

1. **Título:**

CLIMA IN – Um Modelo INteligente para o Projeto CLIMA (Cloud Lariisa Mushup Aplicados)

2. **Resumo:**

O GISSA (Governança Inteligente de Sistemas de Saúde) é um sistema de inovação tecnológica destinado ao programa Rede Cegonha do Ministério da Saúde, financiado pela FINEP, em desenvolvimento pelo Instituto Atlântico com a parceria do IFCE. O projeto CLIMA IN propõe um modelo integrado dos mecanismos inteligentes para o projeto GISSA. Ele reúne esforços das equipes de pesquisa do IFCE na área de ontologia / mineração de dados e da equipe da UFC na área de integração de dados com linked data / mushups. O resultado será a definição e implementação do modelo de inteligência baseado em inferência por regras, busca semântica e mineração de dados. A experimentação deste modelo faz uso inédito de uma metodologia em mineração de dados criada no IFCE e de nova plataforma de interoperabilidade semântica criada pela equipe da UFC/PUC-Rio. O CLIMA IN é uma continuidade dos projetos CLIMA (2014) e Point G CLIMA (2015), suportados pela PRPI/IFCE com a chancela da FUNCAP/CNPq.

Palavras-chave: Sistemas inteligentes, ontologia, mineração de dados, TI em saúde.

3. **Introdução / Justificativa**

O GISSA (Governança Inteligente dos Sistemas de Saúde) é um framework para construção de sistemas de informação para apoio ao processo de tomada de decisão, em diversos níveis de seus atores (secretário, agente de saúde, paciente, etc.) no contexto da saúde pública [1]. Trata-se de um sistema de inovação tecnológica destinado ao programa Rede Cegonha do Ministério da Saúde, financiado pela FINEP, em desenvolvimento pelo Instituto Atlântico com a parceria do IFCE. Uma prova de Conceito (PoC) do framework GISSA está sendo implementada no município de Tauá (Ce). Para tanto, o GISSA é formado por um conjunto de componentes que permitem a coleta, integração, transformação, inferência e visualização de informações de forma a fornecer aos usuários finais fatos e dados necessários às diversas decisões relacionadas a saúde pública.

A. Integração de Sistemas de Saúde

O tema Integração de Sistema de Saúde está presente em todo o processo discursivo, ocorrido na década de 70, que fundamentou a criação do Movimento da Reforma Sanitária Brasileira e subsequente criação do Sistema Único de Saúde – SUS [2]. Marcou profundamente a própria denominação deste sistema na década de 80 quando este foi firmado na Constituição Federal Brasileira. A década de 90 foi marcada pelo processo de municipalização e o esforço de construir um sistema integrado em um país federativo. Os serviços de saúde de sistemas integrados, como é o caso do brasileiro, devem ser organizados em redes e necessitam serem operacionalizados de maneira integrada.

Um sistema de saúde para funcionar como se fosse um único sistema (interligado, integrado, interdependente) necessita manter uma integração fundada em cinco pontos: integração sistêmica; integração normativa; integração funcional; integração clínica; e integração de cuidados, conforme mostra a figura 1 [3].

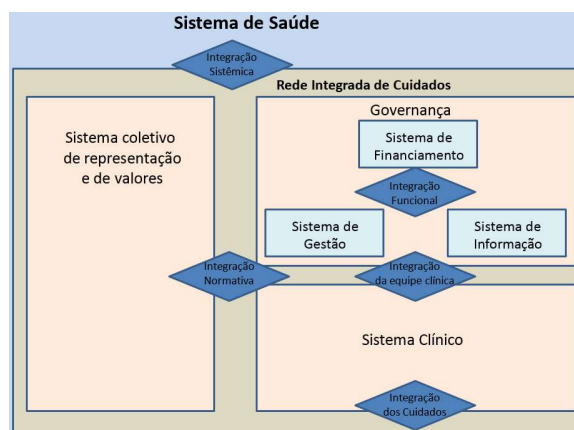


Figura 1 – Dimensões de um sistema integrado de saúde.

O sistema de saúde atual é constituído pelos níveis primário, secundário e terciário, categorizados conforme a complexidade de suas atividades.

- O nível primário é voltado para a atenção básica, compreende atividades de promoção, proteção e recuperação, em nível ambulatorial desenvolvidas por pessoal elementar ou médio, médicos generalistas e odontólogos.
- O nível secundário, além das atividades do nível primário, engloba também atividades assistenciais nas quatro especialidades médicas básicas: clínica médica, clínica cirúrgica, gineco-obstetrícia e pediatria e, especialidades estratégicas nas modalidades de atenção ambulatorial, internação, urgência e reabilitação.
- O nível terciário é o de maior capacidade resolutiva dos casos mais complexos do sistema, nas modalidades de atendimento ambulatorial, internação e de urgência. Os estabelecimentos deste nível são os ambulatórios de especialidades, hospitais especializados e de especialidades.

Desta forma, o GISSA promoverá a integração sistêmica, o que pressupõe um alinhamento de todo processo de pensar e executar o Sistema de Saúde; suportará a integração normativa, permitindo que se expressem de forma homogênea os valores da sociedade, das organizações e dos atores envolvidos; induzirá a integração funcional, permitindo a manifestação do conjunto de elementos presentes nos serviços de saúde, referentes ao suporte operacional ou executivo; articulará a integração clínica de equipes multidisciplinares, fornecendo mecanismos que mobilizam a diversidade de competências e conhecimentos e coordenará a integração de cuidados de saúde, organizando de forma sustentável as práticas clínicas em torno de problemas de saúde específicos de cada paciente [1][3].

