnto:hasDimension, kpionto#TempoPorAno;
rdfs:label "Numero do Ano"^^xsd:string ;
sio:hasValue 2021 ;
sio:isAttributeOf <http://hadatac.org/ont/kpionto#kpitmi > ;
prov:generatedAtTime "2021-05-22T03:12:39+00:00"^^xsd:dateTime .

TempoMedioDeInternacao a kpiOnto:hasAggFunction, kpionto#TempoMedioDeInternacao> ;
rdfs:label "Tempo Medio de Internacao"^^xsd:string ;
sio:hasValue "21.45"^^xsd:float ;
sio:isAttributeOf <http://hadatac.org/ont/kpionto#kpitmi> ;
prov:generatedAtTime "2021-05-22T03:12:40+00:00"^^xsd:dateTime .
    
```

- Visualização dos dados. Um painel em formato de dashboard foi construído em MS-PowerBI<sup>12</sup> para conectar-se ao Virtuoso por meio do componente ODBC do Windows e executar consultas SPARQL, ilustrando como os dados, manipulados a partir do grafo, podem ser visualizados de diferentes formas. Uma consulta SPARQL é utilizada para extrair os dados do grafo de conhecimento. O MS-PowerBI é usado para carregar os dados em formato RDF e permitir a sua transformação para dados tabulares, a fim de gerar os indicadores e métricas.

Os indicadores gerados pela aplicação de exemplo, como o TMI por município, podem ser visualizados na Figura 3.

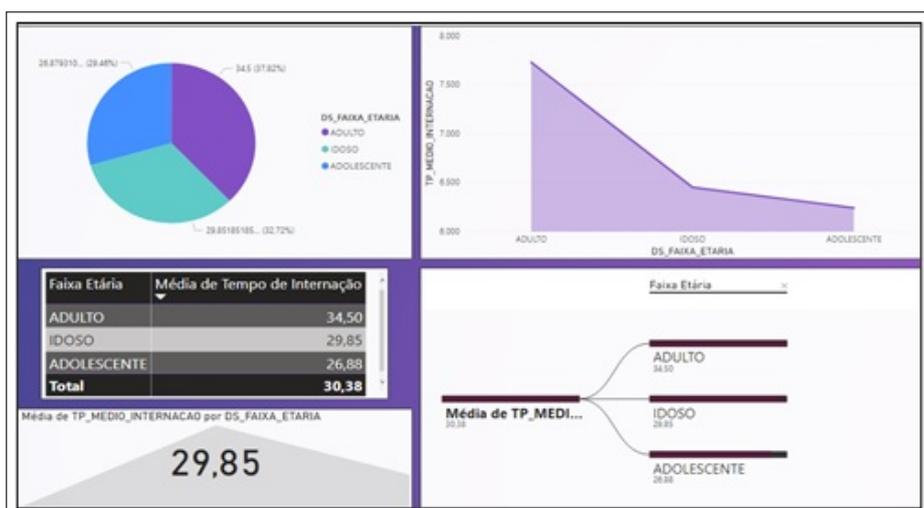


Figura 3. Modelo dimensional de dados para o indicador TMI

Para ilustrar outra possibilidade de visualização dos dados armazenados no formato RDF, a Figura 4 apresenta o TMI por faixa etária do paciente.

<sup>12</sup><https://powerbi.microsoft.com/pt-br/>



Figura 4. Modelo dimensional de dados para o indicador TMI

## CONSIDERAÇÕES FINAIS

No contexto organizacional, a modelagem conceitual adequada dos dados envolve a interpretação e negociação de significados sobre entidades, relacionamentos e regras de negócios, que ocorrem naturalmente na comunicação entre os vários atores (ou partes interessadas). As vantagens do SDD atingem pleno potencial em cenários onde a integração de fontes de dados diversas (internas ou externas) se faz necessária para enriquecer os dados que se quer analisar. O processo apresentado visa organizar etapas para anotação com SDDs e geração do grafo de conhecimento em RDF, representando formalmente o conjunto de fatos originados da combinação de dados de diferentes fontes. Um exemplo de geração de KPI, usando como fonte um modelo dimensional, para avaliar critérios de desempenho em função do tempo médio de internação produzido por hospitais e com base nas regras da Agência Nacional de Saúde, ilustrou o processo, constituindo uma validação preliminar do método.

Conforme visto, os SDDs podem contribuir para organizar e integrar dados oriundos de diferentes fontes de um hospital, gerando informações que estruturam conhecimentos sobre diversos indicadores. Isso facilita os alinhamentos semânticos sobre os KPIs a partir de uma abordagem de modelagem de dados ampla, considerando-se ainda a importância de existir uma modelagem conceitual que trate os dados em uma abordagem *top-down* e não apenas *bottom up*. O processo proposto permite associar dados de datasets a conceitos consensuados a fim de gerar KPIs, enriquecendo-os e formalizando-os com ontologias.

