

## **PROJETO GISSA:**

### **META FÍSICA 3 – Atividade 3.8**

#### **Definir Modelo do Sistema Integrado Inteligente de Saúde**

**Draft 2.0**

**Novembro 2015**

## RESERVADO

<b>Responsáveis:</b> Mauro Oliveira & Vania Vidal		<b>Modelo:</b> B.01
<b>Elaborador(es):</b>	<b>e-mail</b>	<b>Telefone</b>
Mauro Oliveira	<a href="mailto:amauroboliveira@gmail.com">amauroboliveira@gmail.com</a>	(85) 99705 4321
Valdir Silveira	<a href="mailto:valdirsr@gmail.com">valdirsr@gmail.com</a>	
Vânia Vidal	<a href="mailto:vaniap.vidal@gmail.com">vaniap.vidal@gmail.com</a>	
Anilton Sales Garcia	<a href="mailto:aniltonsallesg@gmail.com">aniltonsallesg@gmail.com</a>	
Ronaldo Ramos	<a href="mailto:ronaldo.ramos@gmail.com">ronaldo.ramos@gmail.com</a>	
Tiago Prince Sales	<a href="mailto:tiago_sales@atlantico.com.br">tiago_sales@atlantico.com.br</a>	
Gabriel Lopes	<a href="mailto:gabriellopes9102@gmail.com">gabriellopes9102@gmail.com</a>	

## ÍNDICE

1	Histórico .....	4
2	Introdução .....	4
2.1	Sobre deste documento .....	4
2.2	Escopo do Modelo do Sistema Integrado Inteligência Saúde no GISSA .....	5
2.3	Glossário Resumido .....	5
3	Referencial Teórico .....	7
3.1	Integração de Sistemas de Saúde.....	7
3.1.1	Protocolos da Saúde e Barramento SUS.....	9
3.1.2	Interoperabilidade.....	10
3.2	Diagramas de Integração de BDs.....	11
3.3	Linked Data & Mushups .....	15
4	O Framework GISSA.....	17
4.1	Funcionalidades .....	17
4.2	Componentes do framework GISSA.....	19
5	modelo integrado de inteligência do gissa .....	21
5.1	Linked Data e Mashup no GISSA.....	21
5.3	Modelo de Inteligência do GISSA.....	29
5.4	Arquitetura do Modelo Integrado de Inteligência do GISSA.....	32
6	O GISSA como um SISTEMA INTEGRADO DE INTELIGÊNCIA .....	34
7	Conclusão .....	36
	Referências .....	37

## 1 HISTÓRICO

Data	Versão	Responsável	Alteração
06/04/2015	A.01	Ronaldo Ramos	- Criação do documento base 2 (Análise de Risco Sistêmico utilizando Mineração de Dados)
11/08/2015	A.02	Tiago Prince	- Criação do documento base 1 (Framework de Gestão do Conhecimento)
29/11/2015	A.03	Mauro Oliveira Anilton Garcia	- Criação do Documento Meta Física 3 0 Atividade 3.1 (Modelo de Inteligência de Gestão de Saúde)
30/11/2015	A.04	Anilton Garcia	- Revisão do documento META 3 – Atividade 3.1
02/12/2015	A.05	Mauro Oliveira Vânia Vidal	- Criação do Documento Meta Física 3 0 Atividade 3.8 (Modelo de Sistema Integrado de Inteligência de Saúde) – Draft 1.0
03/12/2015	A.06	Vânia Vidal Gabriel Lopes	- Revisão do documento META 3 – Atividade 3.8 Draft 1.1

## 2 INTRODUÇÃO

### 2.1 Sobre deste documento

Este documento tem como objetivo apresentar os resultados da META FÍSICA 3 – Atividade 3.8 que definem o Modelo do Sistema Integrado Inteligente de Saúde, no contexto do Projeto GISSA. A ideia central consiste em mostrar como os três componentes que compõem o modelo de inteligência de gestão na saúde do GISSA (Inferência por Regras, Busca Semântica e Mineração de Dados), objetos da META FÍSICA 3 – Atividade 3.8, agem de forma integrada no GINGA.