B. Escopo do Modelo do Sistema Integrado Inteligência Saúde no GISSA

Estruturalmente, o GISSA será constituído por um conjunto de componentes habilitados para coletar informações em tempo real de contextos inerentes às situações monitoradas de saúde e alimentar mecanismos inteligentes capazes de produzir informações qualificadas, as quais municiam diversas aplicações voltadas para os mais variados atores envolvidos: usuários finais, profissionais de saúde, gestores tomadores de decisão do sistema de saúde [1].

O projeto envolverá áreas de fronteira do conhecimento técnico-científico, em dois aspectos: de um lado, no campo da saúde coletiva, o planejamento e gestão em saúde pública e saúde da família; de outro lado, no campo da tecnologia da informação, os mecanismos de inferência, baseados em técnicas de inteligência artificial e modelos de ontologias. Ele será constituído por um conjunto de componentes habilitados para coletar informações em tempo real de contextos inerentes às situações monitoradas de saúde e alimentar mecanismos inteligentes capazes de produzir informações qualificadas, as quais municiam diversas aplicações voltadas para os mais variados atores tomadores de decisão do sistema de saúde.

C. Funcionalidades

O framework GISSA prevê cinco principais classes de funcionalidades [4][5]:

Relatórios listam um conjunto de eventos, pessoas ou itens e suas características relevantes, como, por exemplo: óbitos, nascimentos, gestantes, crianças, etc. Relatórios são as formas de apresentação de informações mais granulares do GISSA. Sua função é fornecer aos usuários a capacidade de realizar análises específicas sobre itens em particular. Para exemplificar um cenário de uso da função Relatório, considere-se uma listagem das gestantes de risco, disponibilizada para um secretário de saúde municipal. Com ela, um gestor de saúde, por exemplo, é capaz de acessar as características de uma gestante de risco em particular, permitindo-o entender os fatos que levaram o sistema a classificá-la como tal. Com esse conhecimento, o gestor de saúde pode solicitar ao agente de saúde responsável que realize, por exemplo, um acompanhamento mais próximo de determinado paciente.

Indicadores fornecem informação agregada sobre um determinado fato. Podem ser taxas (ex: taxa de mortalidade infantil), porcentagens (ex: cobertura vacinal de sarampo), totalizadores (ex: número de gestantes com data provável de parto para o próximo mês), gráficos (ex: um gráfico de barras com as divisões dos óbitos por causas básicas), entre outros. Ao mostrar informações agregadas de forma simples e direta, e por possuírem parâmetros definidos, os indicadores permitem identificar rapidamente se existe ou não um problema. Como cenário do uso dessa classe de funcionalidade, considere um indicador de taxa de partos cesáreos no último mês, disponibilizado para um secretário de saúde. Como existe um incentivo governamental para partos vaginais, uma taxa de 80% de partos cesáreos mostra, por exemplo, que o município não está cumprindo sua meta, evidenciando a necessidade de ações de retificação do problema.

Dashboards são funcionalidades de apresentação de informações onde se combinam múltiplos indicadores e relatórios, permitindo a construção de painéis que mostrem, de forma agregada, a situação de determinado evento. No GISSA, os dashboards não provêm novas informações, e sim apresentam de forma simultânea as funcionalidades de relatórios e indicadores. O GISSA prevê pelo menos dois níveis de dashboard. O primeiro nível orientado para os domínios de gestão pública de saúde: clínico-epidemiológico, técnico-administrativo, normativo, gestão compartilhada e gestão do conhecimento. E um segundo, orientado para determinados tópicos dentro de cada domínio. Exemplo de tópicos no domínio clínico-epidemiológico são: pré-natal, puericultura, nascimento e óbito.

Alertas são as funcionalidades orientadas a geração de mensagens para informar aos usuários da ocorrência de um evento em particular (ex: nascimento da pessoa A, entrada no hospital da pessoa B), ou da configuração de uma determinada situação (ex: estoque baixo do medicamento X, cobertura vacinal baixa). Alertas podem ser gerados automaticamente pelo sistema ou a partir de entradas manuais de dados.

Alertas também são enviados para diferentes tipos de usuários de acordo com sua gravidade. Por exemplo, um alerta de óbito infantil acidental é enviado apenas para agentes de saúde, enquanto um causado por desnutrição também é enviado para o governador. O objetivo desse tipo de funcionalidade é agilizar a disponibilização de informação aos atores da saúde pública, chamando atenção para necessidade de ações em particular.

Busca visa auxiliar os usuários na recuperação de artefatos que contém informação não estruturada, isto é, texto, imagens e vídeos. As buscas são voltadas principalmente para três dos cinco domínios do GISSA:

- normativo, no qual se deseja recuperar conhecimentos contidos em portarias, leis, decretos e resoluções;
- gestão compartilhada, no qual se deseja recuperar documentos produzidos nas comissões intergestores, como atas e deliberações;
- gestão do conhecimento, no qual a recuperação é orientada para protocolos de saúde, os quais contém conhecimentos clínicos consolidados para orientar procedimentos médicos em situações específicas, como pré-natal de alto risco.

D. Componentes do framework GISSA [4][5]:

Para dar suporte ao desenvolvimento das funcionalidades descritas o GISSA possui quatro grandes componentes: coleta, persistência, inteligência e visualização. O agrupamento das atividades do GISSA nesses componentes é representado na figura 2. Nela é mostrada a visão lógica do framework, destacando os componentes do framework de gestão do conhecimento (Inteligência) em azul, objeto do propósito do projeto CLIMA IN.

