# 6
# Conclusão

O uso dos padrões da Web Semântica, como o RDF e RDFa, na publicação de informações na Web vêm demonstrando ser a única forma viável de garantir a interoperabilidade [34][53][80-83] de dados na Web. As soluções existentes, baseadas em padrões da Web Semântica, no entanto, ainda apresentam dificuldades no que tange a escalabilidade, extensão e apoio fornecido aos usuários, o que dificulta a adoção e disseminação da mesma.

## 6.1
## Contribuições

Neste cenário propomos Babel, um framework que permite que especialistas, bem como não-especialistas, a converter informações armazenadas em planilhas, bancos de dados relacionais, ou texto, para o formato RDF (N3 e N4), por meio de templates.

Babel possui uma máquina de processamento de templates que permite trabalhar com formatos da família XML (RDFa, HTML, XHTML, RDF), bem como triplas (N3 e N4). Babel é extensível, o que é particularmente interessante para garantir a compatibilidade com outros tipos de fontes de dados, tais como arquivos em formatos proprietários.

Embora os problemas da visualização e da persistência de Linked Data venham sendo estudados como problemas distintos [84], nosso estudo demonstrou que a utilização da heurística baseada em templates permite endereçar ambos.

Nosso trabalho também permite realizar o mapeamento, não apenas de fontes de dados tradicionais (banco de dados relacionais e planilhas), mas também permite extensões de forma a incluir novos formatos de dados.

A fim de facilitar a publicação dos dados por usuários não especialistas, foi desenvolvida uma interface gráfica que facilita a seleção e publicação da informação seguindo uma abordagem orientada a objetos, onde é permitido, dentre outras, a utilização de algoritmos que utilizam técnicas de mapeamento na

criação dos templates, além da publicação dos dados em repositórios de triplas bem difundidos, tais como o Virtuoso.

Também foi possível observar, através do Gráfico 3 (Capítulo 1), que a simples conversão da informação em RDF é feita de forma linear e não constitui o gargalo nos repositórios de triplas RDF.

Desta forma, a utilização de uma abordagem SPARQL, baseada na conversão e na tradução da consultas SPARQL para SQL, pode representar uma possível solução para o problema de desempenho. Neste caso, o fator limitador são técnicas eficientes que possibilitem a tradução de consultas SPARQL para SQL [12-13].

## 6.2
## Consulta de bases através de mapeamentos

Ainda que o framework ofereça uma interface Web para a publicação de informações, a adoção de protocolos e padrões tais como o protocolo SPARQL é um caminho importante na busca pela integração e padronização. Várias ferramentas já oferecem esse recurso, entretanto, muitas delas, não são compatíveis com as bases não-RDF dos usuários.

A utilização de templates abriu espaço não apenas para o mapeamento das bases para RDF, mas também para o caminho inverso, o que permite, em um futuro próximo, a identificação dos conjuntos de dados de uma base relacional em consultas SPARQL.

A Máquina de Processamento baseada em templates mostrou que é possível gerar conteúdo com base em uma Collection definida por uma cláusula restritiva. Um estudo mais aprofundado, utilizando as técnicas de conversão SPARQLtoSQL existentes, pode levar a uma solução que possibilite a utilização da base relacional original, como exemplificado na Figura 24.