Este documento tem, portanto, como base:

- o relatório dos resultados da Atividade 3.1, Meta Física 3, que trata e define o **Modelo de Inteligência de Gestão na Saúde do Projeto GISSA**,
- O trabalho **Framework para Especificação e Manutenção Incremental de Mashup de dados Ligados**, desenvolvidos pelas equipes de pesquisa do LAR-A (IFCE) e ARIDA (UFC), coordenadas pelos Profs Mauro Oliveira e Vania Vidal, respectivamente;

## 2.2 Escopo do Modelo do Sistema Integrado Inteligência Saúde no GISSA

O GISSA fornecerá inteligência de governança na tomada de decisão nos cinco domínios clássicos da área de Gestão em Saúde: sistêmico, normativo, funcional, clínico e de cuidados, integrando os sistemas de informação em saúde no âmbito do Sistema Único de Saúde (SUS) e Cartão Nacional de Saúde (CNS).

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 do sistema, profissionais de saúde, gestores de instituições de saúde e gestores governamentais, tomadores de decisão do sistema de saúde. 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.

## 2.3 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) [1].
- **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.
- **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*.

- **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, extraindo ou ajudando a evidenciar padrões nestes dados e auxiliando na descoberta de conhecimento (ver conceito no item 3.1).
- **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.
- **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.

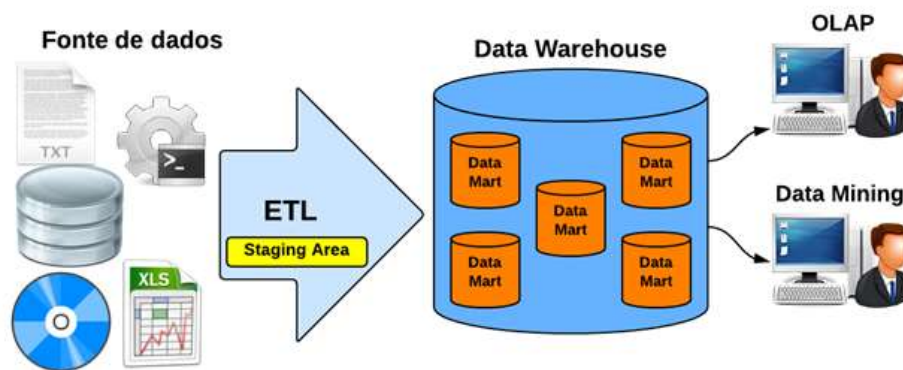


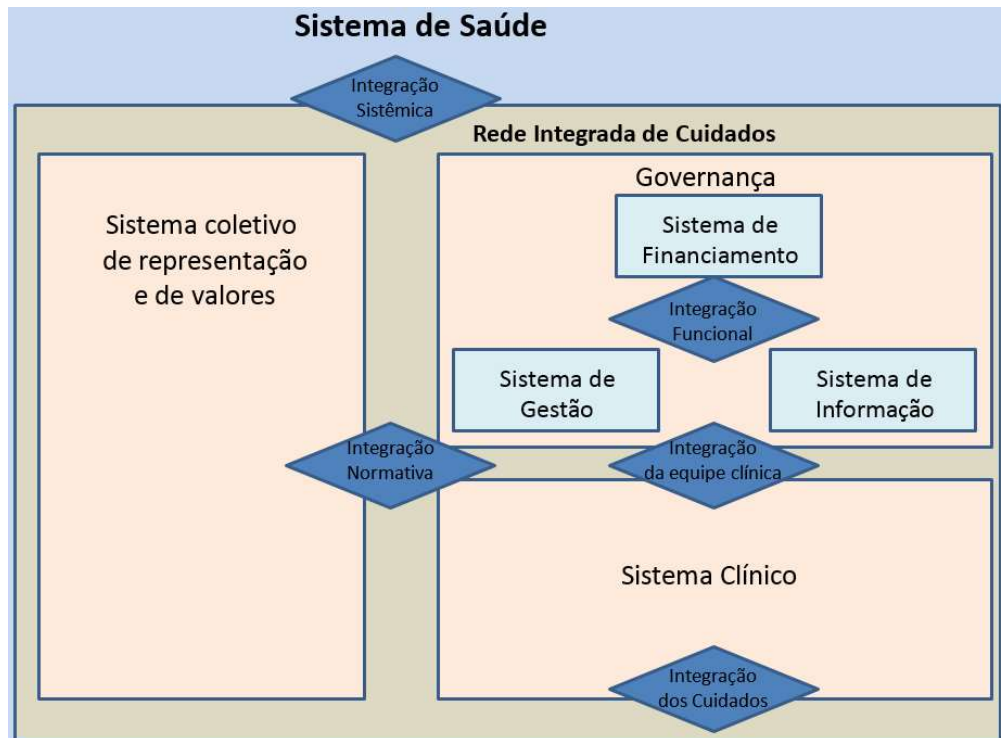
Figura 0. Fluxo de dados: OLTP/ ETL/ Datawarehouse/ DataMining-OLAP