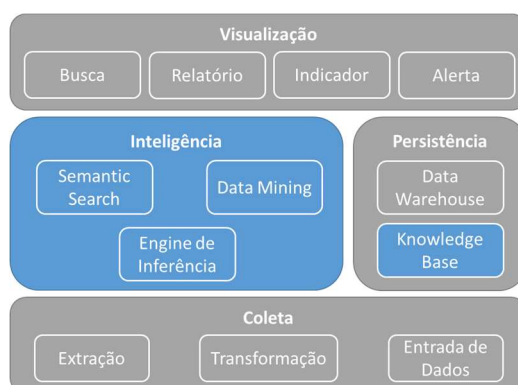


Figura 2. Visão lógica do GISSA framework, com ênfase no framework de GC

- **Coleta:** componente responsável por recuperar dados das diversas fontes produtoras e transformá-los no formato do GISSA. O componente auxilia a extração de dados das mais diversas fontes que incluem: sistemas locais, como e-SUS e SINASC; sistemas web, como SISPRENATAL e HORUS; barramento de web services, como o CADSUS; e fontes de dados públicas disponíveis na internet, como o repositório do bolsa família. O componente inclui scripts de ETL genéricos para a transformação dos dados para o formato padrão do GISSA.

- **Visualização:** componente responsável por prover bibliotecas para a construção das interfaces com usuário, sendo otimizado para a construção das cinco classes de funcionalidades descritos na seção anterior. São exemplos de partes desse bloco: componente de dashboard genérico, componentes de formulário (ex: date picker), componente de fachada de serviço REST.
- **Inteligência:** parte integrante do framework de gestão do conhecimento, o componente de inteligência contém mecanismos de geração de novos conhecimentos, como o engine de inferência e o componente de mineração de dados, além do engine de busca semântica, que permite o uso do conhecimento formalizado na recuperação dos artefatos com informação não estruturada.
- **Persistência:** componente de armazenamento de informações que contém os diversos repositórios de dados utilizados pelo GISSA, como o Datawarehouse, a Knowledge Base e a Base para as entradas de dados auxiliares.

Para o framework de Gestão de Conhecimento deverão ser explicitamente definidos quais os relatórios, alertas, consultas e indicadores que serão baseados em no conceito de inteligência computacional. Devem também estar definidas as tecnologias a serem usadas.

Assim, ao modelo de inteligência do GISSA, através de uma saída refinada, é incumbida a tarefa de aplicar Mineração de Dados e Inferência das Informações com o intuito de:

- Gerar de Alarmes relacionados à Morte Materna e Morte Infantil. A especificação de quais alertas farão parte da engine de inteligência tem como base os requisitos que serão implementados e priorizados no backlog.
- Realizar de Buscas Textuais e Mineração de Textos.
- Possuir algum tipo de feedback, característica imprescindível para um sistema inteligente
- Possuir a capacidade de adaptação

E. O GISSA como um Sistema Integrado de Inteligência [4][5]:

Data mining pode ser considerado um processo de extração de potenciais conhecimentos (reconhecimento de padrões, por ex.) úteis a um determinado universo de interesse (tomada de decisão em sistemas de saúde, por ex.) a partir de variáveis ou campos de bases com uma enorme quantidade (eSUS, SINASC, por ex.) de dados (estruturados ou brutos). Ontologia, é classicamente definida como uma especificação formal (entendida pelo computador) de uma conceitualização (modelo abstrato de algo real) explícita (pode ser compartilhada) sobre um domínio de interesse. Ela se apresenta, portanto, como um mecanismo eficiente na estruturação do conhecimento relativo a estes dados a serem tratados pelas ferramentas de Data Mining.

A figura 3, extraída de [6], mostra o Data Mining na melhoria do processo de armazenamento e integração de dados do Data Warehouse onde é proposto que Ontologia seja usada para agregar valor na captura de conhecimento compartilhado de interesse relativo aos dados existentes no Data Warehouse. Um exemplo de um serviço agregado seria encontrar páginas com expressões sintaticamente diferentes, mas com a mesma semanticamente. Esta é a base da web semântica [7] [8].

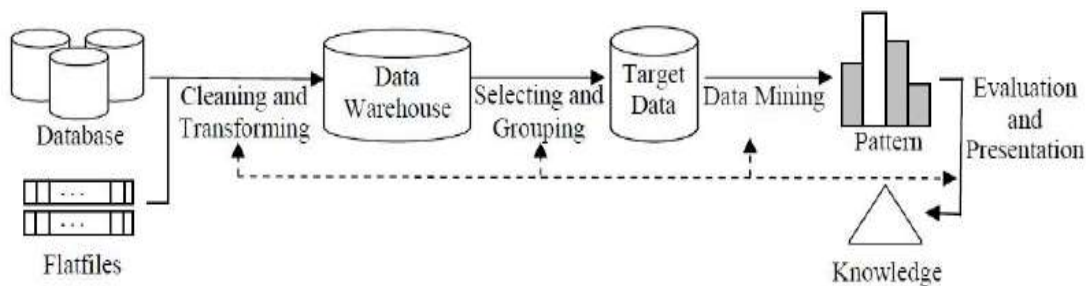


Fig 3: Processo inteligente de Data Mining apoiado por Ontologias

O resultado é um Data Warehouse inteligente centrado na ontologia do usuário, o que pode ajudá-lo a construir modelos que evitem a geração de padrões ineficientes na tarefa de predição do Data Mining. Como mostra a figura 3, este novo processo apoiado por ontologias pode estender novas regras produzindo um mecanismo ativo para a redescoberta de conhecimentos.

Em [6] são apresentadas duas formas de relacionamento de Ontologias com Data Mining:

- Ontologias para Data Mining: agrega conhecimento no processo Data Warehouse/Data Mining com o uso de Ontologias, como explicado anteriormente.
- Data Mining para Ontologias: incluindo conhecimento do domínio logo na entrada do sistema a ser modelado com Ontologias. Neste caso, a análise final é feita sobre as Ontologias (Knowledge Mining).
- A segunda abordagem acima é constituída de duas etapas:
- A construção do bloco Data Mining, que envolve a preparação de dados, seleção e extração do conhecimento.
- A construção do bloco Ontologia que terá como entrada o conhecimento extraído do bloco Data Mining, como mostrado na figura 4, extraída de [6].

