

# ONTOVID - UMA ABORDAGEM PARA CONSTRUÇÃO DE GRAFOS DE CONHECIMENTO SEMÂNTICO COM ENFOQUE EM NOTIFICAÇÕES E ÓBITOS RELACIONADOS AO NOVO CORONAVÍRUS (COVID-19)

Nacles B. P. Gomes<sup>1</sup>, Sandro de C. Franco<sup>1</sup>, Laís do N. Salvador<sup>1</sup>

<sup>1</sup>Universidade Federal da Bahia (UFBA) – Ba – Brazil

**Abstract.** *Brazilian municipalities face many difficulties in obtaining strategic information that enables decision-making due to the decentralized management model and the lack of interoperable systems that allow complex analyses. With the novel coronavirus (COVID-19) pandemic, obtaining data quickly and reliably has become even more necessary due to the escalation of the disease and the need for rapid interventions. Therefore, this work, through the construction of Semantic Knowledge Graphs (SGG), which integrated the SIM databases (MIS - Mortality Information System) and the e-SUS System (COVID Notifications), had as its main objective to mitigate the deficiencies arising from the extraction of important indicators on deaths, vaccinations, and hospitalizations. For the construction of ONTOVID, the NeOn methodology and the OBDI (Ontology-Based Data Integration) approach were used to integrate the bases. As a way of evaluating the accuracy of this automatic approach, the data obtained in the integration were presented and validated by the managers of the Camaçari Health Department, who identified some divergences not detected by the manual process of collecting information previously carried out by the team.*

**Keywords:** *Semantic Knowledge Graphs (SCM), Integration of Heterogeneous Data Sources, Ontologies.*

**Resumo.** *Muitas são as dificuldades enfrentadas pelos municípios brasileiros na obtenção de informações estratégicas que possibilitem a tomada de decisões devido ao modelo de gestão descentralizado e a falta de sistemas interoperáveis que facilitem a realização de análises complexas. Com a pandemia do novo coronavírus (COVID-19), a obtenção de dados de forma rápida e confiável tornou-se ainda mais necessária em função da escalada da doença e da necessidade de rápidas intervenções. Portanto, esse trabalho, mediante a construção de Grafos de Conhecimento Semântico (GCS), a partir da integração das bases de dados do SIM (Sistema de Informações sobre Mortalidade) e do e-SUS (Notifica), teve como objetivo principal mitigar as deficiências decorrentes da extração de indicadores importantes sobre os óbitos, vacinados e internações. No processo de construção da ONTOVID foram utilizadas a metodologia NeOn e a abordagem OBDI (Ontology-Based Data Integration). Como forma de avaliar a acurácia desta abordagem automática, os dados obtidos na integração foram apresentados e validados pelos gestores da Secretaria de Saúde de Camaçari, que identificaram algumas divergências não detectadas pelo processo manual de coleta de informações executado anteriormente pela equipe.*

**Palavras-chave:** *Grafos de Conhecimento Semântico (GCS), Integração de Fontes de dados Heterogêneas, Ontologias.*

## 1. Introdução

A evolução e popularização das Tecnologias de Informação e Comunicação (TIC), nas últimas décadas, têm promovido um aumento na produção e disponibilidade de dados. Este cenário apontou para uma necessidade recorrente de interoperar sistemas ou serviços com o intuito de manipular dados de forma transparente e distinta de suas fontes, tornando-se uma necessidade estratégica para diversos segmentos, inclusive para a saúde pública em suas diversas esferas governamentais. Nesse aspecto, o Sistema Único de Saúde Brasileiro (SUS), em função de seu modelo descentralizado de gestão e do grande número de sistemas de informação que possui, necessita com frequência operar uma grande quantidade de dados que permitam a realização de análises complexas para a tomada de decisão, o que é extremamente difícil em função da falta de interoperabilidade entre os sistemas existentes, sejam eles do Ministério da Saúde, próprios ou terceiros.

A necessidade de produzir informações na área de saúde pública tem impulsionado o desenvolvimento de novas soluções que facilitem o processo de integração de dados, que podem ser complexas e onerosas devido à heterogeneidade semântica ou à sua distribuição. Uma solução que tem se mostrado eficiente para o processo de integração, conforme, por exemplo, os trabalhos de [Gabriela et al. 2016], [Gabriel et al. 2017], [Pereira 2019], [Andrade et al. 2019] e [Ferronato et al. 2016], são abordagens que empregam ontologias. As ontologias têm sido utilizadas com sucesso no processo de integração de fontes de dados heterogêneas como linguagem padrão para representar metadados e por fornecer uma visão global que assume o papel de mediador nas consultas aos dados. Além disso, a camada ontológica pode evitar a desatualização durante as consultas se ocorrerem alterações nas fontes de dados. Sem a ontologia, aplicações que passam por um processo denominado ETL (Extract Transform Load) ou simplesmente extração, transformação e carga podem apresentar dados desatualizados [Calvanese et al. 2018].