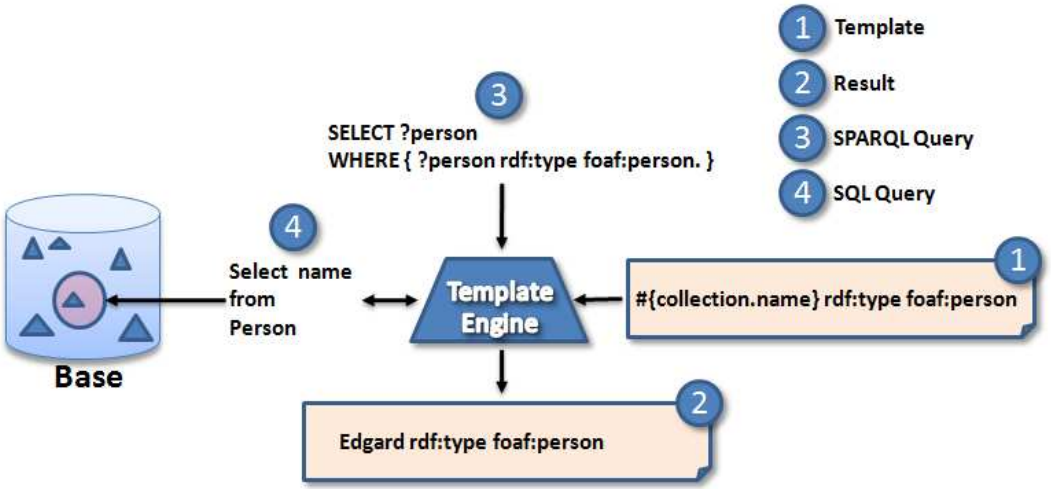


**Figura 24**. Ilustração da conversão de uma base para RDF e de uma consulta SPARQL para uma consulta SQL.

## 6.3 Mashups

Em geral, um Mashup é um aplicação Web que consome dados provenientes de outros sistemas, recuperados através de APIs como SPARQL e Webservices, dentre outras. Devido aos diferentes vocabulários utilizados pelos diferentes conjuntos de dados, é difícil encontrar relações e concatenar essas informações. Tim Beners Lee, afirma que o objetivo da Web Semântica é fazer da Web uma uma rede de dados global, interligada e acessível [85]. A possibilidade de realizar Mashups dessas informações é o primeiro passo nessa direção. Para interligar e disponibilizar uma enorme quantidade de dados estruturados na Web, as pessoas devem ser capazes de consultar e interligar os dados de uma maneira eficiente. Há poucos padrões a este respeito, e pode ser muito difícil entender o contexto, definições, e o histórico dos dados que irão ser trabalhados [86].

É exatamente neste ponto que a ferramenta proposta pode ser útil. Babel permite acessar multiplas fontes de dados através de uma interface padrão, não só isso, as informações extraídas dessas fontes de dados podem ser interligadas através em um só vocabulário, como demonstrado na Figura 25. Dessa forma, por exemplo, dados pronvidos do IBGE poderiam ser interligados e convertidos para o mesmo vocabulário utilizado por Data.gov[41], ou ainda, converter os dados do

[41] Sitio de dados abertos oficial do governo dos Estados Unidos

IBGE e Data.gov para um terceiro vocabulário utilizado, por exemplo, pela DBPedia. O que possibilitaria o *mashup* e o processamento dessas informações, mesmo estando armazenadas em bases distintas e organizadas em estruturas diferentes.