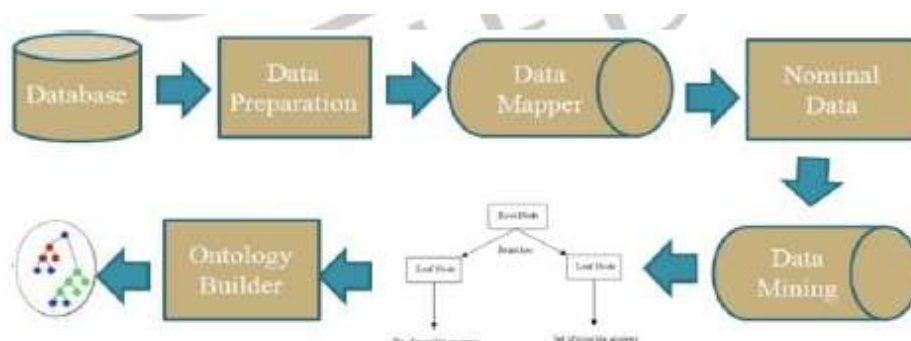


Fig 4: Construção da Ontologia a partir do Data Mining

4. OBJETIVOS GERAL & ESPECÍFICOS:

O projeto CLIMA IN tem como objetivo especificar e implementar, a partir de um estudo analítico e experimental das tecnologias de Data Mining, Data Warehouse, Ontologias e Mushups, o modelo integrado de inteligência do GISSA permitindo ao framework um apoio mais eficiente à tomada de decisão em sistemas de saúde para o programa Rede Cegonha do Governo Federal. Uma prova de conceito do GISSA será implementada no município de Tauá (Ce). Além da definição e implementação do modelo de inteligência do GISSA, o CLIMA IN fará uso inédito de um framework de integração de dados com interoperabilidade semântica baseado em linked data.

Ao final, três focos terão norteado os objetivos a serem alcançados no projeto:

- A definição e implementação do modelo de inteligência baseado em inferência por regras, busca semântica e mineração de dados a partir das ideias apresentadas na figura 3, propostas em [6].
- A validação da ferramenta de integração de dados com interoperabilidade semântica baseado em linked data, recém desenvolvida por pesquisadores da UFC e PUC-Rio, permitindo a disponibilidade pública das ontologias desenvolvidas no GISSA.
- A elaboração de estratégias para harmonizar a proposta de integração de dados baseada em Linked Data na definição do modelo de inteligência do GISSA baseado em Mineração de Dados e Ontologia.

Objetivos específicos:

- Análise do projeto GISSA em seus aspectos conceituais, funcionais, arquiteturais e tecnológicos [2][3][8];
- Análise dos cenários de aplicação da Rede Cegonha, bem como das bases de dados de saúde relacionadas [1];
- Análise de ontologias já desenvolvidas no GISSA para descrição de dados governamentais para saúde num cenário de integração de dados;
- Estudo das tecnologias inteligentes (Mineração de Dados e Ontologias) capazes de dotar o GISSA de mecanismos de inferência por regras, busca semântica e mineração de dados [11];
- Uso da plataforma desenvolvida pela equipe da UFC/PUC-Rio para a integração de dados utilizando as ontologias desenvolvidas no projeto GISSA;
- Avaliação da compatibilidade do modelo CLIMA IN desenvolvido com as funcionalidades e requisitos arquiteturais do GISSA.
- Elaboração de estratégias de harmonização da integração de dados com o modelo de inteligência do GISSA [11][12].

5. REVISÃO BIBLIOGRÁFICA

A. Glossário Resumido

- **Framework:** conjunto de componentes de software e/ou processos combinados para a solução de uma classe de problemas. Um framework fornece pontos de extensão para a construção de soluções customizadas para problemas específicos. O framework possui uma arquitetura que define as relações entre seus componentes, a forma de utilizá-los e como implantar a solução final.
- **Gestão do Conhecimento (GC):** processo cíclico composto principalmente por atividades de criação, integração e disseminação do conhecimento em uma organização (ou um conjunto delas).
- **Ontologia:** modelo que formaliza de forma explícita e inequívoca a visão de mundo de uma comunidade sobre um domínio de interesse. Uma ontologia define um conjunto de conceitos, suas relações e propriedades [9].
- **Inferência:** conclusão de novas informações a partir de informações já existentes. Inferências podem ser feitas por meio de raciocínio automatizado, utilizando um conjunto de regras formalizado em uma ontologia, ou por meio de algoritmos de mineração de dados e machine learning [10].
- **Data Mining:** conjunto de ferramentas, técnicas e metodologia baseadas em algoritmos de aprendizagem ou em classificação (redes neurais e estatística) capazes de explorar um conjunto de dados, extraíndo ou ajudando a evidenciar padrões nestes dados e auxiliando na descoberta de conhecimento [11][12].
- **Business Intelligence:** refere-se ao processo de coleta de grande volume de dados, organização, análise, compartilhamento e monitoramento de informações (normalmente contidas em um Data Warehouse/Data Mart), analisando-os e desenvolvendo percepções e entendimentos a seu respeito, tomada de decisão.
- **Data Warehouse:** possibilita a análise de grandes volumes de dados, coletados dos sistemas transacionais (OLTP). São as chamadas séries históricas que possibilitam uma melhor análise de eventos passados, oferecendo suporte às tomadas de decisões presentes e a previsão de eventos futuros [6][7].