Este trabalho, por sua vez, tem como objetivo principal, propor uma abordagem para minimizar as dificuldades recorrentes de gestores do Sistema Único de Saúde (SUS) na obtenção de dados para análises sobre o COVID-19. Os Grafos de Conhecimento Semânticos (GCS) aqui propostos, mediante a integração das fontes de dados do e-SUS Notifica (Sistema de Monitoramento e Notificações do COVID-19) e do SIM (Sistema de Informação de Mortalidade), fornecerão visões ampliadas sobre as relações entre os casos graves, óbitos, ciclo vacinal e de contaminação. Ademais, os mapeamentos semânticos poderão proporcionar ainda um vislumbre sobre diversos aspectos que envolvem a letalidade da doença. Para tanto, foi utilizado um recorte temporal das bases de dados supracitadas, as quais foram obtidas junto ao município de Camaçari-Ba<sup>1</sup> para realização do estudo. Para a construção da ONTOVID, empregamos a metodologia NeOn [Suarez–Figuerola et al. 2008a] em conjunto com a abordagem OBDI (*Ontology-Based Data Integration*) para a integração das fontes de dados do e-SUS Notifica e do SIM na geração dos grafos a cerca das Notificações, Vacinação e Mortalidade relacionados ao COVID-19, visando demonstrar a eficiência do modelo para a obtenção de indicadores e relatórios.

Este trabalho é composto por quatro seções. A segunda seção apresenta uma fundamentação teórica sobre Grafos de Conhecimento Semântico e Ontologias. Em se-

---

<sup>1</sup>Bases de dados obtidas através de Carta de Anuência celebrada entre Secretaria de Saúde de Camaçari e a Universidade Federal da Bahia - UFBA

guida, na terceira seção, retratamos a abordagem utilizada para a construção da ONTOVID, detalhando a metodologia empregada. Por fim, na quarta seção, são apresentadas as conclusões e trabalhos futuros seguidas pelas referências bibliográficas.

## 2. Grafos de Conhecimento e Ontologias

Os Grafos de Conhecimento Semântico (GCS) são reconhecidos como uma abordagem eficaz no processo de integração semântica de dados heterogêneos, apesar em de ser considerados como um paradigma relativamente novo. O GCS tem sido amplamente utilizado como elemento essencial para mapear, organizar e produzir informações. Neste contexto, as ontologias desempenham um papel essencial, sendo utilizadas como um vocabulário unificado, que permite a combinação e enriquecimento de informações para a realização de consultas complexas e inferências sobre os dados integrados [Calvanese et al. 2018].

Dentre as múltiplas tarefas envolvidas no processo de integração de dados, uma das mais importantes é a compreensão dos conceitos e relações que existem por trás dos registros, o que é fundamental para a construção dos cenários que se deseja representar ou obter informações. Essa atividade é desafiadora porque a semântica é um elemento implícito em nomes de variáveis ou tabelas cujos modelos de dados podem ser diversos, com propriedades e terminologia heterogêneas. Mesmo que o acesso aos dados e a barreira do seu contexto sejam superados, estabelecer relações significativas entre eles é uma atividade complexa [Cheatham and Pesquita 2017].

Na integração de dados, o GCS é utilizado como uma camada que fornece uma visão global que media consultas a diferentes fontes, tornando-se uma camada de representação do conhecimento, abstraindo totalmente o armazenamento de dados, independentemente de sua organização. O usuário pode acessar um vocabulário feito para o domínio representado, facilitando o processo de recuperação da informação. A utilização de GCS podem ser vistos em importantes pesquisas sobre integração de fontes de dados heterogêneas. Em [Lopes et al. 2017], por exemplo, é apresentada uma estrutura conceitual denominada LAIS, com o objetivo de criar um Sistema de Apoio à Saúde para facilitar o processo de criação de visões integradas das bases de dados do Sistema Único de Saúde (SUS). Em [Ciriaco et al. 2020], uma abordagem ontológica adaptada de [Ekaputra et al. 2017] é utilizada para integrar fontes de dados dos sistemas SIM e SINASC na cidade de São Paulo, cujo objetivo foi obter o indicador de saúde sobre Dias Potenciais Perdidos de Gravidez (DPGP).

Na construção dos GCS, as ontologias são importantes em várias fases do processo de integração de fontes de dados heterogêneas, seja na visão global ou no acesso aos dados. As soluções que empregam tecnologias para criar visualizações unificadas são chamadas de OBDI [Ekaputra et al. 2017]. Uma das principais características dos modelos que empregam OBDI é a independência conceitual dos dados, pois permite a adição de novas visualizações e fontes de dados. De acordo com [Ekaputra et al. 2017], a OBDI possui pelo menos quatro componentes considerados essenciais para o modelo: **A**: camada de fonte de dados heterogênea; **B**: camada de ontologia local representando o conteúdo das fontes de dados; **C**: camada de ontologias globais que são semanticamente amplas para representar dados de todas as fontes de dados a serem integradas e **D**: camada de aplicativos que acessam os dados integrados com OBDI.

Segundo [Ekaputra et al. 2017], arquiteturas que empregam a OBDI podem ser

dividas em quatro abordagens, que são:

- **Abordagem de ontologia única (1):** a ontologia global é mapeada acessando diretamente as fontes de dados, sem ontologias intermediárias.
- **Abordagem de ontologias múltiplas (2):** uma ontologia local é criada para cada fonte de dados, onde a integração entre elas ocorre mediante alinhamento ou mapeamentos semânticos.
- **Abordagem híbrida (3):** utiliza os mecanismos das abordagens (1) e (2), simplificando a adição de novas fontes de dados. Todavia, não emprega a técnica de mapeamento entre as ontologias locais. Em vez disso, utiliza um vocabulário compartilhado para ser usado e estendido dentro das ontologias.
- **Abordagem GAV (4):** nesta abordagem são criadas ontologias locais a partir das fontes de dados, como na abordagem híbrida, e uma ontologia global que pode ser acessada através de algum aplicativo.