A abordagem descrita pode contribuir para integrar fontes de dados heterogêneas, além de abrir caminho para a comunidade da KPIOnto na utilização do SDD para anotação semântica dos dados sobre KPIs. Pode haver necessidade da rede hospitalar trocar informações sobre KPIs, como forma da população e do poder público acompanharem a qualidade do atendimento à saúde. Os projetos de SDDs podem permitir o alinhamento conceitual de estados informacionais entre os hospitais. Nesse contexto, o governo brasileiro já considera a dimensão semântica no desenvolvimento, manutenção de ontologias e outros recursos de organização da informação, visando melhorar a interoperabilidade e a troca de informações, por meio do e-Ping - Padrões de Interoperabilidade de Governo Eletrônico (Comitê Executivo de Governo Eletrônico, 2018). Contribuiu-se com a curadoria dos dados, já que SDDs seguem boas práticas de modelagem (princípios FAIR). Finalmente, é necessária a proteção e privacidade das informações gerenciais integradas, em que tecnologias como *Blockchain* e *Smart Contracts* têm potencial de contribuir positivamente para o controle das informações gerenciais hospitalares (Xavier & Duque, 2021).

Estudos futuros pretendem investigar como a modelagem por meio SDD, tal qual a apresentada neste trabalho, podem constituir alternativa vantajosa à modelagem dimensional clássica (do tipo "*data smart*" ou "*data warehouse*"). A flexibilidade de modelos conceituais ontológicos "livres de esquemas" (*schema free*) é uma hipótese que se apresenta vantajosa para a geração de KPIs no contexto analisado. A abordagem apresentada também pode permitir a evolução do conhecimento sobre os indicadores hospitalares de forma mais organizada, flexível, incremental e semanticamente enriquecido pela explicitação de sua semântica formal, advinda do uso de ontologias representadas em Lógica de Descrições (*Description Logic*) (Krotzsch, Simancik, & Horrocks, 2012).

## REFERÊNCIAS

- Agência Nacional de Saúde Suplementar. (2020). *Qualidade hospitalar: Ans divulga as fichas dos indicadores*. Recuperado de <http://www.ans.gov.br/aans/noticias-ans/qualidade-da-saude/5478-qualidade-hospitalar-ans-divulga-as-fichas-dos-indicadores>
- Agresti, A. (2003). *Categorical data analysis*. New York: John Wiley & Sons.
- Bax, M. P., & Silva, E. d. O. (2019, nov. 27-29). *Anotação de dados para geração de indicadores de desempenho em organizações*. Brasília. Recuperado de [http://widat2019.fci.unb.br/images/Anais\\_WIDaT\\_2019.pdf#page=21](http://widat2019.fci.unb.br/images/Anais_WIDaT_2019.pdf#page=21)
- Bax, M. P., & Silva, E. d. O. (2020). *Dicionários semânticos de dados para integrar dados de prontuários eletrônicos de pacientes*. Recuperado de <http://ceur-ws.org/Vol-2728/doctorate1.pdf>
- Comitê Executivo de Governo Eletrônico. (2018). *e-ping: padrões de interoperabilidade de governo eletrônico*. Recuperado de <http://smeduquedecaxias.rj.gov.br/nead/Biblioteca/Forma%C3%A7%C3%A3o%20Continuada/Tecnologia/Leis/e-PING%20v2011%2003%2012%202010.pdf>
- Dang, J., Hedayati, A., Hampel, K., & Toklu, C. (2008). An ontological knowledge framework for adaptive medical workflow. *Journal of Biomedical Informatics*, 41(5), 829–836. doi: 10.1016/j.jbi.2008.05.012
- Davrieux, C. F., Palermo, M., Serra, E., & Giménez, M. E. (2019). Etapas e fatores do “processo perioperatório”: pontos em comum com a indústria aeronáutica. *Arquivos Brasileiros de Cirurgia Digestiva*, 32(1). doi: 10.1590/0102-672020180001e1423
- DBPEDIA. (2020). *About: triplestore*. Recuperado de <http://dbpedia.org/page/Triplestore>
- Diamantini, C., Potena, D., & Storti, E. (2016). Sempi: a semantic framework for the collaborative construction and maintenance of a shared dictionary of performance indicators. *Future Generation Computer Systems*, 54, 352–365. doi: 10.1016/j.future.2015.04.011
- Erling, O., & Mikhailov, I. (2009). Rdf support in the virtuous dbms. In P. T. P., A. Sören, T. Klaus, & S. S. (Eds.) (Eds.), *Networked knowledge-networked media: Integrating knowledge management*. doi: 10.1007/978-3-642-02184-8\_25
- Hogan, A., Blomqvist, E., Cochez, M., & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys (CSUR)*, 54(4), 1–37. doi: 10.1145/3447772
- Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: the definitive guide to dimensional modeling*. New York: John Wiley & Sons.
