

Um Enfoque Incremental para Construção do Grafo de Conhecimento do SUS

Tulio Vidal Rolim¹, Caio Viktor S. Avila¹, Narciso Arruda¹,
José Wellington F. da Silva², José Gilvan R. Maia¹,
Mauro Oliveira³, Luiz Odorico M. Andrade⁴, Vânia M. P. Vidal¹

¹Universidade Federal do Ceará (UFC)
Fortaleza, CE, Brasil.

²Universidade Federal do Ceará (UFC)
Crateús, CE, Brasil.

³Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) - Aracati
Aracati, CE, Brasil.

⁴Fundação Oswaldo Cruz
Eusébio, CE, Brasil

{tuliovr, caioviktor}@alu.ufc.br,

{narcisoarruda, jwellingtonfranco, gilvanmaia

odorico0811, amauroboliveira, vaniapvidal}@gmail.com

Abstract. *In Brazil, public health data are distributed on several heterogeneous bases belonging to the Unified Health System (SUS). Due to the interdependence between these data, the analysis of any problem is preceded by the integration of some of these bases. As an alternative, corporate knowledge graphs (enterprise knowledge graphs) may enable the semantic integration of SUS sources. This work presents an approach for building a graph of organizational knowledge of SUS. This approach is supported through the incremental construction based on pay-as-you-go through the combination of Ontologies and the Semantic Integration process, providing a semantic layer to SUS data. For this, the steps of the approach are presented together with the construction of the SUS enterprise knowledge graph (KG SUS). KG SUS integrates two data sources: SIM and SINASC, as validation KG SUS was consulted through questions of competence in SPARQL. The approach presented promotes flexibility and extensibility for new SUS data sources to be used.*

Resumo. *Os dados de saúde pública no Brasil estão distribuídos em diversas bases heterogêneas pertencentes ao Sistema Único de Saúde (SUS). Por conta da interdependência entre esses dados, a análise de qualquer problema é precedida pela integração de algumas dessas bases. Como alternativa, os grafos de conhecimento corporativos (enterprise knowledge graphs) podem ser utilizados para viabilizar a integração semântica das fontes do SUS. Este trabalho apresenta um enfoque para construção de um grafo de conhecimento corporativo do SUS. Esse enfoque é apoiado através da construção incremental baseada no pay-as-you-go por meio da combinação de Ontologias e do processo de Integração Semântica, fornecendo uma camada semântica aos dados*

do SUS. Para tanto, são apresentados os passos do enfoque juntamente com a construção do grafo de conhecimento corporativo do SUS (KG SUS). KG SUS realiza a integração de duas fontes de dados: SIM e SINASC. Como validação, KG SUS foi consultado através de questões de competência em SPARQL. O enfoque apresentado promove flexibilidade e extensibilidade para que novas fontes de dados do SUS sejam utilizadas.

1. Introdução

A quantidade de dados públicos disponíveis que estão relacionados ao domínio da Saúde tem crescido significativamente nos últimos anos [Viacava et al. 2018]. A tarefa de exploração e análise de dados do Sistema Único de Saúde (SUS) é fundamental para a descoberta de conhecimento que pode ser utilizado para o desenvolvimento de políticas públicas com impacto direto sobre a saúde da população. No entanto, essa tarefa possui alguns desafios associados que precisam ser considerados. Por exemplo, a eficácia da descoberta de conhecimento depende da preparação adequada dos dados e da interpretação dos resultados, o que apresenta alguns desafios, tais como fontes de dados heterogêneas e distribuídas.

Um dos principais desafios na integração dos dados do SUS está relacionado ao problema da interoperabilidade de dados presentes em fontes heterogêneas. De acordo com [Bishr 1998], a heterogeneidade pode ser sintática, esquemática ou semântica. A heterogeneidade sintática é causada pelo uso de diferentes modelos para representar os dados. A heterogeneidade de esquema é resultante de diferenças estruturais entre as bases. Por fim, tem-se a heterogeneidade semântica que é causada pelos diferentes significados e interpretações dos dados em diferentes contextos. Para que se possa alcançar a interoperabilidade de dados é necessário integrar semanticamente as fontes de dados.

Compreendemos como **integração semântica** o processo que faz uso de uma representação conceitual dos dados e seus relacionamentos para eliminar possíveis heterogeneidades. A representação conceitual pode ser feita por meio do uso de ontologias que são, por definição, uma representação formal e explícita de uma conceitualização compartilhada [Studer et al. 1998].

Nessa conjuntura, os grafos de conhecimento corporativo ou (*Enterprise Knowledge Graphs* - **EKG**) estão sendo utilizados como um mecanismo para consolidar e integrar semanticamente um grande número de fontes de dados heterogêneas em um espaço de dados abrangente [Gomez-Perez et al. 2017].

Deste modo, um EKG pode ser construído visando integrar os dados do SUS fornecendo uma visão semanticamente conectada dos dados de modo a prover um acesso integrado às fontes de dados por meio do uso de aplicações semânticas. Entretanto, o processo de construção de um EKG não é trivial, apresentando desafios para integrar os dados advindos de sistemas, fontes e departamentos distintos, bem como na escassez de ferramentas e métodos formais.