Com base nos modelos apresentados da OBDI, que implementam abordagens para muitos dos problemas de uniformidade semântica como solução para questões relacionadas a interoperabilidade, este projeto, conforme pode ser observado na Figura 1, optou pela **abordagem de ontologia única** na solução para integração entre duas bases importantes do Sistema Unico de Saúde (SUS), a base sobre mortalidade (SIM) e a de notificações do COVID-19 (e-SUS Notifica), para análises sobre indicadores relacionados a pandemia do novo coronavírus (COVID-19).

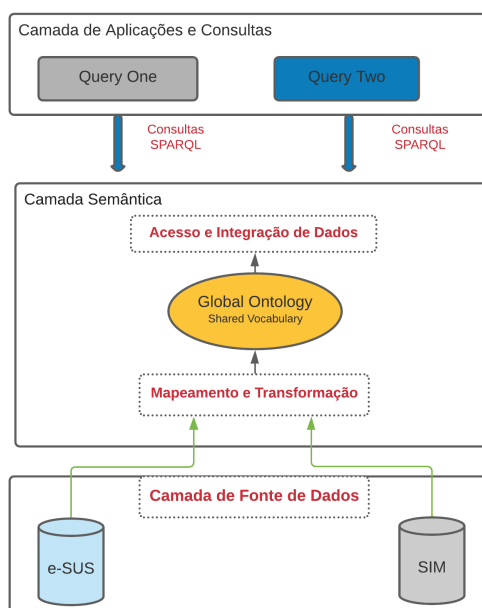


Figura 1. Modelo em 3 (três) camadas para construção dos Grafos de Conhecimento Semântico (GCS). Fonte: Autores.

### 3. Desenvolvimento da abordagem para a construção dos GCS (ONTOVID)

Na busca por uma metodologia que apoiasse o processo de construção do ONTOVID, identificamos a metodologia NeON [Suarez–Figuerola et al. 2008b] como a mais adequada devido ao conjunto de processos e atividades que disponibiliza. O Projeto NeON<sup>2</sup>

<sup>2</sup><http://neon-project.org/nw/Ontology.html>

tem a capacidade de permitir a gestão de múltiplas ontologias de forma colaborativa, dinâmica e evolutiva. Em vez de propor um único caminho para o desenvolvimento de uma ontologia, a NeOn propõe nove cenários que podem ser individuais ou combinados para cobrir problemas comuns que ocorrem na engenharia de ontologias, conforme descritos por [Suárez-Figueroa et al. 2012], que são:

Cenário 1: Construir ontologia nova, sem reutilizar os recursos de conhecimento existentes; Cenário 2: Reutilizando e refazendo a engenharia de recursos não ontológicos; Cenário 3: Reutilizando recursos ontológicos; Cenário 4: Reutilizando e fazendo reengenharia de recursos ontológicos; Cenário 5: Reutilizando e mesclando de recursos ontológicos; Cenário 6: Reutilizando, mesclando, e realizando reengenharia de recursos ontológicos; Cenário 7: Reutilizando padrões de projeto de ontologia; Cenário 8: Reestruturando recursos ontológicos; e Cenário 9: Localizando recursos ontológicos;

Para este projeto o Cenário 4 da Metodologia NeOn foi escolhido porque possibilita o reuso e reengenharia de recursos ontológicos, já que este trabalho estende a ontologia produzida e disponibilizada pelo portal SemanticSUS<sup>3</sup>, que foi desenvolvida inicialmente para a extração de indicadores sobre o Risco de Morte Neonatal (RMS), a partir das bases de dados dos sistemas SIM (Sistema de Informações sobre Mortalidade) e o SINASC (Sistema de Informações sobre Nascidos Vivos). Para o Cenário 4 a metodologia NeOn propõe quatro processos básicos, os quais foram adaptados para este trabalho, que são, conforme [Suárez-Figueroa et al. 2012]:

a) Re-especificação: processo de atualização da especificação ou engenharia reversa para formalização ou conceitualização do projeto, adaptando-as (especificações) as novas necessidades; b) Re-conceitualização: processo de engenharia reversa da implementação ou formalização. A ontologia será conceitualizada conforme as novas especificações do projeto, formalizada e implementada; c) Re-formalização: reimplementação da ontologia conforme a nova formalização; d) Re-implementação: a ontologia é modificada.

A partir da definição do Cenário da metodologia NeOn que seria utilizado, iniciamos o processo de construção da ONTOVID com o estudo de viabilidade e aquisição do conhecimento, que teve como objetivo principal identificar o problema para o qual o modelo de integração semântica seria construído. Assim, iniciamos o levantamento junto aos técnicos do município de Camaçari-Ba para avaliar quais informações deveriam e poderiam ser obtidas, além das fontes de dados que seriam utilizadas. Mediante essa avaliação, identificamos as fontes dos sistemas e-SUS Notifica e o Sistema de Informações sobre Mortalidade (SIM) como essenciais para obtenção dos dados sobre o perfil epidemiológico da pandemia por COVID-19.

