

A WEB SEMÂNTICA APLICADA NA RECUPERAÇÃO DE INFORMAÇÃO: UM ESTUDO DE CASO NO CONTEXTO ESTATÍSTICO DE USO DE LIVROS DIGITAIS POR ALUNOS DE GRADUAÇÃO

Alessandra Monteiro Pattuzzo Caetano

Henrique Monteiro Cristovão

Stella Schwanz Dias de Assis



INTRODUÇÃO

A oferta de bibliotecas virtuais vem aumentando no país, e as instituições de ensino superior precisam inovar o processo de acesso e disponibilização da informação. Este é um cenário que não tem volta, e os profissionais bibliotecários, além de disponibilizar, precisam apresentar e acompanhar o uso efetivo dos livros digitais para tomadas de decisão em planejamentos administrativos, financeiros e pedagógicos em suas unidades de informação.

Em relação ao desenvolvimento de coleções, as decisões de cancelar ou permanecer com o acervo virtual, ocorre durante seu tempo de vigência por meio de controle estatístico de uso e se o acervo disponibilizado é viável para a composição das referências bibliográficas dos planos de ensino das disciplinas dos cursos ofertados pela Instituição de Ensino Superior (IES).

Observa-se uma emergencial necessidade de geração de indicadores que permitam aos profissionais, que atuam em bibliotecas, a elaboração de relatórios e estatísticas de uso dos livros virtuais para a criação de um cenário econômico sustentável pela IES e para dar suporte pedagógico na elaboração dos documentos educacionais essenciais para os funcionamentos dos cursos ofertados pela IES.

INTRODUÇÃO

A recuperação de informação (RI) trata da representação, armazenamento, organização e acesso a itens de informação, como documentos, páginas da Web, catálogos online, registros estruturados e semiestruturados, objetos multimídia. A representação e organização dos itens de informação devem ser tais que proporcionem aos usuários fácil acesso às informações de seu interesse

(BAEZA-YATES; RIBEIRO-NETO, 2011).

Nesse contexto de demandas da RI a Web semântica surgiu com o intuito de melhorar a organização da Web e, conseqüentemente, tornar os seus dados mais facilmente localizáveis inserindo semântica nesses dados com o objetivo que eles sejam melhores compreendidos pelas máquinas (BERNERS-LEE; HENDLER; LASSILA, 2001).

Problema de Pesquisa

Como se utilizar dos recursos da Web semântica para organizar os dados disponíveis nas bases de uma biblioteca a fim de fornecer relatórios e estatísticas que auxiliam a gestão da IES ao qual ela pertence?

Objetivo de Pesquisa

Mostrar o processo de mapeamento dos dados existentes para dados ligados na Web semântica dando ênfase ao estabelecimento de interoperabilidade nas camadas sintática, estrutural e semântica, e uma prova de conceito para a recuperação de informações a partir de uma linguagem de consulta específica de dados ligados.

O presente estudo usa abordagem qualitativa, natureza aplicada, sendo utilizado também procedimentos de estudo de caso.

Utilizou-se uma prova de conceito pela escrita de consultas de forma a obter um resultado composto de dados de outras bases interligadas por meio de metadados interoperáveis.

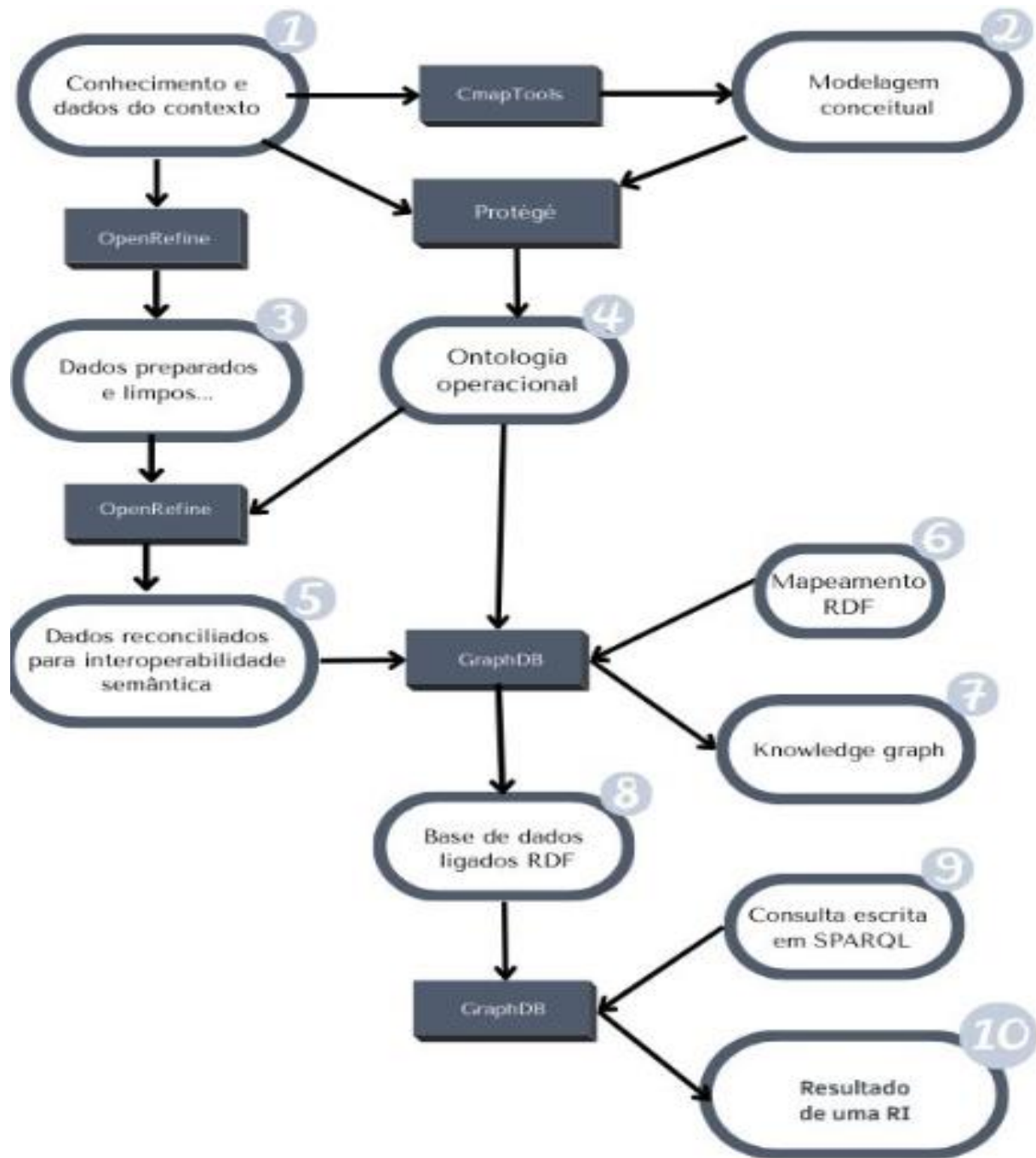
O contexto delimitado para a pesquisa está nos relatórios de acesso da biblioteca digital da FAESA - Centro Universitário Espírito Santense, sendo que os dados foram gerados por relatórios da Vital Source, provedora do acervo digital, referentes aos acessos do ano de 2021 no primeiro semestre (janeiro a junho 2021).