Neste trabalho é apresentada uma solução baseada em ontologias para a construção incremental do EKG do SUS (*Semantic SUS*). O enfoque proposto é baseado em quatro pontos principais:

- Uso de uma arquitetura de Integração de Dados baseada no enfoque ODBA;

- Uso de um método, baseado no enfoque *pay-as-you-go* [Madhavan et al. 2007] e tecnologias da Web semântica para a Integração semântica das fontes de dados;
- Uso de uma ontologia (EKG-O) [Ver Seção 3] para representar a estrutura do EKG (*i.e.*, os artefatos gerados em cada passo do processo de construção do EKG). O propósito da EKG-O é identificar e representar os principais conceitos e axiomas no domínio de EKG, e dentro do escopo da arquitetura e método propostos; e
- Uso de um *Knowledge Graph* (KG SUS), o qual é uma instanciação da ontologia de EKG-O, e descreve os componentes do EKG do SUS. Nesse contexto, um KG que representa semanticamente os componentes do KG SUS, é fundamental para facilitar o uso e evolução do EKG automatizar alguns passos do processo de construção de um EKG para o SUS, identificar inconsistências e avaliar a qualidade dos artefatos gerados durante cada passo da construção/evolução do EKG.

O restante deste artigo está organizado como se segue. A Seção 2 apresenta a arquitetura do EKG Semantic SUS. Por conseguinte, na Seção 3 apresenta-se uma visão geral do enfoque e seu processo. Já na Seção 4 são expostos os passos para especificação e publicação das visões exportadas e de *linksets*. Na Seção 5 é realizada a validação do EKG construído sob a realização de consultas SPARQL com base em questões de competência. Por fim, a Seção 6 apresenta as considerações finais do trabalho.

2. Arquitetura do EKG Semantic SUS

O SUS é composto por diversas fontes de dados heterogêneas contendo vocabulários distintos¹. Nessa perspectiva, foram utilizadas inicialmente no *EKG Semantic SUS* as fontes de dados do Sistema de Informação sobre Mortalidade (SIM)² e do Sistema de Informações sobre Nascidos Vivos (SINASC)³.

O *EKG Semantic SUS* consiste em um EKG representado por um *Knowledge Graph* (KG) construído através das visões exportadas e de *linksets* juntamente com uma ontologia de domínio. Esse EKG integra as duas fontes de dados, SIM e SINASC, como um modelo para descoberta de informações para a problemática de gestantes de risco.

A arquitetura do *EKG Semantic SUS* é apresentada na Figura 1. Essa arquitetura é dividida em quatro camadas: Camada de Fontes de Dados; Camada de Publicação de Dados; Camada de Acesso aos Dados; e por fim, a Camada de Aplicações. Estas camadas são apresentadas em mais detalhes a seguir.

2.1. Camada de Fontes de Dados

Esta camada é composta pelas fontes de dados relativos ao domínio da saúde, integrantes do SUS. O formato de armazenamento destas fontes pode estar em diferentes formatos, *e.g.*, bancos de dados relacionais, CSV, *triple stores* RDF, documentos JSON, etc.

Estas fontes podem conter informações complementares sobre objetos em comum advindos de departamentos, sistemas e fontes distintas. Consequentemente, a recuperação de todas as informações disponíveis para um mesmo objeto pode possibilitar a implantação de aplicações e estudos mais sofisticados, profundos e impactantes. Surge assim a necessidade do acesso integrado a estas fontes.

¹<http://tabnet.datasus.gov.br/tabdata/livroidb/2ed/fontes.pdf>

²<http://sim.saude.gov.br/>

³<http://www2.datasus.gov.br/DATASUS/index.php?area=060702>

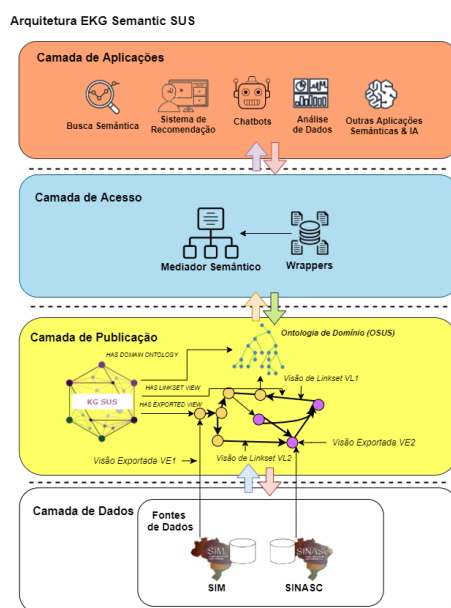


Figura 1. Arquitetura do EKG Semantic SUS.

No entanto, o acesso integrado a essas fontes é desafiador, pois cada uma possui um mecanismo de acesso diferente, o que demanda o uso de diferentes técnicas. Além disso, cada fonte pode ser estruturada seguindo um vocabulário diferente, o que dificulta a compreensão e a relação entre os dados de fontes diferentes.