### 3 REFERENCIAL TEÓRICO

Conforme dito no item 1, a ideia central deste documento consiste em mostrar como os três componentes que compõem o modelo de inteligência de gestão na saúde do GISSA (Inferência por Regras, Busca Semântica e Mineração de Dados), objetos da META FÍSICA 3 – Atividade 3.8, agem de forma integrada no framework GISSA.

#### 3.1 Integração de Sistemas de Saúde

O tema Integração 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 (NUNES, 2006). Marcou profundamente a própria denominação deste sistema na década de 80 quando este foi firmado na Constituição Federal Brasileira (BRASIL, 1988). A década de 90 foi marcada pelo processo de municipalização e o esforço de construir um sistema integrado em um país federativo (VASCONCELOS 2004). 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 (HARTZ, CONTANDRIOPOULOS, 2004).



**Figura 1** – Dimensões de um sistema integrado de saúde (CONTANDRIOPOULOS et al, 2001).

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 a figura 1

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 (CONTANDRIOPOULOS et al, 2001).

### **3.1.1 Protocolos da Saúde e Barramento SUS**

O Brasil tem longa tradição no uso de Sistemas de Informação em Saúde. Recentemente, tem havido iniciativas para a construção de um sistema informatizado do SUS que contemple a interoperabilidade dos sistemas do SUS e o Registro Eletrônico de Saúde (RES) do cidadão (Cartão Nacional de Saúde). Em consequência, cresce no País a expectativa de uso da e-Saúde. Desde maio de 2012 acontecem oficinas de RES com objetivo de construir essa visão estratégica.

São os seguintes os principais sistemas envolvidos com o SUS:

CADSUS (Sistema de Cadastramento de Usuários do Sistema Único de Saúde)  
SCPA (Sistema de Controle e Permissão de Acesso) SINASC (Sistema de Informações sobre Nascidos Vivos) SIM (Sistema de Informação sobre Mortalidade)  
CNES (Cadastro Nacional de Estabelecimentos de Saúde) SISRCA (Sistema de Regulação, Controle e Avaliação) SAI (Sistema de Informação Ambulatorial) SIH (Sistema de Informação Hospitalar) BPA (Boletim de Produção Ambulatorial) APAC (Autorização de Procedimentos de Alta Complexidade) SISREG (Sistema Nacional de Regulação) FARMÁCIA POPULAR HORUS (Sistema Nacional de Gestão da Assistência Farmacêutica) SAMU (Serviço de Atendimento Móvel de Urgência).

O plano de trabalho proposto para a área de Protocolos da Saúde & Barramento SUS tem como objetivo analisar os Protocolos de envolvidos no SUS, em especial os que se referem ao projeto Rede Cegonha.