Procedimentos Metodológicos



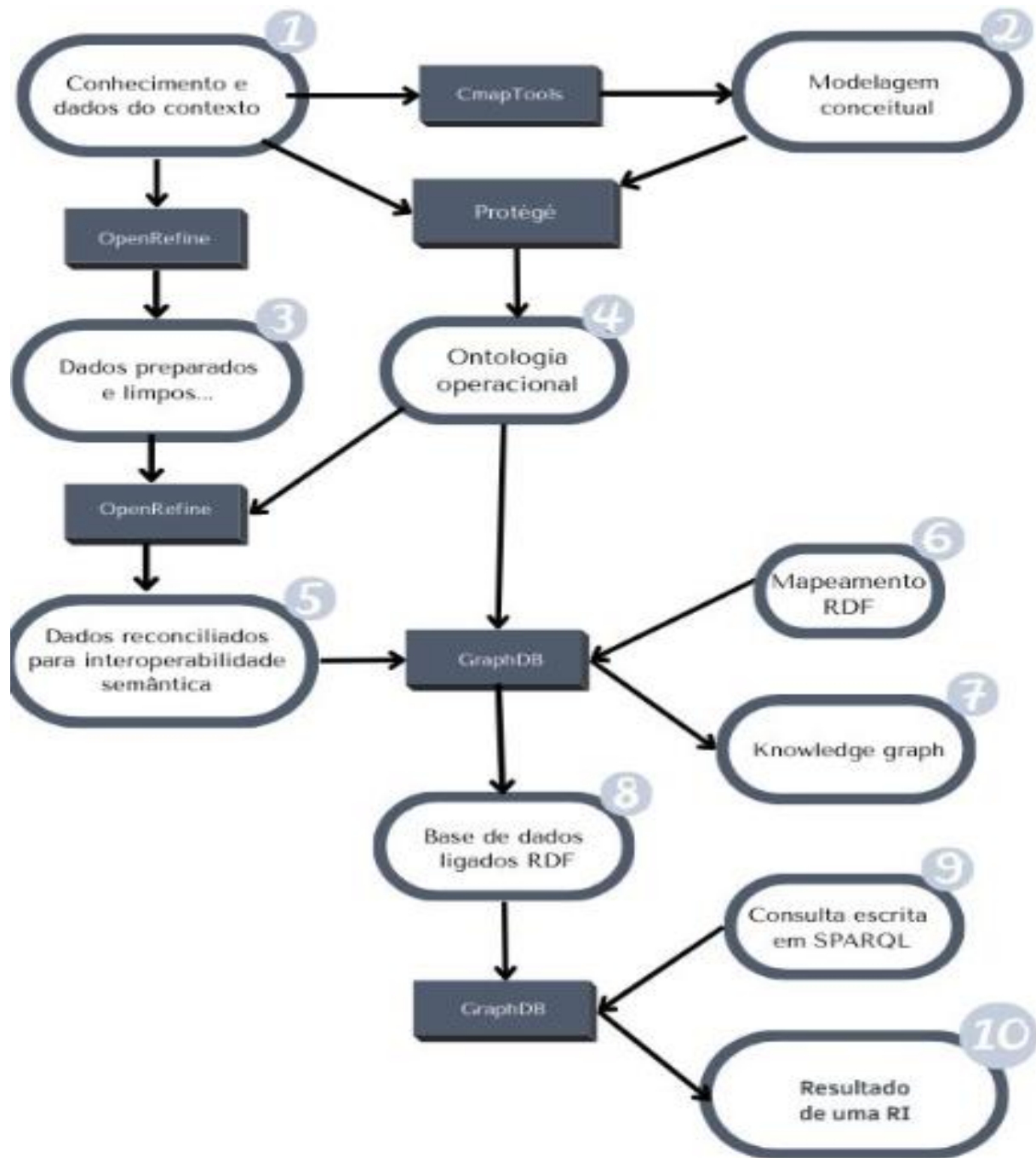
Composto por 10 etapas para a recuperação da informação através de construção de ontologias e os softwares necessários nessas etapas. Pode-se exemplificar, de maneira breve, a descrição de cada etapa da seguinte maneira:

Procedimentos Metodológicos



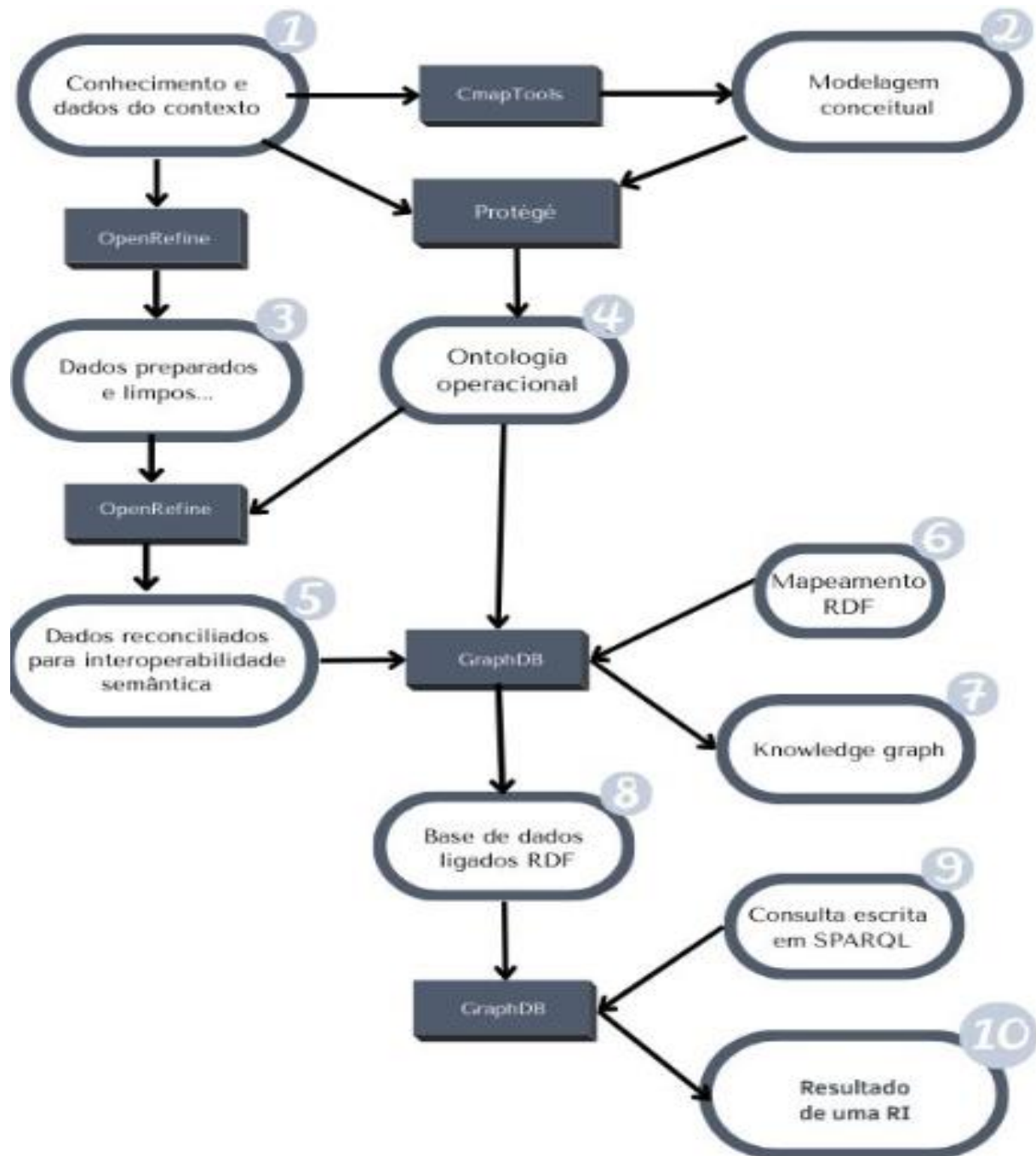
A PRIMEIRA etapa seria a definição, conhecimento e exploração inicial dos dados do contexto e do referencial necessário, realizando um tratamento e filtragem da base.

Procedimentos Metodológicos



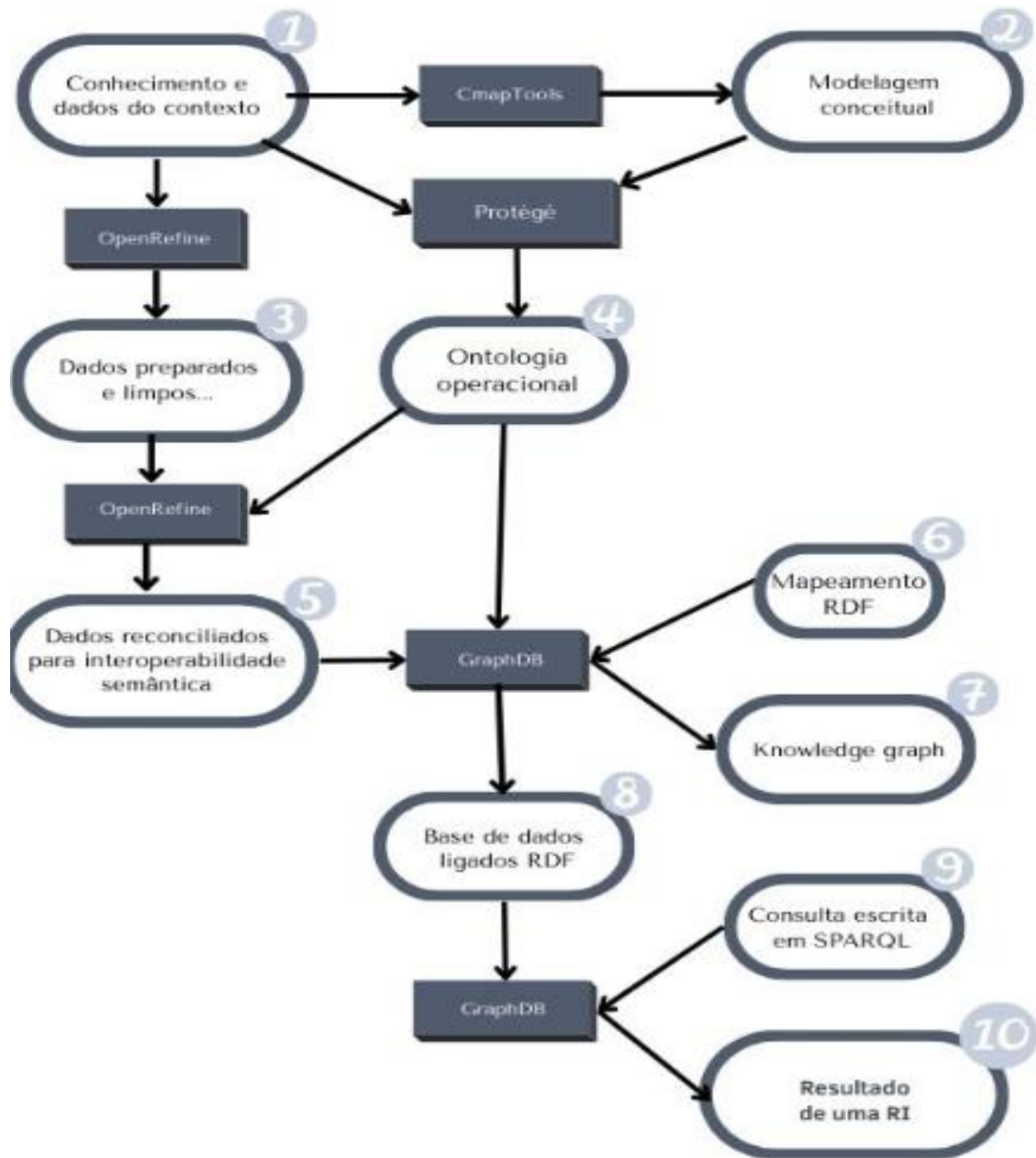
A **SEGUNDA** etapa passa por uma modelagem preliminar da ontologia, onde começa-se um esboço da conexão entre os dados trabalhados.