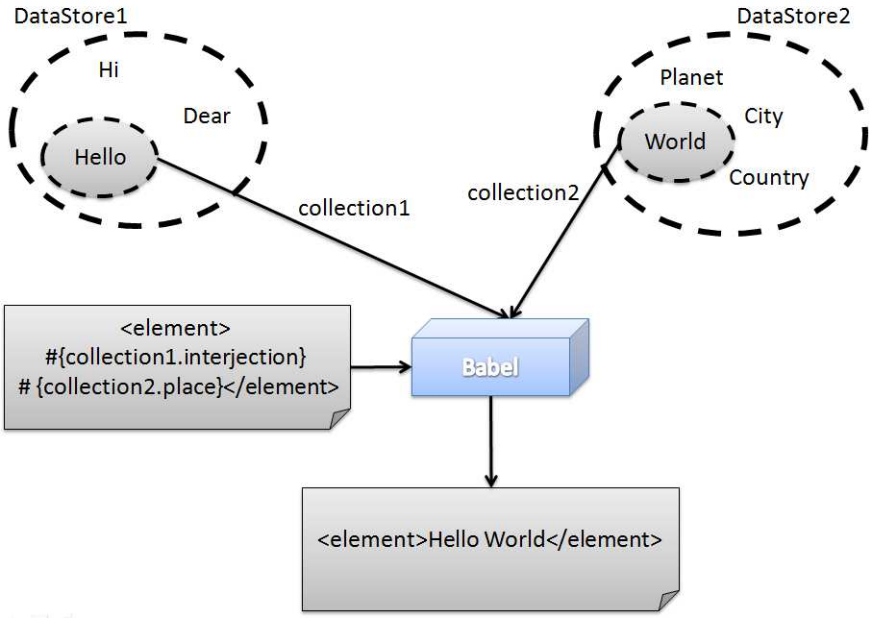


**Figura 25.** Figura ilustrativa de um Mashup realizado a partir de duas bases de dados utilizando Babel.

## 6.4
## Trabalhos Futuros

Apesar dos resultados obtidos, ainda podemos fazer muito para melhorar a qualidade e usabilidade do Framework, e de seus diversos componentes. Embora seja possível adicionar novos DataStores de maneira não intrusiva, o processamento de novas estruturas de templates só é possível mediante uma alteração no código da Máquina de Processamento. Para tornar a ferramenta mais flexível seria necessário que as máquinas de processamento também fossem implementadas como plugins.

Outro aspecto importante, diz respeito ao suporte integral das regras estabelecidas pela TML, que possibilitará o uso de estruturas mais complezas como templates. A versão atual da Máquina de Processamento não permite a utilização da segunda regra.

Também não é possível utilizar mais de uma chave identificadora (CollectionID) para o relacionamento entre as coleções, um avanço importante seria possibilitar o estabelecimento de mais de uma chave entre o relacionamento das coleções, permitindo assim, o uso de chaves primárias compostas.

Por mais que a Interface Gráfica atual tenha contribuído para que usuários pouco familiarizados com as tecnologias de Web Semântica pudessem definir Collections, criar templates e arquivos de configuração há vários pontos que ainda podem ser melhorados:

- Em linhas gerais, algumas funcionalidades podem ser desenvolvidas de forma extensível, como:
    - A interface de seleção das Collections: a interface de seleção das coleções deverá sofrer alterações a cada nova implementação de DataStore adicionada. Atualmente a adição de novos DataStores só é permitida mediante intervenção manual no código;
    - Estratégias de mapeamento: com o passar do tempo, as estratégias de mapeamento poderão mudar. Estratégias poderão surgir, tornarem-se obsoletas ou simplesmente serem atualizadas. Seria importante que essas estratégias fossem isoladas em um módulo diferente do principal.
    - Repositórios de triplas: existem vários repositórios de triplas que variam quanto ao desempenho, escalabilidade e usabilidade, esses fatores podem levar um usuário a optar pela utilização de um determinado repositório a outro, facilitar a adição de novos repositórios promove uma maior versatilidade da ferramenta.
- A adição de outras técnicas de mapeamento, além do mapeamento direto, como, por exemplo, a utilização do Vocabulário RDF de Dados em Cubos[42] (*RDF Data Cube Vocabulary*) desenvolvido para mapeamento de dados estáticos como arquivos CSV´s e planilhas.

[42] http://publishing-statistical-data.googlecode.com/svn/trunk/specs/src/main/html/cube.html#dsd-example

- A seleção de outras coleções que não somente à fonte de banco de dados relacionais.
- Facilitar a criação do template através de uma interface orientada a objetos que possibilite que usuários não familiarizados com as estruturas pudessem criá-las de uma forma mais natural. Nesse sentido, é possível encontrar várias implementações que já foram propostas [51][72].
- Introduzir algoritmos que selecionem o vocabulário de uma forma automática ou semi-automática, com base no esquema da fonte de dados, o que reduziria a lentidão e a necessidade de intervenções manuais no processo de publicação, que permitiria a usuários desconhecedores do domínio a utilizar a ferramenta.