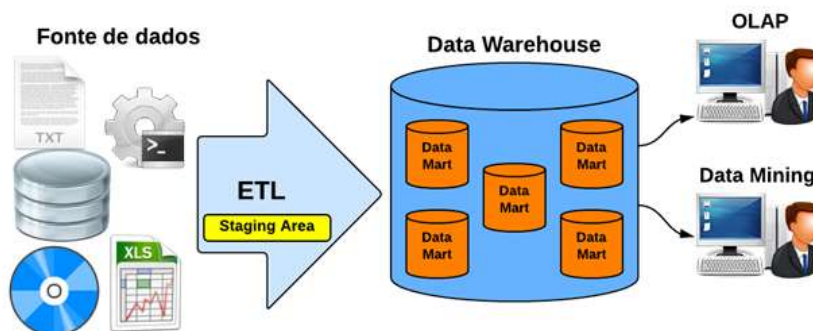


Figura 5. Fluxo de dados: OLTP/ ETL/ Datawarehouse/ Data Mining-OLAP

- **Tomada de Decisão:** conjunto de elementos que envolvem gestão do conhecimento, mecanismo de inferência, captura de informações (contexto) de uma forma sistêmica a permitir a conversão destas informações em ações.
- **OLTP (Online Transaction Processing):** são sistemas que têm a tarefa de monitorar e processar as funções básicas e rotineiras de uma organização, tais como processamento da folha de pagamento, faturamento, estoque, etc. Os fatores críticos de sucesso para este tipo de sistema são: alto grau de precisão, integridade a nível transacional e produção de documentos em tempo hábil.
- **OLAP (Online Analytical Processing):** trata-se da ferramenta mais popular para exploração de um data warehouse. OLAP fornece para organizações um método de acessar, visualizar, e analisar os dados corporativos com alta flexibilidade e performance, via um modelo de dados natural e intuitivo.

B. Linked Data & Mashups

O Linked Data trata de um conjunto de práticas que utiliza os padrões da Web Semântica para publicar e ligar dados heterogêneos. Diferentes fontes de dados estruturadas conforme padrões e devidamente ligadas geram como resultado uma expansão da rede de dados e abrangendo diversos domínios, assim como permitem um melhor processamento desses dados por máquinas, através de consultas inteligentes, inferências e ferramentas de apoio a decisão.

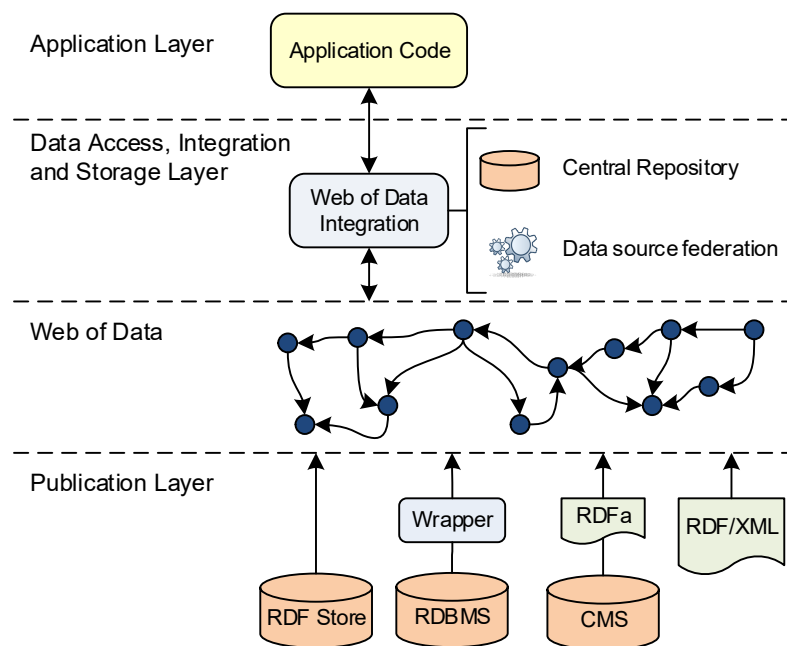


Figura 6 – Arquiteturas para Aplicações de Mashup de Dados Ligados

O Mashup de Dados Ligados refere-se às aplicações que conseguem fazer uso do Linked Data e construir visualizações a partir de diversas fontes, oferecendo um novo paradigma da transformação e integração de dados (fig 6).

Na figura 7 apresentamos um exemplo de Mashup de Dados com a integração de dados de companhias farmacêuticas, testes de remédios, mecanismos de ação das drogas e informações de segurança.

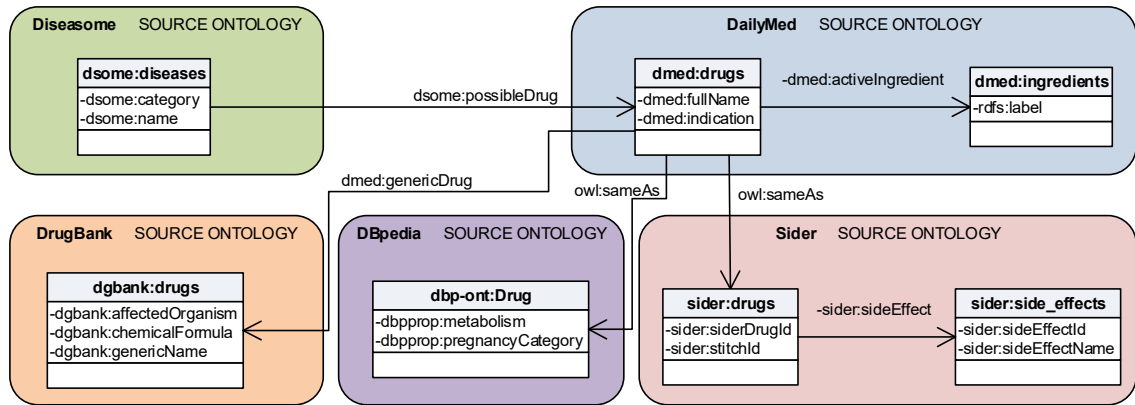


Figura 7 – Exemplo de Mashup

Mesmo com essa estrutura de Linked Data, existem algumas dificuldades inerentes a utilização de diversas fontes, e fontes heterogêneas. A primeira trata da identificação de fontes de dados relevantes, que venham a complementar a informação adquirida. Também a diversidade de vocabulários e significados (questão semântica) e a qualidade dos dados (problemas como fragmentação, incoerência e inconsistência).