Além das questões tratadas no contexto da interoperabilidade semântica, é de se atentar para a interoperabilidade sintática a ser provida pelo chamado “Barramento”, em seus seguintes aspectos:

Mecanismo de Interoperabilidade Infraestrutura tecnológica (implantado)  
Middleware de mensageria (implantado) Serviços (em desenvolvimento contínuo)  
Provido pelo DATASUS Serviços em desenvolvimento: CADSUS (Abr/2012) MPI (PIX/PDQ) (Fev/2013) CNES / RES Produção (Fev/2013) RES Clínico – primeira versão (Out/2013) Fase N: evoluções incrementais futuras (desde 2014).

Parte significativa do desenvolvimento do RES Nacional consiste na disponibilização de serviços de negócio no barramento

### **3.1.2 Interoperabilidade**

A interoperabilidade pode ser formalmente definida como “a habilidade de dois ou mais sistemas ou componentes de trocarem informações e utilizarem essa informação que foi trocada” (IEEE, 1990). Assim, para que esses sistemas ou componentes interoperem é preciso que a troca de informações entre eles seja livre de ambiguidades, ou seja, que as informações recebidas por um sistema receptor sejam compreendidas exatamente como foram enviados pelo sistema emissor.

Esse conceito nos leva a dois pontos fundamentais: (1) a troca de informação, que corresponde à interoperabilidade técnica; (2) a capacidade do receptor de utilizar a informação, que corresponde à interoperabilidade semântica.

- A interoperabilidade sintática é a habilidade de mover um dado de um sistema computacional x para um sistema y. Neste caso, a troca de dado não leva em consideração o significado do que está sendo trocado entre os sistemas e sim a convergência desses dados. Assim, nesta atividade, é preciso focar no problema de conectar sistemas distribuídos em uma rede.
- A interoperabilidade semântica vai um passo à frente da interoperabilidade sintática, abordando a necessidade de interoperabilidade de conceitos e vocabulários que são trocados entre os sistemas de forma que o emissor e o

receptor interpretem e entendam o dado da mesma forma. Neste caso, um alto nível de interoperabilidade é indispensável quando se deseja maximizar a utilidade da informação compartilhada e oferecer suporte à tomada inteligente de decisão. Portanto, quanto maior for o nível de interoperabilidade semântica (em nível de software), menor será a necessidade de haver “dados processados por humanos”. Conseqüentemente, pode-se por exemplo evitar erros humanos na entrada e análise de dados. Ao mesmo tempo, caso a solução não seja cuidadosamente desenvolvida, testada e implantada, poderá criar oportunidades para a inclusão de informações enganosas (intrusão) e, até mesmo, pro exemplo, de políticas equivocadas em processos de atendimento ao paciente.

No contexto do GISSA, a área de pesquisa em Interoperabilidade está focada na busca e/ou desenvolvimento de soluções (i.e., padrões, tecnologias, ferramentas e/ou linguagens) para o desafio da interoperabilidade técnica e semântica de sistemas de saúde que se relacionam para facilitar a interoperação entre os diferentes processos de Gestão em Saúde (legados e novos) no Brasil.

### **3.2 Diagramas de Integração de BDs**

As figuras 2,3,4 e 5 detalham a integração de bancos de dados (BDs) à arquitetura. Estas figuras, auto-explicativas, mostram o relacionamento dos grupos de BDs legados e BDs Gissa com a arquitetura via agentes de Integração de BDs.