- Krotzsch, M., Simancik, F., & Horrocks, I. (2012). *A description logic primer*. doi: 10.48550/arXiv.1201.4089
- Medeiros, C. B. (2018). *Gestão de dados científicos: da coleta à preservação*. Recuperado de <https://blog.scielo.org/blog/2018/06/22/gestao-de-dados-cientificos-da-coleta-a-preservacao/#.ZBsMx3bMKUk>
- Pan, J. Z. (2017). *Exploiting linked data and knowledge graphs in large organizations*. Heidelberg: Springer.
- Parmenter, D. (2015). *Key performance indicators: developing, implementing, and using winning kpis*. New York: John Wiley & Sons.
- Pereira, R. (2021). Gestão do conhecimento aliada ao crescimento organizacional: perspectivas à prática hospitalar. *AtoZ: novas práticas em informação e conhecimento*, 10(1), 103–112. doi: 10.5380/atoz.v10i1.75082
- Protegé. (2020). *Webprotegé*. Recuperado de <https://protege.stanford.edu/about.php>
- Rashid, S. M., McCusker, J. P., Pinheiro, P., Bax, M. P., Santos, H. O., Stingone, J. A., ... McGuinness, D. (2017). The semantic data dictionary approach to data annotation & integration. *SemSci ISWC*, 47–54. Recuperado de [https://doi.org/10.1162/dint\\_a\\_00058](https://doi.org/10.1162/dint_a_00058)
- Rashid, S. M., & McGuinness, D. L. (2020). The semantic data dictionary: an approach for describing and annotating data. *Data Intelligence*, 2(4), 443–486. Recuperado de <https://direct.mit.edu/dint/article-abstract/2/4/443/94892>
- Roldán-García, M. d. M., García-Niéto, Maté, A. J., & Aldana-Montes, J. F. (2019). Ontology-driven approach for kpi meta-modelling, selection and reasoning. *International Journal of Information Management*, 58. doi: 10.1016/j.ijinfomgt.2019.10.003
- Tsai, T. C., Jha, A. K., Gawande, A. A., Huckman, R. S., Bloom, N., & Sadun, R. (2015). Hospital board and management practices are strongly related to hospital performance on clinical quality metrics. *Health affairs*, 34(8), 1304–1311. doi: 10.1377/hlthaff.2014.1282
- Wetzstein, B., Ma, Z., & Leymann, F. (2008). Towards measuring key performance indicators of semantic business processes. In *International conference on database theory*. Recuperado de [https://link.springer.com/chapter/10.1007/978-3-540-79396-0\\_20](https://link.springer.com/chapter/10.1007/978-3-540-79396-0_20)
- Xavier, A. C. C., & Duque, C. G. (2021). Prontuário eletrônico do paciente: qual a contribuição da arquivística e do smart contracts para a sua gestão na era da saúde 4.0? *AtoZ: novas práticas em informação e conhecimento*, 10(3), 01–10. doi: 10.5380/atoz.v10i3.81267
- Zaidan, F. H., & Bax, M. P. (2013). Linked open data como forma de agregar valor às informações clínicas. *AtoZ: novas práticas em informação e conhecimento*, 2(1), 44–59. doi: 10.5380/atoz.v2i1.41319

---

Como citar este artigo (APA):

Silva, E. de O. da., Pereira, F. C. M., Marques, Y. B., & Melo, E. C. (2023). Dicionário semântico de dados para anotação de indicadores de desempenho para gestão hospitalar. *AtoZ: novas práticas em informação e conhecimento*, 12, 1 – 11. Recuperado de: <http://dx.doi.org/10.5380/atoz.v12.88109>

## NOTAS DA OBRA E CONFORMIDADE COM A CIÊNCIA ABERTA

### CONTRIBUIÇÃO DE AUTORIA

Papéis e contribuições	Evaldo de Oliveira da Silva	Marcello Peixoto Bax	Frederico César Mafra Pereira	Yuri Bento Marques	Eduardo Cardoso Melo
Concepção do manuscrito	X	X	X		
Escrita do manuscrito	X	X	X	X	X
Metodologia	X	X		X	
Curadoria dos dados			X	X	X
Discussão dos resultados	X	X	X	X	X
Análise dos dados	X	X		X	

### Disponibilidade de Dados Científicos da Pesquisa

Os conteúdos subjacentes ao texto da pesquisa estão contidos no manuscrito.

### EQUIPE EDITORIAL

#### Editora/Editor Chefe

Paula Carina de Araújo (<https://orcid.org/0000-0003-4608-752X>)

#### Editora/Editor Associada/Associado

Helza Ricarte Lanz (<https://orcid.org/0000-0002-6739-2868>)

#### Editora/Editor de Texto Responsável

Cristiane Sinimbu Sanchez (<https://orcid.org/0000-0002-0247-3579>)

Seção de Apoio às Publicações Científicas Periódicas - Sistema de Bibliotecas (SiBi) da Universidade Federal do Paraná - UFPR

#### Editora/Editor de Layout

Karolayne Costa Rodrigues de Lima (<https://orcid.org/0000-0002-6311-8482>)