Neste sentido faz-se uso de Ontologias (fig 8) que auxiliem no “entendimento” dos dados e na ligação dos diversos significados, assim como a utilização de frameworks para identificar ligações, limpar dados e especificar regras de inferência.

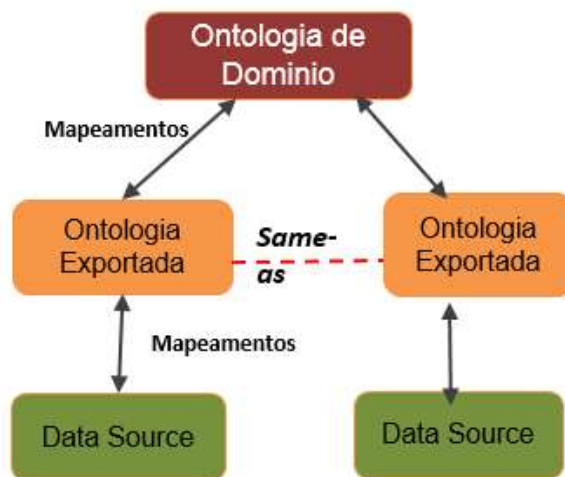


Figura 8 –Mashup Framework

6. MATERIAL E MÉTODOS

Promover a integração entre os diversos sistemas, a exemplo dos sistemas que compõe a estrutura computacional do SUS, é necessário e ao mesmo tempo complexo. A heterogeneidade das estruturas de dados e, mais ainda, da semântica, requer estratégias e ferramentas que consigam realizar a interoperabilidade entre esses sistemas, seus dados e significados.

É foco do Projeto CLIMA IN o desenvolvimento de um modelo de Gestão do Conhecimento do GISSA que faça uso de mecanismos de inferência, data mining e ontologias, assim como dos conceitos e aplicações do Linked Data, permitindo que os dados em diferentes bases possam ser “conectados”, a partir de padrões e ontologias, de forma que informações relacionadas possam ser complementadas e integradas, ampliando a compreensão do contexto sobre o qual as informações estão inseridas e que podem gerar como resultado.

Com o objetivo de ilustrar os métodos a serem usados no CLIMA IN quanto ao uso de Linked Data e Mushup, segue o seguinte cenário no contexto do GINGA:

Objetivo: Permitir a análise de dados sobre gestantes. Exemplos de consultas:

- Quantidade de gestantes potencialmente em risco, segundo uma região, período e/ou faixa-etária?
- Quantidade de mães que vieram a óbito por problema de diabetes, doenças crônicas, ou tabagismo.

Serão usados 6 passos neste cenário:

- Passo 1: Modelagem da Ontologia de Aplicação;
- Passo 2: Seleção das Fontes de Dados;
- Passo 3: Geração das Ontologias Exportadas;
- Passo 4: Identificação das Ligações (links same-as);
- Passo 5: Limpeza dos Dados.
- Passo 6: Especificação das Regras de Inferências

Passo 1: Modelagem da Ontologia de Aplicação:

Ação: Com a Ontologia de Aplicação, especificar formalmente os conceitos do domínio da aplicação em foco. Ela é utilizada como o vocabulário comum de integração de dados.

Modelo: A Ontologia de Aplicação a ser utilizada é mostrada na figura 9.

Passo 2: Seleção das Fontes de Dados;

Ação: Para a etapa de Seleção de Dados, serão escolhidas as fontes para obtenção de informações sobre a gestante e sua gestação (SINASC), bem como gestações passadas. Além disso, também é relevante informações sobre o indivíduo, tais como doenças crônicas, dados sociodemográficos, etc. (E-SUS).

Modelo: As bases utilizadas são do E-SUS e SINASC

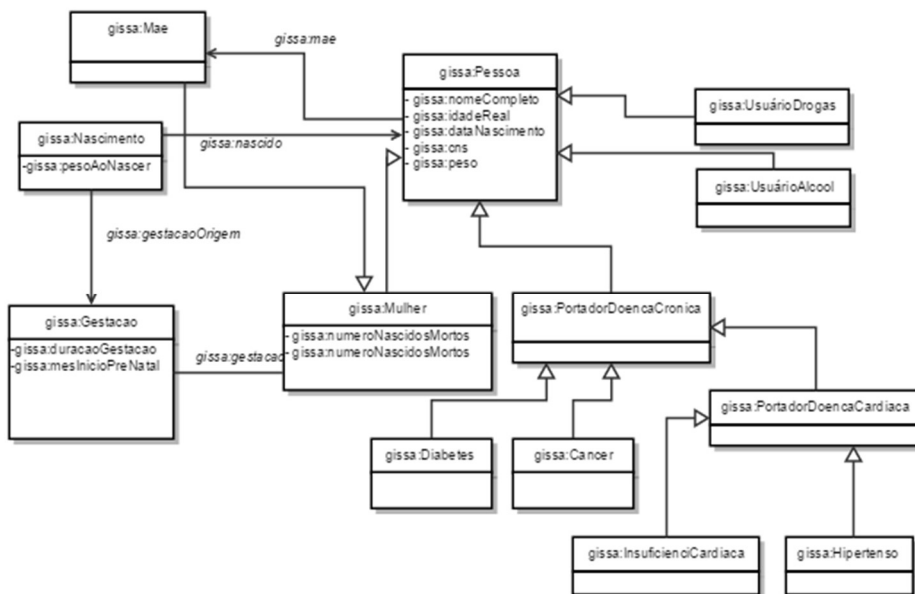


Figura 9 – Ontologia de Aplicação do GISSA

Passo 3: Geração das Ontologias Exportadas [13][16];

Ação: As Ontologias Exportadas descrevem os esquemas das fontes locais em termos da OA. A ontologia exportada é um sub conjunto da ontologia de domínio. Para gerá-las, é necessária a criação de regras de mapeamentos semânticos entre os esquemas das fontes de dados e a OA.