Procedimentos Metodológicos



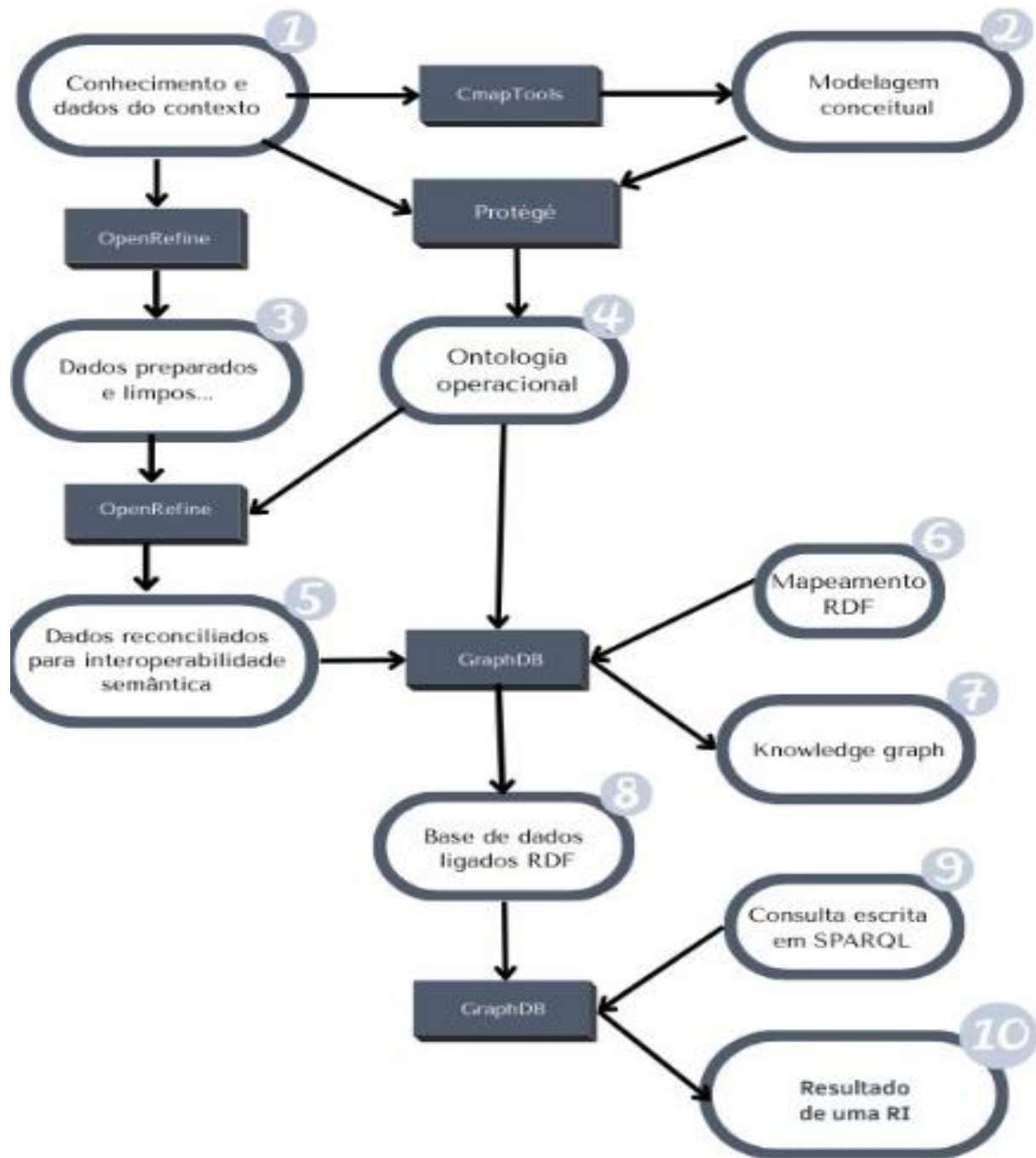
A TERCEIRA etapa é onde se dá a finalização do tratamento dos dados para melhor definição das instâncias da Ontologia.