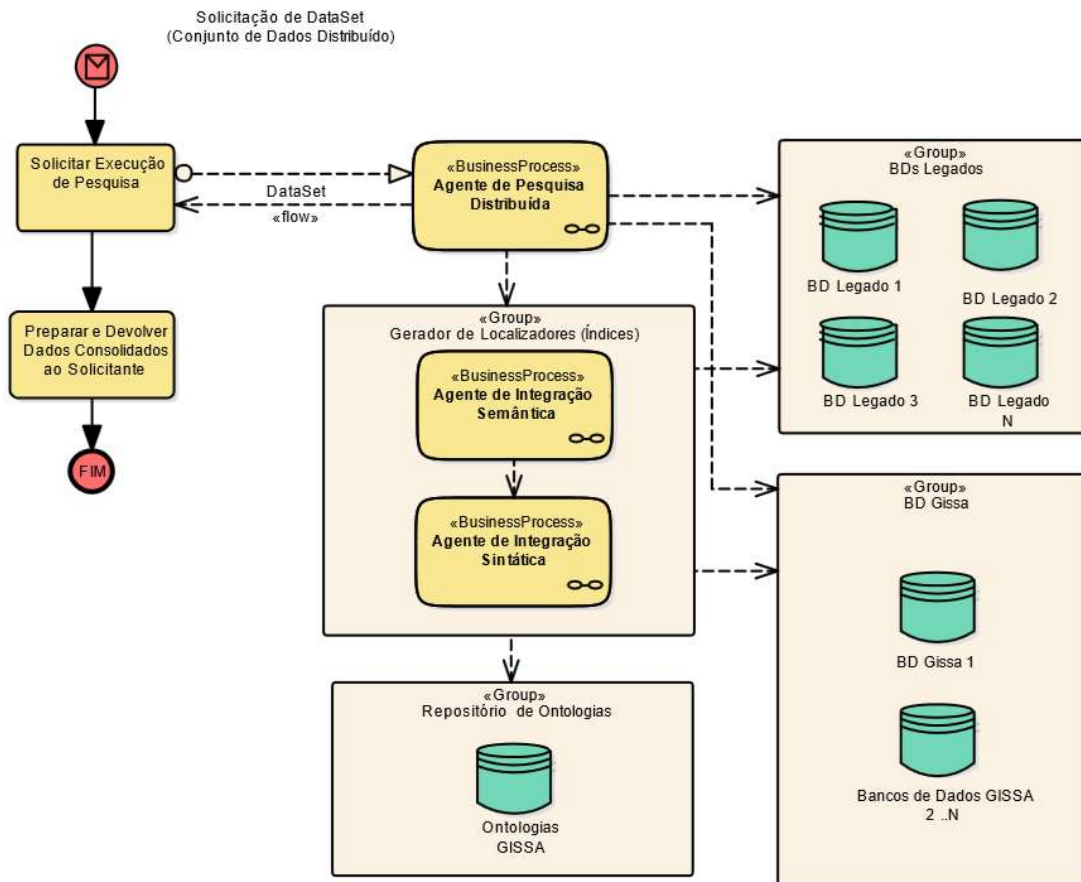


Figura 2 – Agente de Integração de BDs

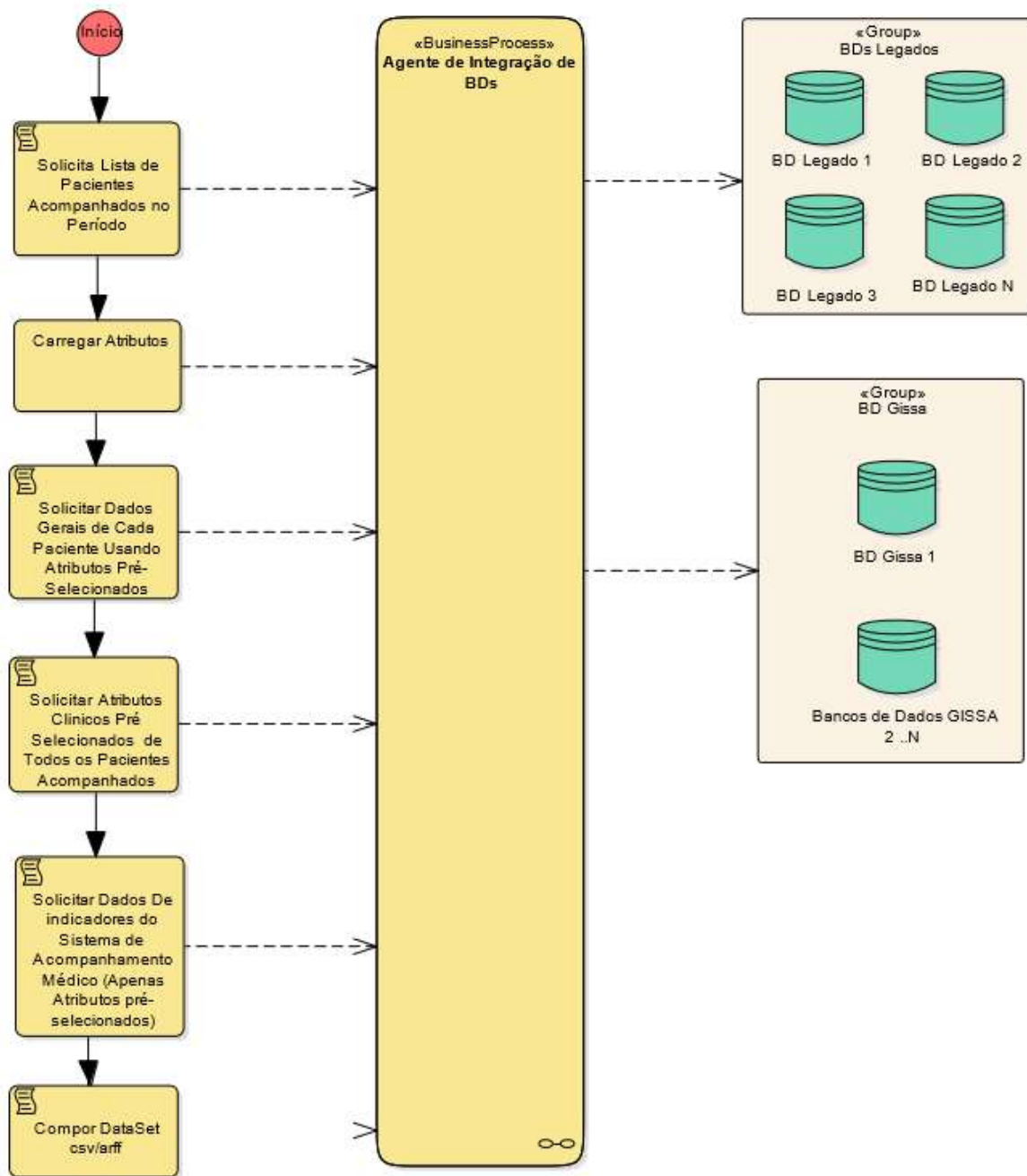


Figura 3 – Composição Dataset Inferência

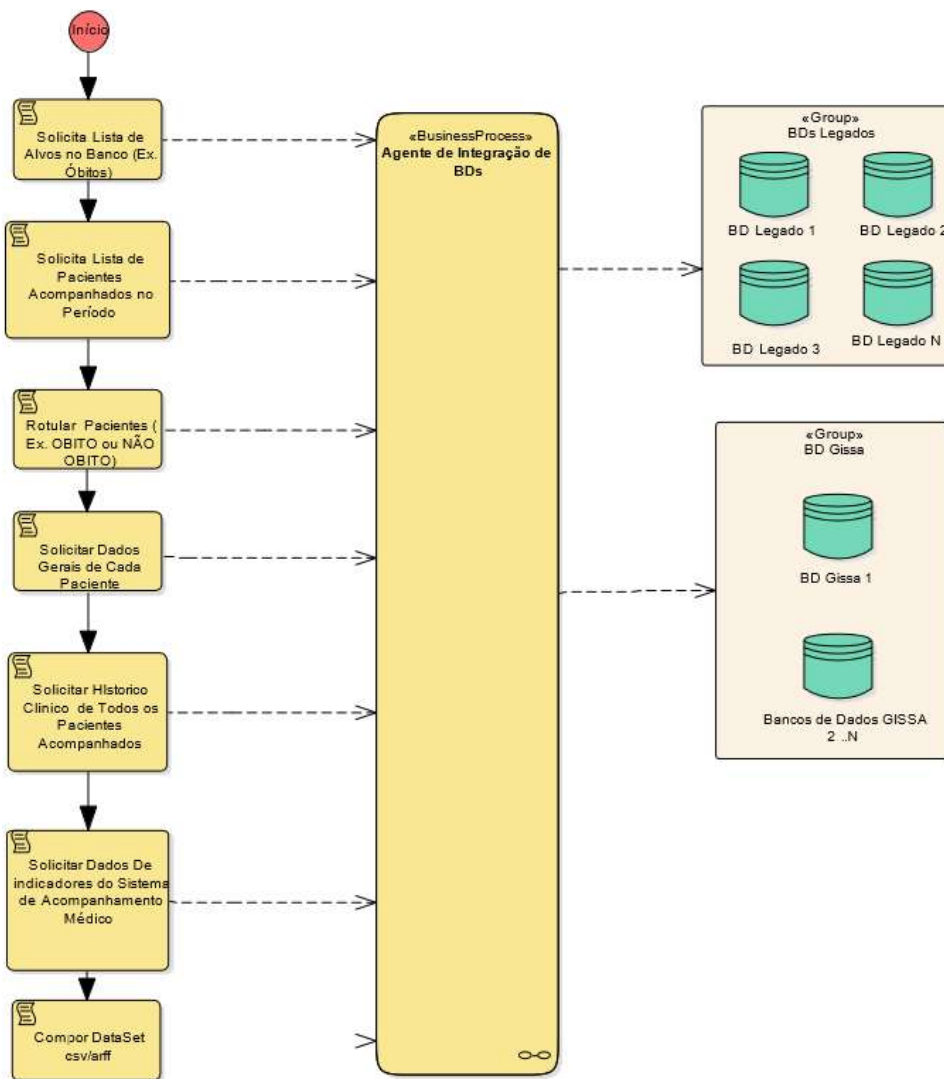