Modelo: Na Figura 10 é mostrado o mapeamento do esquema relacional, GISSA_REL, para a Ontologia Exportada, GISSA_OWL, e as Ontologias Exportadas das bases E-SUS e SINASC.

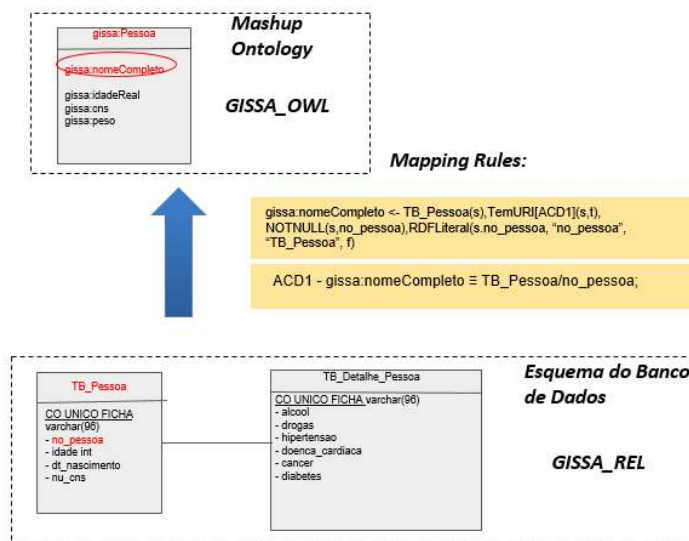


Figura 10 – Ontologia de Aplicação do GISSA

Passo 4: Identificação das Ligações (links same-as);

Ação: Os links *same-as* são responsáveis por identificar entidades iguais entre fontes de dados heterogêneas. A heurística utilizada para encontrar registros idênticos entre as bases selecionadas teve como referencial o match dos atributos Nome Completo, Data de Nascimento e Número CNS da entidade Pessoa.

Modelo: A figura 11 mostra a identificação dos links entre as duas fontes

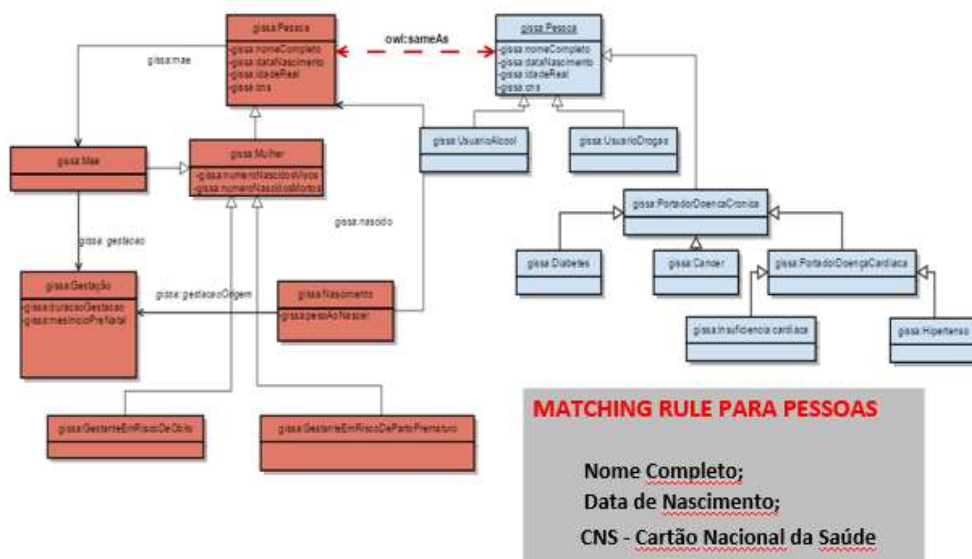


Figura 11 – Descoberta dos Links entre as duas fontes de dados

Passo 5: Limpeza dos Dados.

Ação: Trata-se do processo clássico de higienização das bases envolvidas

Modelo: A figura 12 mostra a explicitações dos links entre as duas bases (E-SUS e SINASC)

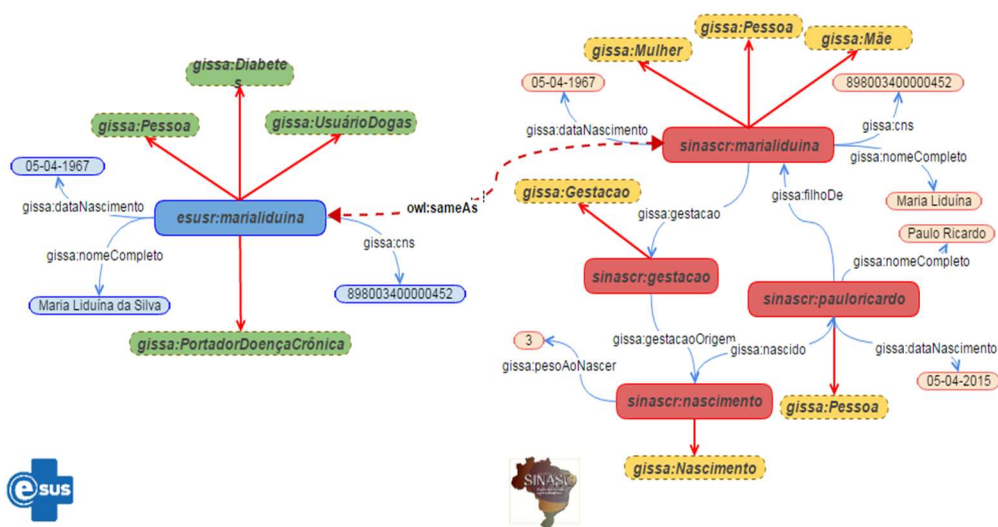


Figura 12 –Instância Exportada com ligação

Passo 6: Especificação das Regras de Inferências [13][16]

Ação: Com as Ontologias populadas e as regras definidas, assim como a ligação entre as fontes de dados estabelecida, inferir sobre os dados. Por exemplo, a quantidade de gestantes potencialmente em risco, segundo uma região, período ou faixa etária, ou ainda a quantidade de mães que vieram a óbito por problema de diabete, doenças crônicas ou tabagismo.