O SIM é um sistema empregado na esfera municipal de saúde, com base de dados local, que permite o compartilhamento de informações sobre mortalidade com o Ministério da Saúde. Essa fonte de dados possibilita a obtenção de indicadores que contribuem para análises complexas sobre a mortalidade no Brasil, favorecendo uma gestão

---

<sup>3</sup><https://semanticsus.github.io/semanticSUS/index.html>

mais eficiente da saúde. O e-SUS Notifica, diferentemente do SIM, não possui base municipal, e foi desenvolvido pelo Ministério da Saúde durante a pandemia com objetivo de coletar dados das notificações de casos de COVID-19. Nesse aspecto, como o município não tem acesso direto a base de dados do e-SUS Notifica, foi utilizado um arquivo CSV, exportado como relatório, para o desenvolvimento desta abordagem.

Durante o levantamento inicial realizado no município de Camaçari (Bahia), foram identificadas necessidades de obtenção de indicadores importantes relacionados a cura e óbitos por COVID, óbitos por outras causas, vacinas aplicadas (contra COVID-19), óbitos em vacinados e taxa da Letalidade; que foram devidamente mapeadas como Questões de Competência (QC) que a ontologia deveria responder. Essas informações e indicadores não foram fáceis de obter, pois exigem a manipulação de diversas planilhas que são constantemente extraídas dos sistemas SIM e e-SUS Notifica. Nesse sentido, o processo de identificação dos casos ocorre por meio do mapeamento do NOME, DATA DE NASCIMENTO, NOME DA MÃE e/ou CPF e/ou CNS.

A partir da obtenção das fontes de dados e da definição da metodologia e abordagem que seriam empregadas, importamos a ontologia disponibilizada pelo portal SemanticSUS<sup>4</sup> e a estendemos para agregar novas classes e propriedades, conforme poderá ser observado na seção (Re)Conceituação, (Re)Formalização e (Re)Implementação. Essas alterações foram necessárias para viabilizar a integração das fontes de dados do SIM e do e-SUS Notifica, bem como a de outras fontes para análises futuras, como a do CNES (Cadastro Nacional de Estabelecimentos de Saúde) e do CBO (Cadastro Brasileiro de Ocupação). Vale observar que a ontologia disponibilizada pelo SemanticSUS não contemplava conceitos sobre Notificações, Teste de COVID-19, Vacinas, Estabelecimentos e Profissionais de Saúde, os quais foram agregados durante o desenvolvimento da abordagem ONTOVID. Para além, os conceitos relacionados ao SINASC, parte da ontologia original do SemanticSUS, foram descartados por não serem objeto deste estudo.

### **3.1. (Re)Especificação**

As três primeiras tarefas realizadas nesta fase do processo de especificação ou re-especificação - evoluindo o SemanticSUS para ONTOVID -, foram: (1) Identificar o propósito, escopo e nível de formalidade; (2) Identificar os usuários e (3) Identificar os usos da ontologia. Em relação ao Propósito, escopo e nível de formalidade, identificamos que além de promover a uniformidade semântica necessária para integração das fontes de dados dos sistemas de notificação e mortalidade do Ministério da Saúde, e-SUS Notifica e SIM, respectivamente, com o objetivo de permitir a extração de indicadores essenciais a gestão municipal de saúde; este projeto também poderá facilitar a interoperabilidade com outros sistemas mediante a integração de conceitos.

Em relação a granularidade da ontologia, ela está relacionada diretamente às Questões de Competência (QC) apresentadas na Tabela 1, conforme a especificação de requisitos realizada a partir dos cenários-problema levantados. É importante ressaltar que as QC apresentadas aqui não abrangem todas as questões levantadas pelos gestores do município de Camaçari, mas representam apenas um subconjunto destas (as principais), como forma de demonstração da viabilidade e funcionalidade da abordagem proposta.

---

<sup>4</sup><https://semanticsus.github.io/semanticSUS/index.html>

Tabela 1: Questões de Competência da ONTOVID obtidas na SESAU de Camaçari. (01/01/2021 até 31/12/2021)).

Questões de Competência
QC 01: Quantos óbitos ocorreram em função do COVID-19?
QC 02: Quantas pessoas se recuperaram plenamente do COVID-19?
QC 03: Quantas pessoas vacinadas pegaram COVID-19 (confirmados em laboratório)?
QC 04: Quantos óbitos ocorreram, por sexo?
QC 05: Quantos pessoas vacinados adquiriram COVID-19?
QC 06: Quantas pessoas vacinadas morreram por complicações do COVID-19?
QC 07: Qual é a taxa de letalidade do COVID-19?

### 3.2. (Re)Conceitualização, (Re)Formalização e (Re)Implementação

Nesta fase, foi avaliada a reutilização de recursos ontológicos existentes, principalmente aqueles que possuem um alto nível de granularidade e qualidade no desenvolvimento. A classificação da qualificação por tipo e nível de granularidade foi especificado por [Suárez-Figueroa et al. 2012], como: i) Gerais/Comuns: quando trata de assuntos diversos ou comuns; ii) Domínio: quando estabelecem conceitos de questões ou contextos específicos. Conforme estabelecido pela metodologia NeOn, algumas tarefas podem ser agrupadas em atendimento a esta fase. Assim, procedemos com as seguintes ações: i) Análise de ontologias gerais relacionadas ao tema/estudo; ii) escolha de uma ou mais ontologias; e iii) Personalização e integração da(s) ontologia(s). Com base nas atividades alçadas, estendemos e personalizamos o GCS disponibilizado pelo SemanticSUS para atender as necessidades aplicadas as novas questões de competência sobre o COVID-19.