2.2. Camada de Publicação de Dados

Esta camada torna as fontes de dados subjacentes transparentes, onde nela os dados são publicados como uma visão de grafo RDF. Esta visão fornece o acesso aos dados através de um único ponto de acesso, representado pelo EKG. Este EKG expõe todos os dados seguindo um único vocabulário comum, definido pela ontologia de domínio. Onde o acesso aos dados é realizado através de um único método de acesso, ou seja, consultas SPARQL.

Além disso, no EKG, as diferentes representações de um mesmo objeto do mundo real através das diferentes fontes são identificadas e conectadas por *links owl:sameAs*. Mais detalhes sobre o processo de construção de um EKG são apresentados na Seção 3.

2.3. Camada de Acesso aos Dados

Esta camada permite o acesso ao KG SUS através do Mediador Semântico (MS) juntamente com os *wrappers* relativo as visões exportadas de cada fonte de dados.

2.4. Camada de Aplicações

Aplicações de busca semântica, Q&A como chatbots, sistemas de recomendação e outras aplicações de inteligência artificial e *machine learning* que utilizam da semântica podem acessar o KG SUS através de um MS e *wrappers*. Isso posto, tais aplicações podem usufruir da semântica provida através dos dados integrados no EKG para realizar consultas de interesse no domínio de saúde.

3. Visão Geral do Enfoque

A construção do EKG é baseado em um enfoque que combina ontologias e dados interligados para enfrentar os desafios no desenvolvimento de aplicações onde existe a necessidade de integrar fontes de dados heterogêneas

O enfoque norteia-se na especificação das visões dos dados e suas ligações (*linksets* para realização da integração semântica dos dados constituintes do EKG juntamente com a publicação das visões.

No enfoque proposto neste trabalho, as visões exportadas e de *linksets* podem ser tratadas de modo materializado ou virtual. Para tanto, a especificação define as fontes, mapeamentos e visões de *linksets* como virtuais ou materializadas, podendo o EKG ser acessado diretamente através de um *triplestore* mediante *endpoint* SPARQL quando materializadas ou acessados diretamente via *Wrappers* quando virtuais.

Neste trabalho, são utilizadas visões virtuais exportadas e de *linksets* para construção do *KG SUS*. A principal motivação concentra-se no *query answering* através da linguagem SPARQL, gerando-se um EKG virtual com base na consulta SPARQL sobre as visões virtuais das fontes de dados. Em tempo consulta, uma consulta SPARQL *SQ* é processada na visão virtual, convertendo-se em uma consulta SQL *Q* que pode ser executada diretamente nas fontes de dados com base no *schema* da ontologia, garantindo a atualização sempre que ocorrerem mudanças nos dados presentes nas fontes.

Para tanto, os passos do enfoque utilizado para construção do *KG SUS* são compreendidos da seguinte forma:

1. **Modelagem da Ontologia de Domínio:** Passo contendo o processo de construção da ontologia de domínio com base nos conceitos presentes nas fontes de dados;
2. **Especificação e Publicação das Visões Exportadas:** Etapa de especificação das visões exportadas, onde são definidas as fontes de dados, a ontologia exportada e os mapeamentos que traduzem os dados da vocabulário da fonte de dados para o vocabulário da visão exportada, e posterior publicação da especificação da visão exportada em um *TripleStore* no caso de visão materializada ou mediante uso de *Wrappers* tratando-se de visões virtuais;
3. **Especificação e Publicação das Visões de *Linksets*:** As quais serão usadas para gerar *links* entre instâncias em diferentes fontes de dados e por conseguinte publicá-las em um Mediador Semântico.

3.1. Representação Semântica do *KG SUS* através da *EKG Ontology*

Durante a construção do *KG SUS*, os passos foram guiados através da representação e especificação por meio da *EKG Ontology* (EKGO)⁴, uma ontologia que descreve os conceitos presentes em um EKG.

EKGO é utilizada como uma ontologia de referência para EKGs nos mais variados domínios, tal aspecto é fundamentado através do fato de que a EKGO baseia-se em uma abordagem formal para construção e manutenção de EKGs, fornecendo ainda a possibilidade de reúso e expansão por meio da adição de novas fontes ou conceitos.

⁴<http://tiny.cc/ekgo>

KG SUS baseia-se na representação ontológica da EKG0, sendo definido por uma tripla = $\{OD, E, L\}$, onde: **OD**: refere-se a ontologia de domínio construída com termos presentes em cada visão exportada. É representada na EKG0 como uma *omv:Ontology* e relacionada com o KG SUS através da *Object Property ekgo:hasDomainOntology*; **E**: refere-se a uma Visão Exportada definida sobre uma fonte de dados *S*, utilizando uma ontologia exportada *OE* e mapeamentos *ME*, formando uma Especificação da Visão Exportada *ES*. **E**: é definida como uma classe *ekgo:ExportedView* relacionando-se com o KG SUS através da propriedade *ekgo:hasExportedView*; **L**: é uma Visão de *Linkset*, as quais especificam como identificar objetos em diferentes fontes de dados que representam o mesmo objeto no mundo real. Na EKG0 uma Visão de *Linkset* *L* é representada como uma *ekgo:LinksetView* e relacionada por *ekgo:hasLinksetView*.