Modelo: Consiste na inferência entre as fontes de dados (não mostrada aqui devido ao espaço)

7. CRONOGRAMA DE ATIVIDADES DOS DOIS BOLSISTAS

São solicitados dois bolsistas para apoio aos trabalhos a serem desenvolvidos pelos pesquisadores envolvidos no projeto CLIMA. Como pode ser observado na tabela a seguir, das 6 atividades previstas para os bolsistas as 4 primeiras são comuns e as duas últimas são mais orientadas para os dois blocos principais do CLIMA IN: Framework de integração de dados (UFC) e para a metodologia de mineração de dados (IFCE).

ATIVIDADE 01: FRAMEWORK	1°	2°	3°	4°	5°	6°
01: Realizar estudo das tecnologias de Data Mining, Ontologia e Linked Data	X	X				
02: Estudo dos aspectos funcionais dos projeto GISSA e Point G CLIMA		X	X			
03: Análise das estórias e dos modelos de ontologias desenvolvidas no GISSA		X	X	X		
04: Estudo do framework de integração de dados desenvolvido pela equipe UFC			X	X		
05: Utilização do framework da UFC com as ontologias e dados do GISSA			X	X	X	
06: Apoio na pesquisa de integração com os mecanismos de inteligência				X	X	X
ATIVIDADE 02: INTELIGÊNCIA	1°	2°	3°	4°	5°	6°
01: Realizar estudo das tecnologias de Data Mining, Ontologia e Linked Data	X	X				
02: Estudo dos aspectos funcionais dos projeto GISSA e Point G CLIMA		X	X			
03: Análise das estórias e dos modelos de ontologias desenvolvidas no GISSA			X	X		
04: Estudo do framework de integração de dados desenvolvido pela equipe UFC			X	X		
05: Utilização da metodologia de Data Mining do IFCE no contexto GISSA			X	X	X	
06: Apoio na pesquisa de integração com os trabalhos do framework da UFC				X	X	X

REFERÊNCIAS

- [1] Relatório FINEP - Projeto GISSA, "User Story & Modelagem de Risco de Óbito Materno". Instituto Atlântico. Fortaleza. Ceará. 2015.
- [2] ANDRADE, L. O. M.; SANTOS, L. "Redes interfederativas de saúde e o SUS". Saúde para Debate, Rio de Janeiro, v. 42, p.27-34, 2008.

- [3] ANDRADE, L. O. M. "Inteligência de Governança para apoio à Tomada de Decisão". *Ciência & Saúde Coletiva*, v. 17, n. 4, p. 829-837, 2012.
- [4] Relatório FINEP - Projeto GISSA.: "Modelo de Inteligência de Gestão de Saúde. Meta Física 3 - Atividade 3.1". Instituto Atlântico. Fortaleza (Ce). 2015
- [5] Relatório FINEP - Projeto GISSA.: "Integração de Dados no Modelo de Inteligência de Gestão de Saúde". Meta Física 3 - Atividade 3.8. Instituto Atlântico. Fortaleza (Ce). 2015
- [6] Atiya Kazi, Prof. D.T. Kurian, "An Ontology Based Approach to Data Mining". *IJEDR* | Volume 2, Issue 4 | ISSN: 2321-9939. 2014.
- [7] Natalya F. Noy, "Semantic Integration: A Survey of Ontology-Based Approaches", *ACM SIGMOD Record*, vol 33, issue 4, pp 65-70, 2004
- [8] Oliveira A, M.; Andrade, L. O. M.; Bringel, Denis J.L. et al. "A Context-Aware Framework for Health Care Governance Decision Make System". *IEEE - Second International Workshop on Interdisciplinary Research on EHealth Services and Systems (IREHSS 2010) – Montreal*. 2010
- [9] G. Guizzardi, "Theoretical foundations and engineering tools for building ontologies as reference conceptual models," *Semant. Web J.*, vol. 1, no. 1,2, pp. 3–10, 2010.
- [10] G. Guizzardi and T. P. Sales, "Detection, Simulation and Elimination of Semantic Anti- patterns in Ontology-Driven Conceptual Models," in *International Conference on Conceptual Modeling (ER)*, 2014.
- [11] Liao, S.H., Ho, H.H., & Yang, F.C. (2009). Ontology-based data mining approach implemented on exploring product and brand spectrum. *Expert Systems with Applications* 36(9), 11730–1174
- [12] Marinica, C., Guillet, F., & Briand, H. (2008). Post-processing of discovered association rules using ontologies. In *Proceedings of IEEE International Conference on Data Mining Workshops* (pp. 126–133).
- [13] W3C, "RDF/XML Syntax Specification," 2014.
- [14] W3C, "OWL 2 Web Ontology Language Document Overview," 2012.
- [15] W3C, "SWRL: Semantic Web Language - Combining OWL and RuleML," 2004.
- [16] J. Hendler, "US Government Linked Open Data: Semantic.data.gov," *IEEE Intell. Syst.*, vol. 27, no. 3, pp. 25–31, Apr. 2012.