No processo de extensão da ontologia SemanticSUS, procedemos com o levantamento e análise das classes, relações entre classes, axiomas e regras; validando tanto a documentação do SemanticSUS quanto os termos utilizados nos sistemas SIM e e-SUS Notifica para a (Re)Formalização e (Re)Implementação, conforme pode ser observado na Figura 2. As classes incorporadas criaram novas relações e regras, dentre as principais classes, temos: a) Notificação: representa o registro dos dados relacionados ao paciente com possível indicação de COVID-19. Possui propriedades que indicam a classificação, critérios de avaliação, evolução do quadro, vacinação e testes realizados em paciente; b) Estabelecimento de Saúde: permite identificar local de testes, internamentos e tratamentos realizados; c) Profissional de Saúde: vínculo com o estabelecimento de saúde, profissão e atividades desenvolvidas. Também permite observar se o óbito de um profissional de saúde está relacionado ao COVID-19; d) Vacina: identifica o tipo de vacina, lote e data de aplicação da vacina em pacientes que foram registrados (notificados) com COVID-19; e e) Teste COVID-19: relaciona os tipos de teste, data e resultados de cada paciente notificado, permitindo análises sobre óbitos em vacinados.

Nesta fase também foram realizadas as validação das descrições semiformais, das propriedades, das restrições e do mapeamento das regras. Assim, concluímos as atividades inerentes ao processo de conceitualização e formalização. É importante destacar que esta fase ocorreu em paralelo com a fase de implementação, fornecendo um modelo em

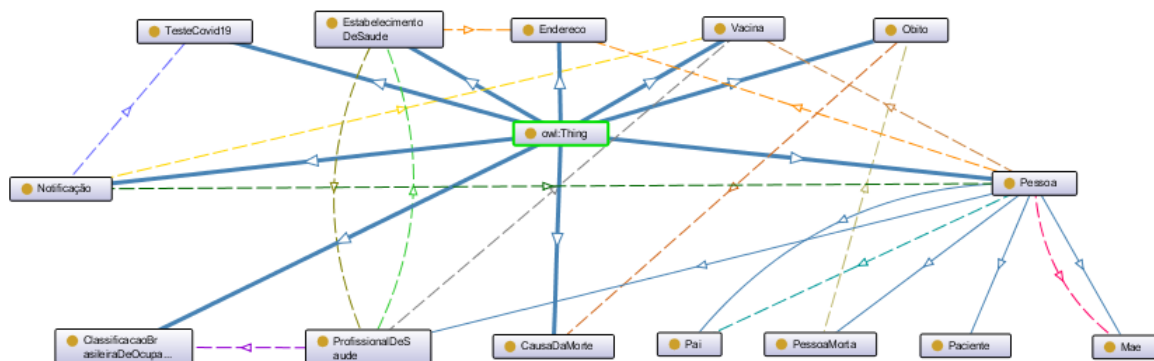


Figura 2. Estrutura da Ontologia ONTOVID. Fonte: Autores

formato computável. Também foram realizados testes de conexão com as fontes de dados e-SUS Notifica e SIM para validar o mapeamento entre a ontologia e os dados. Durante os testes, IDs únicos (identificadores) foram gerados aleatoriamente para cada registro, substituindo os dados pessoais.

Os mapeamentos foram realizados mediante a utilização da sintaxe *Turtle*<sup>5</sup> ou *TurtleSyntax*, que é semelhante a linguagem SPARQL<sup>6</sup>, e também um padrão W3C. Com a finalização dos mapeamentos, passamos para o processo de revisão da ONTOVID, avaliando a documentação e os artefatos. Neste momento também fizemos os últimos ajustes, visando o início das consultas SPARQL para validar as questões de competência da ontologia.

### 3.3. Avaliação

Como um dos processos para validação das Questões de Competência, foram realizadas consultas SPARQL sobre os dados mapeados que nos permitiram verificar a eficácia da abordagem. É importante destacar que para os testes iniciais optamos por trabalhar apenas com dados do ano de 2021. Em relação aos resultados obtidos, identificamos algumas inconsistências importantes que serão abordadas, em detalhes, nos próximos parágrafos. Sobre o número de Notificações que indicavam óbito por COVID-19 no e-SUS Notifica, foram localizados 496 registros, dos quais 9 estavam em duplicidade, porém, com *status* de cancelado ou ignorado. A consulta aos dados em questão pode ser observada através da Figura 3,