Cada Visão Exportada e Visão de *Linkset* possui uma Especificação e uma Publicação. Para construção do KG SUS foram criadas instâncias (Conforme Figura 2) referentes aos passos do enfoque apresentado, sendo: **KG SUS**: Instância do *Knowledge Graph* do SUS que é construído incrementalmente pela ontologia de domínio, visões exportadas e de *linksets*; **OntologiaSUS**: Instância da Ontologia de Domínio do SUS; **VisaoExportadaSIM**: Instância da Visão Exportada da fonte de dados SIM; **VisaoExportadaSINASC**: Instância da Visão Exportada da fonte de dados SINASC; **EspecificacaoVisaoExportadaSIM**: Instância da Especificação da Visão Exportada da fonte de dados SIM; **EspecificacaoVisaoExportadaSINASC**: Instância da Especificação da Visão Exportada da fonte de dados SINASC; **PublicacaoVisaoExportadaSIM**: Instância da Publicação da Visão Exportada da fonte de dados SIM; **PublicacaoVisaoExportadaSINASC**: Instância da Publicação da Visão Exportada da fonte de dados SINASC; **VisaoLinksetSIMSINASC** Instância da Visão de *Linksets* das fontes de dados SIM e SINASC; **EspecificacaoVisaoLinksetSIMSINASC**: Instância da Especificação da Visão de *Linksets* das fontes de dados SIM e SINASC; **PublicacaoVisaoLinksetSIMSINASC**: Instância da Publicação da Visão de *Linksets* das fontes de dados SIM e SINASC.

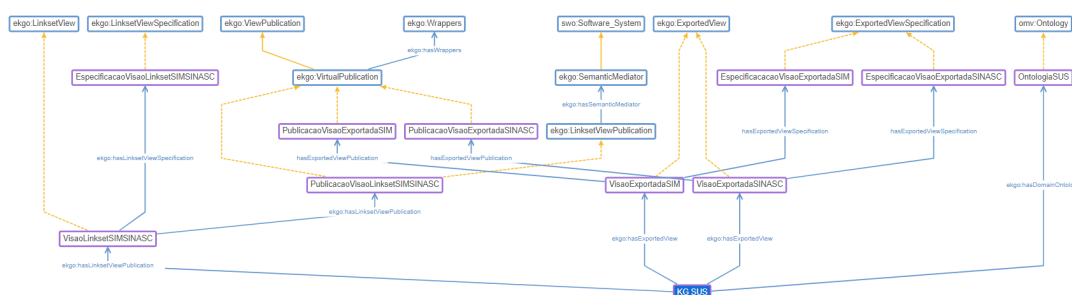


Figura 2. Visão Geral Simplificada do KG SUS com base na EKG0 e suas instâncias.

4. Construção do KG SUS

Nesta seção são apresentados os passos para construção do KG SUS utilizando as fontes de dados do SIM e SINASC com base no enfoque proposto.

4.1. Modelagem Ontologia de Domínio

Nesse passo, as fontes de dados SIM e SINASC foram armazenadas no *SGBD PostgreSQL* onde posteriormente utilizou-se a técnica de *bootstrapping* para geração automática das ontologias exportadas através do Ontop [Calvanese et al. 2017]. Cada ontologia exportada é tratada como um vocabulário parcial do domínio geral, ou seja, um recorte ou visão exportada dos conceitos presentes na ontologia de domínio.

A ontologia de domínio foi instanciada como *OntologiaSUS* posteriormente sendo relacionada com a instância *EKG SUS* através da *Object Property ekgo:hasDomainOntology*.

4.2. Especificação e Publicação das Visões Exportadas

Cada Visão Exportada *E* possui uma Especificação da Visão Exportada *ES* contendo uma fonte de dados, ontologia exportada e mapeamentos. Uma *ES* é então uma tripla (*S, OE, ME*), onde:

- *ES*: é representada através de uma instância *ekgo:ExportedViewSpecification*;
- *S*: é uma fonte de dados representada como uma *drm:DataAsset*, sendo um conceito reutilizado do vocabulário *Data Reference Model drm*:⁵. Por sua vez, um *drm:DataAsset* possui as sub-classes *ldp:RDFSource* para representação de fontes de dados RDF e *ldp:RDFNonSource* para representação de fontes não-rdf, ambas providas pelo vocabulário *Linked Data Platform ldp*: (LDP) ⁶. *ldp:RDFNonSource* possui como tipos especializado a subclasse *rdbso:RelationalDatabase* para representar fontes de dados relacionais através do vocabulário *Relational Database System Ontology rdbso*: [de Aguiar et al. 2018];
- *OE* é a ontologia exportada. *OE* é um recorte de *OD*, o qual contém os termos de *OD* que se relacionam com termos do *schema* de uma fonte *Si*. *OE* é uma classe *omv:Ontology*; e
- *ME*: é um conjunto de mapeamentos que relacionam termos do vocabulário de *OE* com termos do *schema* da fonte *Si*. *ME*: é representada como *ekgo:Mappings* tendo a subclasse *ekgo:R2RMLMappings* para os mapeamentos de fontes relacionais.