Procedimentos Metodológicos



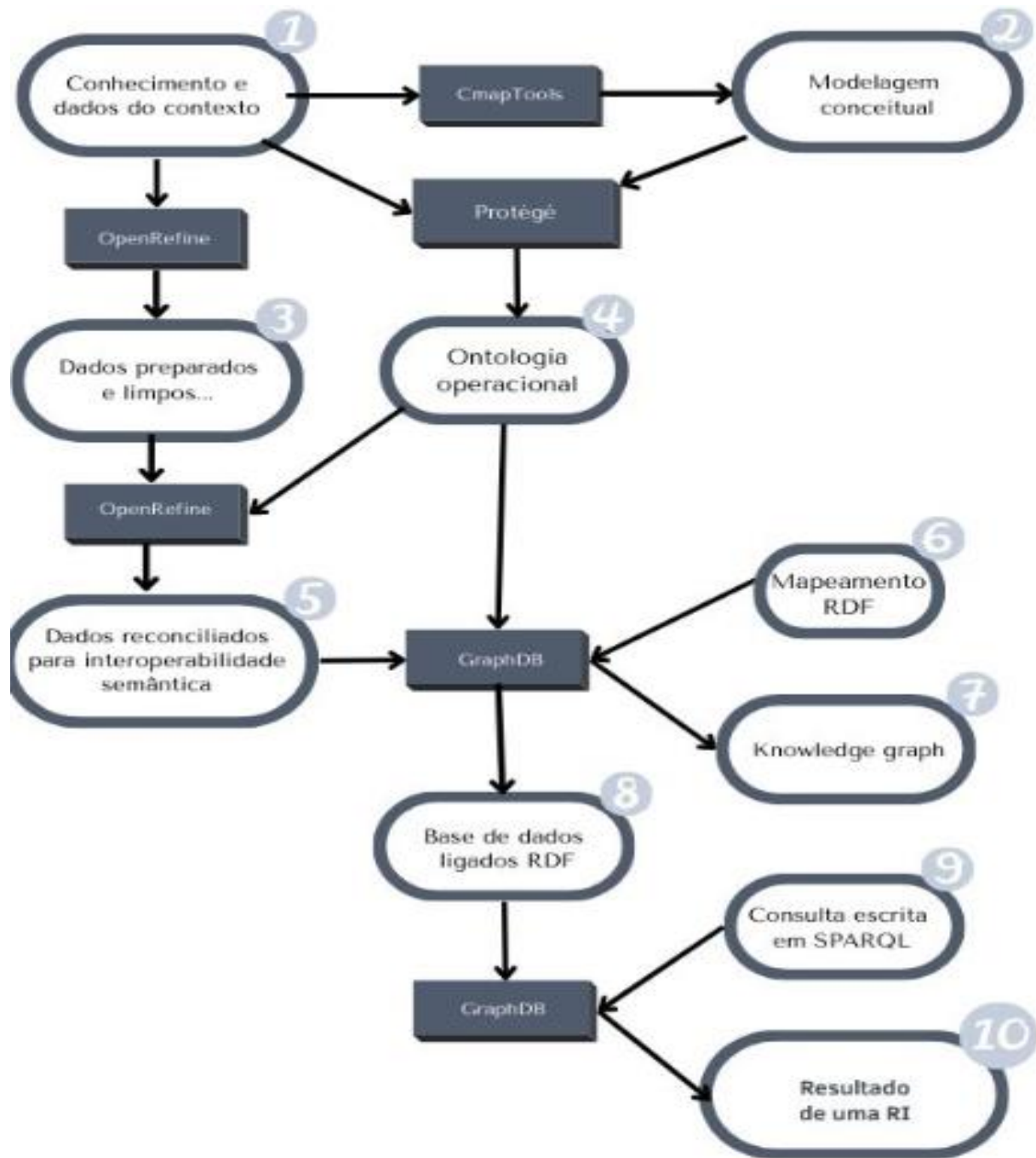
Na QUARTA etapa a Ontologia já pode ser melhor desenvolvida, com base nas escolhas feitas para seleção dos dados.

Procedimentos Metodológicos



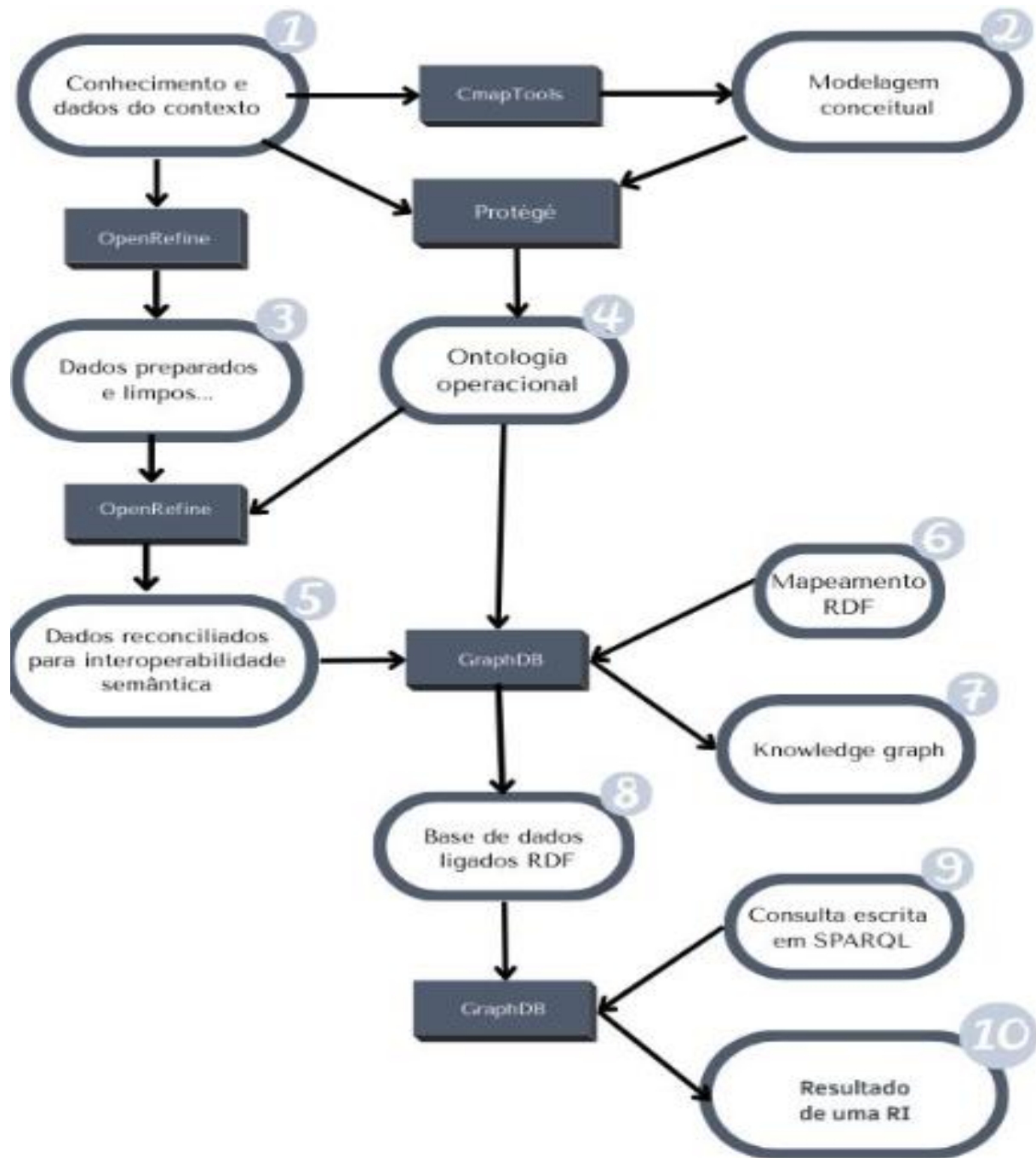
A QUINTA etapa é onde acontece a reconciliação dos dados para agregar interoperabilidade semântica.