Figura 4 – Composição e Rotulação dos Datasets

Seleção do Tipo de DataSet a Preparar

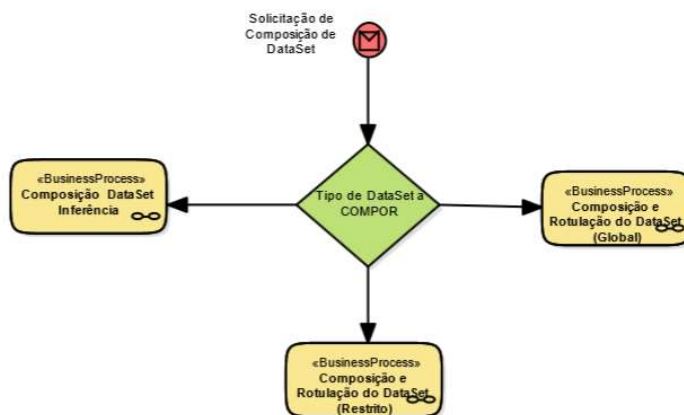
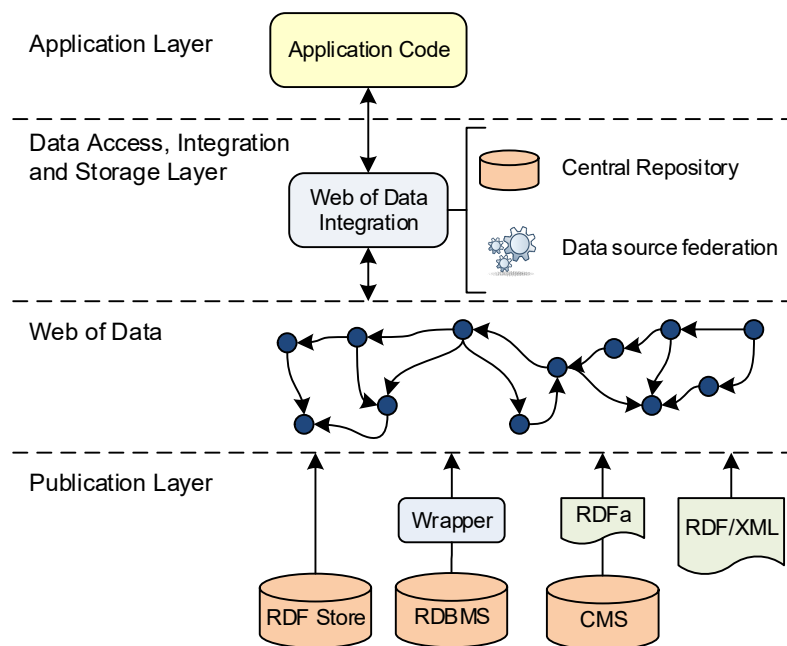


Figura 5 – Seleção de Dataset

### 3.3 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 a 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.

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).