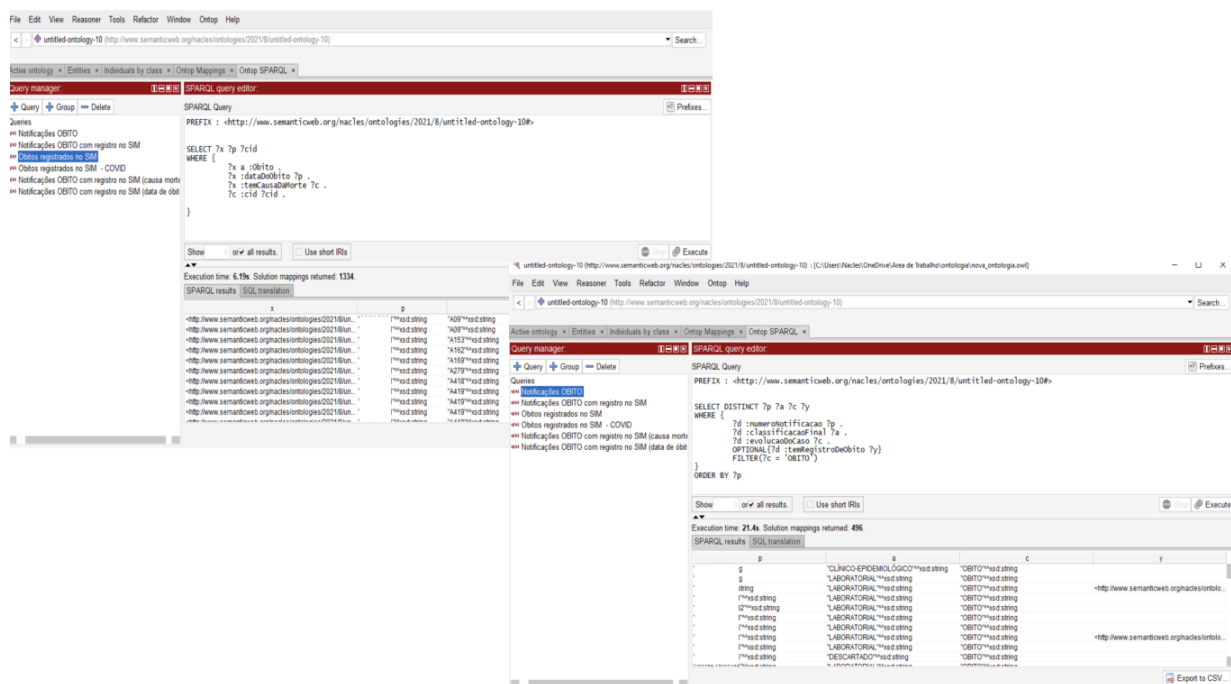
Depois de ajustar as consultas para remoção das duplicidades, realizamos a análise sobre as Questões de Competência (QC) em dois momentos distintos, primeiro em 24/11/2021 e outra avaliação, com novos indicadores, em 15/10/2022. As consultas permitiram identificar os seguintes resultados: i) **CQ 01**: 487 óbitos por COVID-19, levando em consideração apenas os casos confirmados (dados do e-SUS Notifica em 10/05/2022); ii) **CQ 02**: 17.200 curados; iii) **CQ 03**: 99 pessoas vacinadas com, no mínimo, duas doses pegaram COVID-19; iv) **CQ 04**: 257 homens, 229 mulheres e 1 não identificado morreram de COVID-19; v) **CQ 05**: 0 pessoas hospitalizadas (em 24/11/2021)<sup>7</sup>; vi) **CQ 06**: 6 vacinados com, pelo menos, duas doses da vacina vieram a óbito em 2021; vii) **CQ 07**: a

<sup>5</sup><https://www.w3.org/TeamSubmission/turtle/>

<sup>6</sup><https://www.w3.org/TR/sparql11-query/>

<sup>7</sup>Não foi possível avaliar em 10/05/2022, porque o (*status*) de classificação mudou





**Figura 3. Consultas realizadas em SPARQL para identificar as quantidades totais de Notificações e Óbitos em Camaçari-Ba. Fonte: Autores**

taxa de letalidade no município de Camaçari em 2021 foi de, aproximadamente, 2,75% - maior, se comparado a 2020, quando a taxa verificada foi de, aproximadamente, 2,23%.