Procedimentos Metodológicos



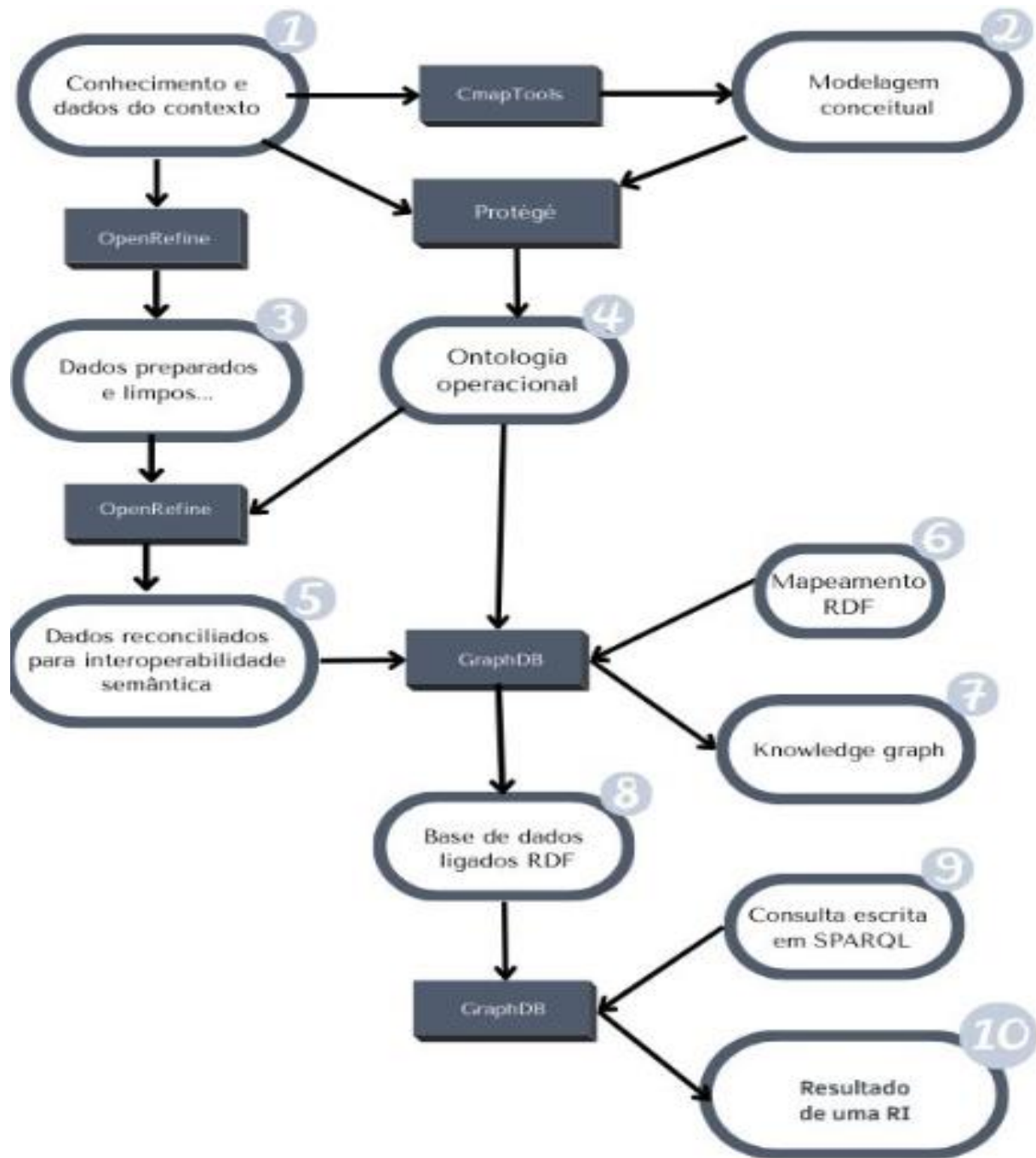
Na **SEXTA** etapa já é possível gerar um *Knowledge Graph* com a representação da ontologia construída.

Procedimentos Metodológicos



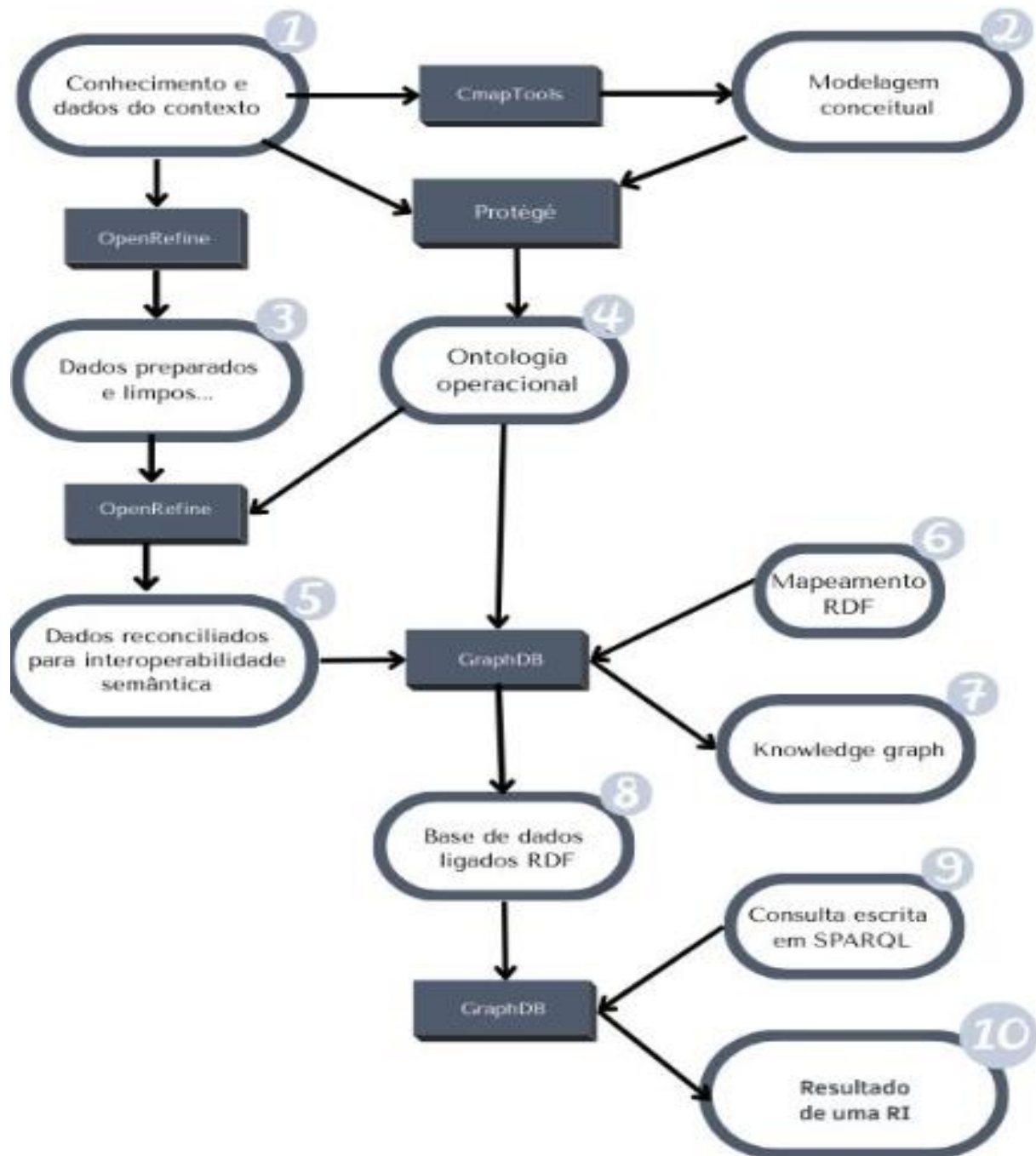
Na **SÉTIMA** etapa ocorre um mapeamento da base de dados ligados utilizando a ontologia criada como referência, possibilitando uma expansão das mesmas com todos os dados trabalhados.