A composição da *ekgo:ExportedViewSpecification* é então formada por: *drm:DataAsset*, *omv:Ontology*, *ekgo:Mappings*, relacionada por meio das propriedades *ekgo:hasDataSource*, *ekgo:hasExportedOntology*, *ekgo:hasMappings*.

Nesse passo, inicialmente os mapeamentos foram gerados através do *bootstrapping* utilizando o *software* MIRROR [de Medeiros et al. 2015] e revisados através da ferramenta Map-On [Sicilia et al. 2017].

De modo a representar os mapeamentos voltados as fontes relacionais SIM e SINASC, foram criadas as instâncias *MapeamentoR2MLSIM* e *MapeamentoR2RMLSINASC* da classe *ekgo:R2RMLMappings*. Após a definição dos mapeamentos, a especificação de cada visão exportada foi instanciada como *EspecificacaoVisaoExportadasSIM* e *EspecificacaoVisaoExportadasSINASC*, sendo relacionada pelas propriedades *ekgo:hasDataSource*, *ekgo:hasExportedOntology*, *ekgo:hasMappings*

⁵<https://lov.linkeddata.es/dataset/lov/vocabs/drm>

⁶<https://www.w3.org/TR/ldp/>

com as instâncias das fontes de dados = $\{FonteSIM \text{ e } FonteSINASC\}$, ontologias exportadas = $\{OntologiaExportadaSIM \text{ e } OntologiaExportadaSINASC\}$ e mapeamentos = $\{MapeamentoR2MLSIM \text{ e } MapeamentoR2RMLSINASC\}$.

Para estabelecer uma relação entre as Visões Exportadas do SIM e SINASC junto com suas especificações, foi criada a instância *VisaoExportadaSIM* e *VisaoExportadaSINASC*, da classe *ekgo:ExportedView* e utilizada a propriedade *ekgo:hasExportedViewSpecification*, após isso, é estabelecida a relação entre um EKG e uma Visão Exportada através da *EKG SUS* utilizando a propriedade *ekgo:hasExportedView* tendo com *rdfs:range* as instância *VisaoExportadaSIM* e *VisaoExportadaSINASC*.

A Figura 3 apresenta uma visão das instâncias envolvidas na especificação das visões exportadas e suas relações com base na EKG.

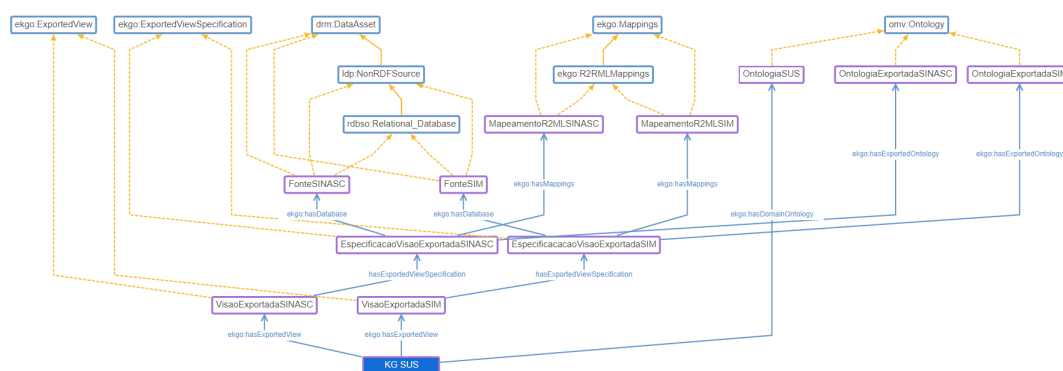


Figura 3. Construção do KG SUS - Especificação das Visões Exportadas.

Após a realização do passo de Especificação das Visões Exportadas, é realizada a Publicação das Visões Exportadas das fontes *SIM* e *SINASC*. Em razão da constante frequência de atualização das fontes, optou-se pelo uso do enfoque virtual para Publicação das Visões Exportadas de modo a visar a recência dos dados.

O processo de publicação das visões virtuais do EKG são baseadas no OBDA, onde através dos mapeamentos uma dada consulta SPARQL é processada através de um *Mediador Semântico* sobre os conceitos e propriedades de uma ontologia de domínio presente em uma visão exportada, provendo o resultado da consulta em um VKG [Xiao et al. 2018]. Em um VKG não se há a necessidade de materialização dos dados das fontes integradas, sua construção é feita *on-the-fly*.

Logo, o VKG construído com base no enfoque desse estudo é consistido na publicação de visões virtuais exportadas representada nas EKG através da classe *ekgo:VirtualPublication*.

Nessa etapa, a publicação das visões exportadas virtuais ocorreu através do *Mediador Semântico* Ontop, em conjunto foi utilizado o Teiid⁷ uma integração do Ontop para realização de consultas federadas sob fontes relacionais. Com base na ontologia *EKGO* criou-se então uma instância *PublicacaoVisaoExportadaSIM* e outra

⁷<http://teiid.io/legacy/>

PublicacaoVisaoExportadaSINASC do tipo *ekgo:VirtualPublication* relacionada com a instância **Mediador Semântico** do tipo *ekgo:SemanticMediator* através da propriedade *ekgo:hasSemanticMediator*. A instância **EKGEstudoCaso** relaciona-se com uma publicação de visão exportada utilizando a propriedade *ekgo:hasExportedView* tendo o *rdfs:range* a instâncias **PublicacaoVisaoExportadaSIM** e **PublicacaoVisaoExportadaSINASC**. Na Figura 4 são apresentados os detalhes da relação das instâncias durante a publicação das visões virtuais exportadas.