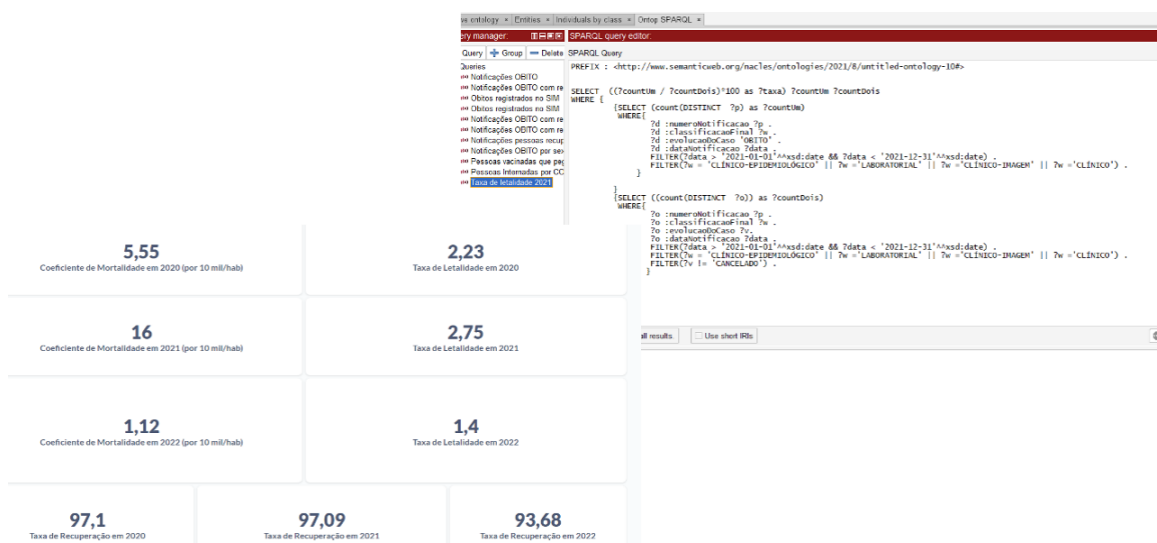
É importante salientar que identificamos divergências nos indicadores de vacinados que pegaram COVID e de mulheres que faleceram em virtude da doença entre as consultas realizadas nos dias 24/11/2021 e 10/05/2022. Todavia, essas diferenças, segundo a Secretaria de Saúde de Camaçari, são admissíveis por conta dos possíveis erros de lançamento, o que implica em atualizações constantes dos registros. O indicador relacionado aos vacinados que pegaram COVID-19 apontou para a existência de 264 pessoas em 24/11/2022 e apenas 99 pessoas em 15/10/2022. Já o total de óbitos em pessoas do sexo feminino, que era de 236 mulheres em 24/11/2021, passou para 229 em 15/10/2022.

Durante o estudo também foram identificados alguns pontos importantes que requerem uma análise criteriosa por parte dos técnicos da Secretaria de Saúde de Camaçari. A primeira refere-se ao número de óbitos por COVID-19: onde, dos 487 óbitos registros no e-SUS Notifica, apenas 412 estão na base do SIM, sendo que na consulta realizada em 24/11/2021, apenas 165 registros haviam sido localizados no SIM, dos 466 identificados no e-SUS Notifica. Todavia, essa mudança nos indicadores, segundo os técnicos da Secretaria de Saúde, pode ter ocorrido em função da retroalimentação com a base nacional que ocorre periodicamente, quando a diferença que era de 301 diminuiu para 75 registros. Para além das diferenças entre óbitos registrados no SIM com os referenciados no e-SUS Notifica, identificamos também óbitos que foram registrados no SIM com *causa base original* diferente do CID (*Classificação Internacional de Doenças*) B-34.2, que se refere à infecção por coronavírus de localização não especificada - lembrando que essa divergência pode não representar erro de registro, apenas que a COVID-19 pode não ter sido considerada como fator base para o óbito.

Dentre os registros identificados na avaliação realizada em 15/05/2022, foram localizados 27 registros que indicavam óbito por COVID-19 associado a outras doenças, como: CIDs A-41.9 (septicemia não especificada), B-20.3 (doença pelo HIV resultando em outras infecções virais), B-20.7 (doença pelo HIV resultando em infecções múltiplas), C-79.5 (neoplasia maligna secundária dos ossos e da medula óssea), D-68.9 (Defeito de coagulação não especificado), I-10 (hipertensão essencial), I-21.9 (infarto agudo do miocárdio), I-26.9 (embolia pulmonar sem menção de cor pulmonale agudo), I-61.9 (hemorragia intracerebral não especificada), J-12.9 (pneumonia viral não especificada), J-15.9 (pneumonia bacteriana não especificada), J-44.0 (doença pulmonar obstrutiva crônica com infecção respiratória aguda do trato respiratório inferior), J-98.8 (outros transtornos respiratórios especificados), K-83.1 (obstrução de via biliar), U-04.9 (síndrome respiratória aguda grave [Severe acute respiratory syndrome] [SARS], não especificada), U-07.1 (não identificado) e W-78.0 (inalação do conteúdo gástrico - residência). Já na consulta realizada em 24/11/2021, 6 óbitos associados ao COVID-19 haviam sido localizados, sendo eles: CID I-10 (hipertensão essencial), CID I-21.9 (infarto agudo do miocárdio), CID I-50.9 (insuficiência cardíaca não especificada), CID J-12.9 (infecção viral não especificada pneumonia), CID O-98.5 (outras doenças virais que complicam a gravidez, parto e puerpério) e CID U-04.9 (grave síndrome respiratória).

Outro ponto de questionamento sobre os resultados obtidos, e que está sendo avaliado pela equipe técnica da Secretaria e Saúde de Camaçari, refere-se aos óbitos com causas diferentes da COVID-19 (CID B-34.2), onde 10 registros foram localizados em ambos sistemas, e-SUS Notifica e SIM, mas não possuíam vinculação ao COVID-19 como causa base ou causa associada ao óbito.

Durante o processo de construção e validação dos GCS, dentre as consultas mais complexas, o cálculo da taxa de letalidade e mortalidade foram os que demandaram mais tempo em função dos critérios envolvidos em sua estruturação, os quais podemos observar através da Figura 4.