Procedimentos Metodológicos



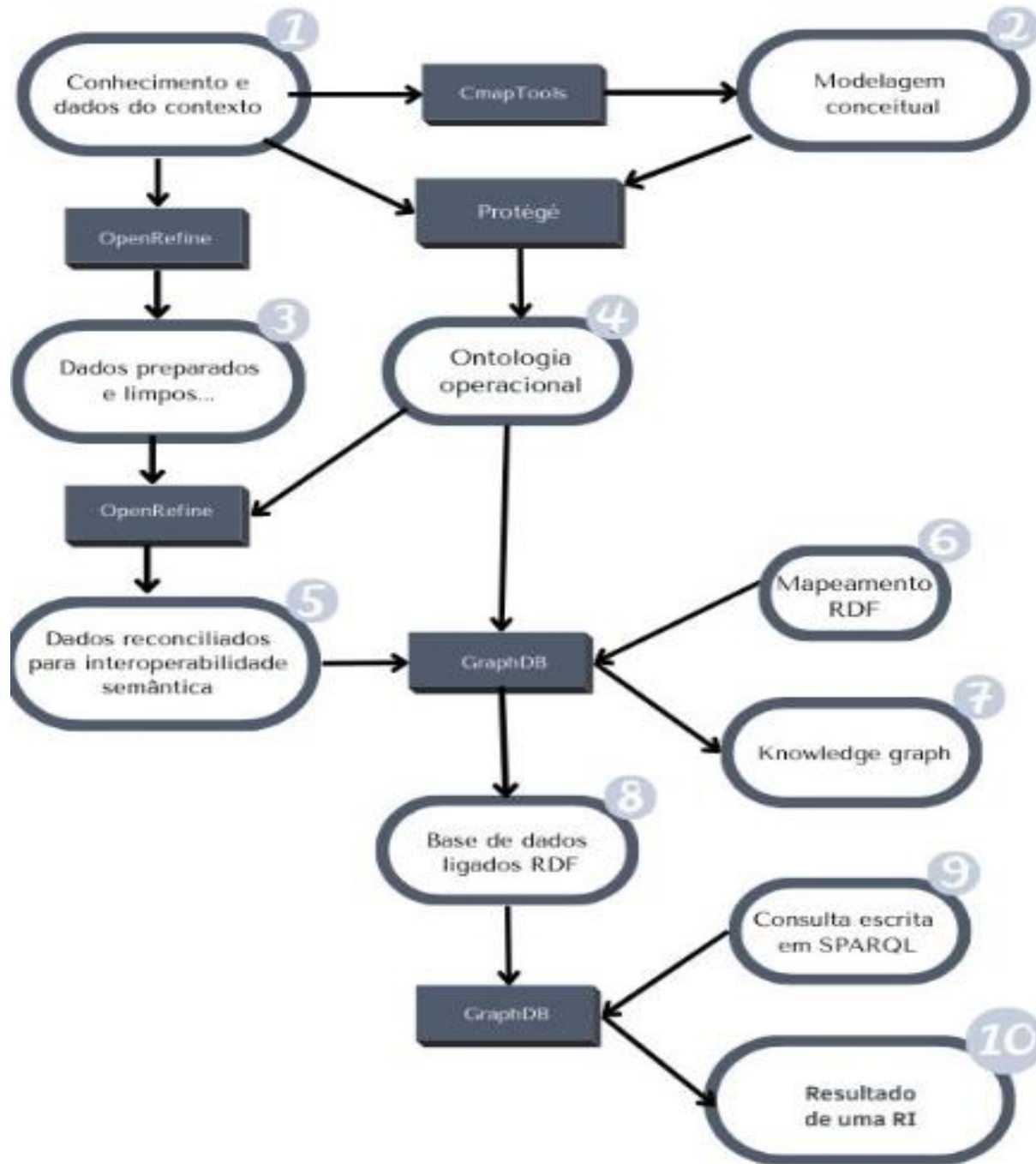
A OITAVA etapa é o momento em que a base de dados ligados é exportada em RDF.

Procedimentos Metodológicos



Na NONA etapa são escritas consultas em SPARQL para análise e teste da base construída, a fim de demonstrar os resultados que podem ser obtidos com base no trabalho construído.

Procedimentos Metodológicos



Na DÉCIMA, e última etapa, com base em todos os procedimentos anteriores, é onde se consolida o resultado de uma Recuperação da Informação.

Resultados



Hierarquia de classes da ontologia operacional

Construção da ontologia em RDF-Turtle ([Base da Ontologia](#)), no software Protégé, onde podem ser observadas as Classes, Indivíduos e algumas das propriedades que criam conexões entre os elementos da base.