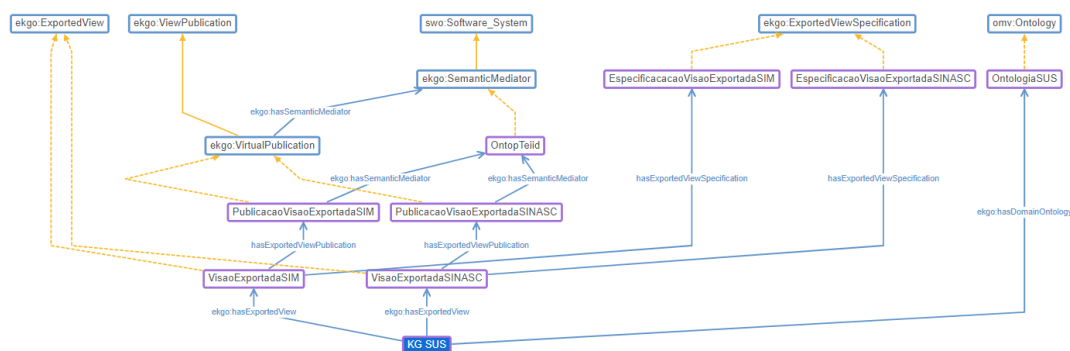


Figura 4. Construção do KG SUS - Publicação das Visões Exportadas.

4.3. Especificação e Publicação das Visões de Linksets

A especificação de *linksets* consiste na definição de um conjunto de regras de linkage para geração de *linksets* entre as classes semanticamente semelhantes das ontologias exportadas. Linksets são usados para ligar objetos em fontes distintas mas que representam o mesmo objeto do mundo real (resolução de entidades). Em OWL⁸, esses links são estabelecidos através da propriedade *owl:sameAs*. O processo de identificação desses *links* é conhecido como “linkage de dados”. Neste passo, são definidas as regras de linkage entre as instâncias das fontes de dados do SIM e SINASC,

No KG SUS foi realizada a relação entre os dados de pessoas mortas advindos do SIM *sim:PessoaMorta* e de recém-nascidos advindos do SINASC *sinasc:RN* mediante uso do *owl:sameAs* utilizando as propriedades *nome* e *registro_de_nascimento*.

As instâncias relativas ao processo de especificação foram criadas como **VisaoLinksetSIMSINASC** do tipo *ekgo:EspecificacaoVisaoLinksetsSIMSINASC* que por sua vez possui uma *linkage rule Equality*. Após, uma visão de *Linkset* foi instanciada como **VisaoLinksetSIMSINASC** do tipo *ekgo:LinksetView* tendo uma **EspecificacaoVisaoLinksetsSIMSINASC** através da propriedade *ekgo:hasLinksetViewSpecification*. As visões exportadas *source* e *target* que fazem parte da especificação são acessadas através das visões exportadas **PessoaMortaSIM** e **PessoaMortaRN** mediante propriedade *ekgo:hasExportedView*. A Figura 5 apresenta instâncias e classes representadas nesta fase.

A publicação das visões de *linksets* ocorreu no Mediador Semântico através da especificação das visões de *linksets* e realizando os *links on-the-fly*, associando as instâncias em comum através de *links owl:sameAs* virtuais.

⁸<https://www.w3.org/OWL/>

Primeiro foi criada uma instância *VisaoLinksetSIMSINASC* do tipo *ekgo:LinksetView* associada com uma instância *EspecificacaoVisaoLinksetsSIMSINASC* através da propriedade *ekgo:hasSpecification*. Após estabelecer a relação entre uma Visão de *Linkset* e uma Especificação ocorre a realização da publicação. Uma publicação de visão de *linkset* virtual é representada por meio de uma instância do tipo *ekgo:LinksetVirtual* chamada *PublicacaoLinksetsSIMSINASC*. Uma *PublicacaoLinksetsSIMSINASC* é publicada em uma instância *MediadorSemantico* do *MS ekgo:SemanticMediator*.

Para publicação dessa visão de *linksets* no KG SUS, a instância *EKGEstudo-Caso* adquire acesso a uma visão de *linkset VisaoLinksetSIMSINASC* contendo uma especificação das visões de *linksets* mediante uso da propriedade *ekgo:hasLinksetView* possuindo como *rdfs:range* uma instância *VisaoLinksetSIMSINASC*. Na Figura 5 são apresentadas as instâncias e classes descritas na etapa de publicação.