**Figura 4. Consulta para calcular a taxa de letalidade e mortalidade do COVID-19 no município de Camaçari-Ba. Fonte: Autores.**

#### 4. Conclusão e Trabalhos Futuros

Com base nas QC alçadas e que foram empregadas no processo avaliativo da ONTOVID, foi possível verificar que a abordagem proposta foi eficaz para a extração de dados entre as bases de Notificação (e-SUS Notifica) e de Mortalidade (SIM). O vocabulário unificado produzido nos permitiu construir Grafos de Conhecimento Semântico (GCS) para avaliar os vários aspectos relacionados à pandemia do novo coronavírus (COVID-19). As consultas complexas realizadas foram devidamente documentadas para oferecerem uma alternativa plausível para os gestores de saúde compreenderem os dados e as possíveis inconsistências.

Um ponto positivo sobre a abordagem empregada foi justamente apontar e identificar automaticamente divergências nas bases de forma automática, que precisam ser avaliadas pela equipe técnica do município de Camaçari, a saber: i) verificar se o problema das divergências relacionadas à quantidade de óbitos entre o sistema e-SUS Notifica e o SIM referem-se, exclusivamente, à questão da sincronização do SIM com a base nacional; e ii) avaliar as distorções entre os óbitos indicados como COVID no e-SUS Notifica e na Certidão de Óbito cadastrada no SIM.

Embora tenhamos obtido sucesso com a abordagem de ontologia única, pretendemos implementar uma abordagem híbrida para o modelo proposto, desenvolvendo ontologias locais para conexão direta com as fontes de dados do e-SUS e SIM. Como trabalho futuro, pretendemos também implementar novas integrações com outras fontes de dados, como a base do Cadastro Nacional de Estabelecimentos de Saúde (CNES) e do e-SUS (APS), por exemplo, a fim de possibilitar a implementação de novas consultas complexas, incluindo dashboards para visualizações gráficas.

#### Referências

- [Andrade et al. 2019] Andrade, L. O. M., Valter, R., Ramos, R., Vidal, V., Andrade, D., and Oliveira, M. (2019). Lariisa: an intelligent platform to help decision makers in the brazilian health public system. *ANAIS DO XXV SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB ARTIGOS CURTOS - SBC*.
- [Calvanese et al. 2018] Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., and Rosati, R. (2018). *Ontology-Based Data Access and Integration*, chapter Ontology-Based Data Access and Integration, page 2590–2596. Springer International Publishing.
- [Cheatham and Pesquita 2017] Cheatham, M. and Pesquita, C. (2017). Semantic data integration. in handbook of big data technologies. *Springer International Publishing*, page 263–305.
- [Ciriaco et al. 2020] Ciriaco, D. L., Pessoa, A., Salvador, L., and Wassermann, R. (2020). Integração semântica das bases de dados do sistema unico de saude: Um estudo de caso com o municipio de sao paulo. *ONTOBRAS 2020 - Ontology Research in Brazil - Proceedings of the XIII Seminar on Ontology Research in Brazil and IV Doctoral and Masters Consortium on Ontologies*, pages 23–26.
- [Ekaputra et al. 2017] Ekaputra, F., Sabou, M., Serral, E., Kiesling, E., and Biffi, S. (2017). Ontology-based data integration in multi-disciplinary engineering environments: A review. *Open Journal of Information Systems (OJIS)*, 4:1–26.

- [Ferronato et al. 2016] Ferronato, A. C. C., Pires, F. R., and Bernardini, F. C. (2016). Um modelo para integração e disponibilização de dados na Área de saúde governamental. *ANAIS DO SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO (SBSI) - SBC*.
- [Gabriel et al. 2017] Gabriel, L., Oliveira, M., Vidal, and Lais, V. (2017). Towards to a linked data framework to support decision-making on healthcare. 5th international workshop on advances in ict infrastructures and services. pages 6, 29, 31.
- [Gabriela et al. 2016] Gabriela, F. L., Vânia, M. P. V., and Mauro, O. (2016). Construção de linked data mashup para integração de dados da saúde pública. In *SBBB*.
- [Lopes et al. 2017] Lopes, G., Vidal, V., and Oliveira, M. (2017). Lais: Towards to a linked data framework to support decision-making on healthcare. *5th International Workshop on ADVANCEs in ICT Infrastructures and Services*.
- [Pereira 2019] Pereira, D. L. N. C. (2019). Integração semântica das bases de dados do sistema Único de saúde: um estudo de caso com o município de são paulo. Master's thesis, Universidade de São Paulo, São Paulo.
- [Suarez–Figuroa et al. 2008a] Suarez–Figuroa, M. C., de Cea, G. A., Carlos Buil, K. D., Fernandez–Lopez, M., Garcia, A., Asuncion Gomez–Perez, G. H., Montiel–Ponsoda, E., Marta Sabou, B. V.-T., and Yufei, Z. (2008a). D5.4.1. neon methodology for building contextualized ontology networks. Technical report.
- [Suarez–Figuroa et al. 2008b] Suarez–Figuroa, M. C., de Cea, G. A., Carlos Buil, K. D., Fernandez–Lopez, M., Garcia, A., Asuncion Gomez–Perez, G. H., Montiel–Ponsoda, E., Marta Sabou, B. V.-T., and Yufei, Z. (2008b). D5.4.1. neon methodology for building contextualized ontology networks. Technical report.
- [Suárez-Figuroa et al. 2012] Suárez-Figuroa, M. C., Gómes-Perez, A., Motta, E., and Gangemi, A. (2012). Ontology engineering in a networked world. *Springer Science and Business Media*.