CLASSE	EQUIVALENTCLASS
Author	https://www.wikidata.org/wiki/Q482980
Book	https://schema.org/Book
ISBN	https://id.loc.gov/ontologies/bibframe.html#c_isbn
Instituição Educacional	https://schema.org/EducationalOrganization
Person	https://schema.org/Person http://xmlns.com/foaf/spec/#term_Person
Publisher	http://mappings.dbpedia.org/server/ontology/pages/OntologyClass%3APublisher

Resultados

Conexões semânticas entre classes

A fim de agregar uma segunda camada de **interoperabilidade semântica**, não apenas conectada às classes, mas à cada um dos indivíduos, foi realizada uma **reconciliação dos dados**, possibilitando uma **conexão com diferentes bases**, onde foi possível encontrar **mais informações sobre elementos reconciliados**, desde mais informações sobre os livros através do ISBN, até a busca pelos autores e editoras.

Resultados

Knowledge graph para conexões de um livro

Foram consolidados também dois Knowledge Graphs com base de Ontologia inicial criada no Protégé, expondo as conexões resultantes de um livro.



Resultados

Editora Blucher X

```
1 PREFIX sioc: <http://rdfs.org/sioc/ns#>
2 PREFIX dc: <http://purl.org/dc/elements/1.1/>
3 PREFIX tto: <http://example.org/tuto/ontology#>
4 select * where {
5     ?Book sioc:num_views ?qtde .
6     ?Book dc:publisher tto:Editora%20Blucher
7 } limit 100
8
```

Table Raw Response Pivot Table Google Chart

Filter query results

	Book
1	tto:ABDI%20e%20APDINS%20-%20RJ
2	tto:Dez%20ensaios%20sobre%20memória%20gráfica
3	tto:Geografia%3A%20Coleção%20A%20Reflexão%20e%20a%20Prática%20no%20Ensino%20Médio
4	tto:Grafos%3A%20Introdução%20e%20prática
5	tto:Introdução%20aos%20processos%20de%20fabricação%20de%20produtos%20metálicos
6	tto:Kadila%3A%20culturas%20e%20ambientes%20-%20Diálogos%20Brasil-Angola

Resultado de uma consulta SPARQL

Foram realizadas 3 consultas SPARQL

diferentes, com base nas editoras da amostragem, utilizando-se da mesma

estrutura de busca, trocando somente

o nome da editora.



Verificação positiva da utilização de elementos da Web semântica para inclusão de dados dinâmicos nos catálogos das bibliotecas digitais.



As soluções de interoperabilidade evoluíram ao longo do tempo, possibilitando a heterogeneidade de acervos e a preservação da semântica dos conteúdos tornados interoperáveis.



Observou-se também que é possível a inclusão de vínculo entre dados de um mesmo catálogo, estabelecendo relações entre registros presentes no acervo e destes com bases de dados externos e de outras ferramentas da Web semântica em bibliotecas, com mostrado pelo knowledge graph da Figura 3.

Considerações Finais



Quanto ao objetivo proposto, a pesquisa conseguiu mostrar, para um pequeno extrato da base escolhida para o estudo de caso, o processo de mapeamento dos dados existentes para dados ligados na Web semântica dando ênfase ao estabelecimento de interoperabilidade nas camadas sintática, estrutural e semântica. Foi também realizada uma prova de conceito para a recuperação de informações a partir da linguagem SPARQL.



Utilizando-se o *workflow* apresentado, seria indicado contemplar a base de dados completa bem como elaborar uma ontologia de domínio que considere todo o cenário.

Considerações Finais

Considerações Finais

Dessa forma, seria possível fazer um acompanhamento estatístico de uso dos livros virtuais disponíveis em uma biblioteca virtual, de forma personalizada para cada objetivo estratégico estabelecido pelo bibliotecário gestor. Seria também necessário implementar uma interface para uso da linguagem de consulta SPARQL uma vez que ela não é apropriada para o usuário final.

Referências

- BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Recuperação de informação: conceitos e tecnologia das máquinas de busca**. 2. ed. Porto Alegre: Bookman, 2013.
- BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked data - the story so far. **International Journal on Semantic Web and Information Systems (IJSWIS)**, v. 5, n. 3, p. 1–22, 2009. Disponível em: <https://www.igi-global.com/article/linked-data-story-far/37496>. Acesso em: 1 mar. 2021.
- FALBO, Ricardo de Almeida. SABiO: Systematic approach for building ontologies. Em: 2014, Rio de Janeiro, RJ. **Anais [...]**. Em: 1st Joint Workshop ONTO.COM/ODISE on Ontologies in Conceptual Modeling and Information Systems Engineering co-located with 8th International Conference on Formal Ontology in Information Systems. Rio de Janeiro, RJ: CEUR Workshop Proceedings, 2014. p. 14. Disponível em: http://ceur-ws.org/Vol-1301/ontocomodise2014_2.pdf. Acesso em: 13 nov. 2021.
- HEATH, Tom; BIZER, Christian. Linked data: evolving the web into a global data space. **Synthesis Lectures on the Semantic Web: Theory and Technology**, v. 1, n. 1, p. 1–136, 2011. Disponível em: <https://www.morganclaypool.com/doi/abs/10.2200/S00334ED1V01Y201102WBE001>. Acesso em: 2 mar. 2021.
- NISO. **Understanding metadata**. Bethesda (EUA): National Information Standards Organization, 2004. Disponível em: [http://www.niso.org/publications/press/Understanding Metadata.pdf](http://www.niso.org/publications/press/Understanding%20Metadata.pdf). Acesso em: 13 nov. 2021.
- SERRA, Liliana Giusti. A web semântica na gestão de livros digitais licenciados: uma proposta de modelo. 2019. Tese (Doutorado em Ciência da Informação), Programa de Pós-Graduação em Ciência da Informação da UNESP, São Paulo, 2019. Disponível em: <https://repositorio.unesp.br/handle/11449/183526>. Acesso em: 02 nov. 2021.
- YIN, Robert K. **Estudo de caso: planejamento e métodos**. 3. ed. Porto Alegre, RS: Bookman, 2001
- ZENG, Marcia Lei. Interoperability. **Knowledge Organization**, v. 46, n. 2, p. 122–146, 2019. Disponível em: <http://www.isko.org/cyclo/interoperability>. Acesso em: 13 nov. 2021.
-