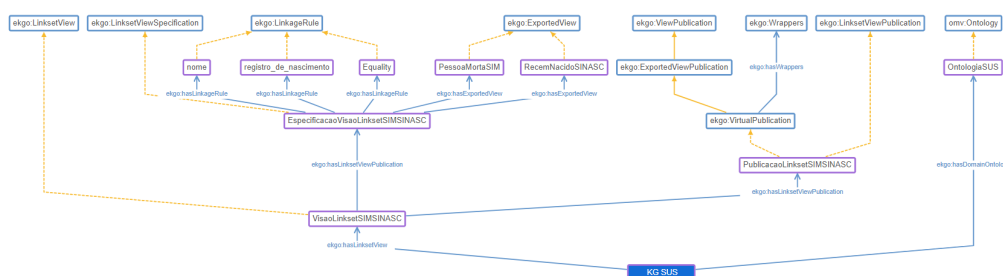


Figura 5. Construção do KG SUS - Publicação das Visões de Linksets.

5. Consultando o EKG (KG SUS)

De modo validar a capacidade funcional do KG SUS, o EKG foi povoado através do processo de integração com dados advindos das fontes do SIM e SINASC.

Como estudo de caso foi proposto a utilização de uma questão de competência [Grüninger and Fox 1995] para formulação de uma consulta apta a responder a seguinte pergunta: **(QC1)** “Quais as informações sobre o parto das crianças que vieram a óbito com menos de 28 dias e tiveram asfixia ao nascer?”.

O processo para realização da consulta é feito sequencialmente adotando-se um conjunto de passos para construção e processamento da consulta da seguinte forma:

1. Uma consulta Q é formulada através dos conceitos com base no *schema* da ontologia de domínio;
2. O Mediador Semântico (MS) traduz a consulta SPARQL Q para uma consulta SQL Q' sobre um *schema* global que representa todas as fontes de dados como uma única fonte;
3. O MS identifica as fontes não triplicadas NS (SIM e SINASC) Q' e para cada fonte NS_j , onde $1 \leq j \leq n$, produz uma sub-consulta Q'_j sobre as visões exportadas do EKG SUS advindas das fontes locais NS_j através do *Ontop* juntamente com o Teiid que traduz a subconsulta em Q'_j uma consulta sobre o *schema* das fontes locais;

4. O MS executa cada uma das consultas Q'_1, Q'_2, \dots, Q'_n em seus respectivos *wrappers* e recupera as tuplas de resposta com base nas visões exportadas do SIM e SINASC;
5. O MS realiza o processo de agregação com base nas visões dos *linksets owl:sameAs*; e
6. O resultado é retornado, conforme apresentado na Figura 6.

http://200.19.182.252/resources/BN#ANTONIA%20SILVANEIDE%20CAVALCANTE%2012-04%2000%3A00%3A00.00000527573242	http://200.19.182.252/resources/Nascimento#ANTONIA%20SILVANEIDE%20CAVALCANTE%2012-04%2000%3A00%3A00.00000527573242
http://200.19.182.252/resources/BN#KILVIA%20MIRELLA%20VERISSIMO%20REIS%2011-03-06%2000%3A00%3A00.00000527570862	http://200.19.182.252/resources/Nascimento#KILVIA%20MIRELLA%20VERISSIMO%20REIS%2011-03-06%2000%3A00%3A00.00000527570862
http://200.19.182.252/resources/BN#MURIELLE%20SOARES%20DE%20FREITAS%2011-10-24%2000%3A00%3A00.00000579445122	http://200.19.182.252/resources/Nascimento#MURIELLE%20SOARES%20DE%20FREITAS%2011-10-24%2000%3A00%3A00.00000579445122
http://200.19.182.252/resources/BN#MARIA%20ZILDA%20DE%20SOUZA%2012-01-18%2000%3A00%3A00.00000579447402	http://200.19.182.252/resources/Nascimento#MARIA%20ZILDA%20DE%20SOUZA%2012-01-18%2000%3A00%3A00.00000579447402
http://200.19.182.252/resources/BN#ADRIANA%20MARIAS%20DF%20OLIVEIRA%2008-06-05%2000%3A00%3A00.0000044255972	http://200.19.182.252/resources/Nascimento#ADRIANA%20MARIAS%20DF%20OLIVEIRA%2008-06-05%2000%3A00%3A00.0000044255972
http://200.19.182.252/resources/BN#ALEXSANDRA%20DOS%20SANTOS%20BARBOSA%2014-06-30%2000%3A00%3A00.00000644573632	http://200.19.182.252/resources/Nascimento#ALEXSANDRA%20DOS%20SANTOS%20BARBOSA%2014-06-30%2000%3A00%3A00.00000644573632
http://200.19.182.252/resources/BN#ANA%20CARLA%20SOUZA%20VIAN%2014-01-29%2000%3A00%3A00.00000609788272	http://200.19.182.252/resources/Nascimento#ANA%20CARLA%20SOUZA%20VIAN%2014-01-29%2000%3A00%3A00.00000609788272
http://200.19.182.252/resources/BN#ANA%20CLELIA%20GOMES%20DUARTE%2016-02-13%2000%3A00%3A00.00000703183832	http://200.19.182.252/resources/Nascimento#ANA%20CLELIA%20GOMES%20DUARTE%2016-02-13%2000%3A00%3A00.00000703183832
http://200.19.182.252/resources/BN#ANA%20KAROLENE%20ROARFES%20SILVA%2016-05-01%2000%3A00%3A00.000007217520972	http://200.19.182.252/resources/Nascimento#ANA%20KAROLENE%20ROARFES%20SILVA%2016-05-01%2000%3A00%3A00.000007217520972

Figura 6. Resultado da Consulta.

6. Considerações Finais

Em [Vidal et al. 2015] os autores apresentam uma especificação formal para EKG. Ela permite a materialização automática da EKG. Essa especificação é usada como referencia para criação da representação da EKG. Dessa forma, a materialização da EKG pode ser feita de forma automática com o uso dessa especificação.

Em [Arruda et al. 2020] os autores apresentaram um vocabulário para representação da especificação da EKG virtual e especializado. Esse trabalho, no entanto não trata o processo de criação da EKG, ele não apresenta uma representação para publicação da EKG.

Já em [Rolim et al. 2019], foi apresentado uma proposta de integração semântica entre as bases do Cadastro de Empresas Inidôneas e Suspensas (CEIS), do Cadastro Nacional de Empresas Punidas (CNEP) e do Sistema Integrado de Administração de Serviços Gerais (SIASG) criando uma visão integrada dessas bases chamada SemanticSEFAZ. Em [da Cruz et al. 2019], os autores também construíram uma EKG sobre as bases SIM e SINASC. No entanto, esses trabalhos não apresentam um vocabulário para representar a especificação e o processo de criação da EKG.

Neste trabalho foi apresentado um enfoque incremental para construção de um EKG com dados enriquecidos semanticamente para o SUS. Para tanto, a proposta foi apresentada e validada, tendo como estudo de caso a problemática de Gestantes de Risco através das fontes do SIM e SINASC. A validação do enfoque sob o estudo de caso mostraram que o trabalho apresentou-se viável como um modelo para construção de EKGs no domínio da saúde, sendo capaz de adaptar-se também a outras problemáticas, tais como para fornecer um EKG com dados de interesse no combate ao novo COVID-19.

Como trabalhos futuros, pretende-se criar uma plataforma semântica interativa para suportar o enfoque de construção de EKGs. Ainda, visa-se realizar outros experimentos com especialistas no domínio de saúde e Engenheiros do Conhecimento e outros

profissionais com domínio em tecnologias semânticas, ontologias e *linked data* de modo a aprimorar o estudo.

Referências

- Arruda, N., Venceslau, A. D., da Cruz, M. M. L., Vidal, V. M. P., and Pequeno, V. M. (2020). Publishing and consuming semantic views for construction of knowledge graphs. In *ICEIS (1)*, pages 197–204.
- Bishr, Y. (1998). Overcoming the semantic and other barriers to gis interoperability. *International journal of geographical information science*, 12(4):299–314.
- Calvanese, D. et al. (2017). Ontop: Answering sparql queries over relational databases. *Semantic Web*, 8(3):471–487.
- da Cruz, M. M. L., Avila, C. V. S., Vidal, V. M. P., and Junior, N. M. A. (2019). Semanticus: Um portal semântico baseado em ontologias e dados interligados para acesso, integração e visualização de dados do sus. In *Anais Estendidos do XIX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 13–18. SBC.
- de Aguiar, C. Z., de Almeida Falbo, R., and Souza, V. E. S. (2018). Ontological representation of relational databases. In *ONTOBRAS*, pages 140–151.
- de Medeiros, L. F., Priyatna, F., and Corcho, O. (2015). Mirror: Automatic r2rml mapping generation from relational databases. In *International Conference on Web Engineering*, pages 326–343. Springer.
- Gomez-Perez, J. M., Pan, J. Z., Vetere, G., and Wu, H. (2017). Enterprise knowledge graph: An introduction. In *Exploiting linked data and knowledge graphs in large organisations*, pages 1–14. Springer.
- Grüninger, M. and Fox, M. S. (1995). The role of competency questions in enterprise engineering. In *Benchmarking—Theory and practice*, pages 22–31. Springer.
- Madhavan, J., Jeffery, S. R., Cohen, S., Dong, X. L., Ko, D., Yu, C., and Halevy, A. (2007). Web-scale data integration: You can only afford to pay as you go.
- Rolim, T. V., Vidal, V. M. P., Avila, C. V. S., Cruz, M. M. L. d., Barrio, M., and Queiroz, D. (2019). Semanticsefaz: an ontology-based semantic portal for the government spending. In *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, pages 493–496. ACM.
- Sicilia, Á., Nemirovski, G., and Nolle, A. (2017). Map-on: A web-based editor for visual ontology mapping. *Semantic Web*, 8(6):969–980.
- Studer, R., Benjamins, V. R., and Fensel, D. (1998). Knowledge engineering: principles and methods. *Data & knowledge engineering*, 25(1-2):161–197.
- Viacava, F. et al. (2018). Sus: supply, access to and use of health services over the last 30 years. *Ciencia & saude coletiva*, 23(6):1751–1762.
- Vidal, V. M. et al. (2015). Specification and incremental maintenance of linked data mashup views. In *CAiSE*, pages 214–229. Springer.
- Xiao, G., Calvanese, D., Kontchakov, R., Lembo, D., Poggi, A., Rosati, R., and Zakharyashev, M. (2018). Ontology-based data access: A survey. *IJCAI*.