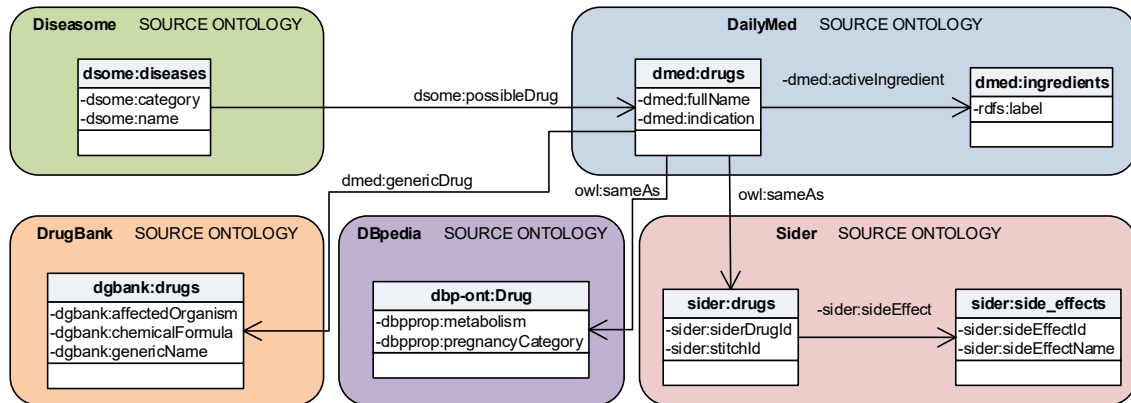


Figura 7 – Exemplo de Mashup

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.

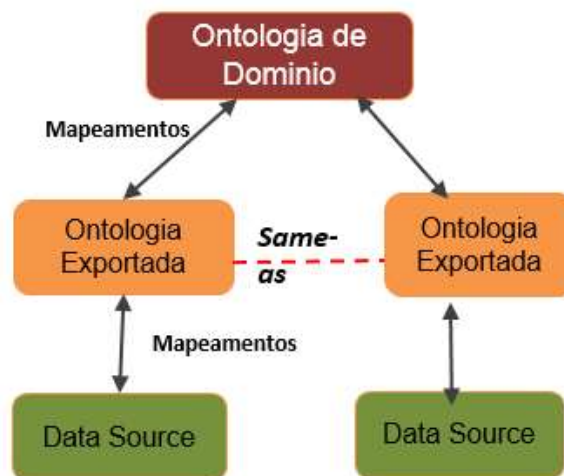


Figura 8 –Mashup Framework



## 4 O FRAMEWORK GISSA

O framework GISSA (de Governança Inteligente dos Sistemas de Saúde) é uma solução genérica para construção de sistemas de informação que apoiem o processo de tomada de decisão no contexto da saúde pública. Como prova de Conceito (PoC) do framework será implementada uma solução para apoiar a rede cegonha do município de Tauá – CE. O framework é formado por um conjunto de componentes para permitir 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.

### 4.1 Funcionalidades

O framework prevê cinco principais classes de funcionalidades, descritas a seguir.

**Relatórios** são funcionalidades que 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 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 um acompanhamento mais próximo.

**Indicadores** são funções que fornecem informação agregada sobre um determinado fato. Indicadores 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 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 relatório, 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.

São classificadas como **alertas** as funcionalidade 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.

A última macro funcionalidade que o GISSA apoia é a de **buscas**. Essa classe de solução 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; e
- *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.

## 4.2 Componentes do framework GISSA

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 9. Nela é mostrada a visão lógica do framework, destacando os componentes do framework de gestão do conhecimento (Inteligência) em azul.

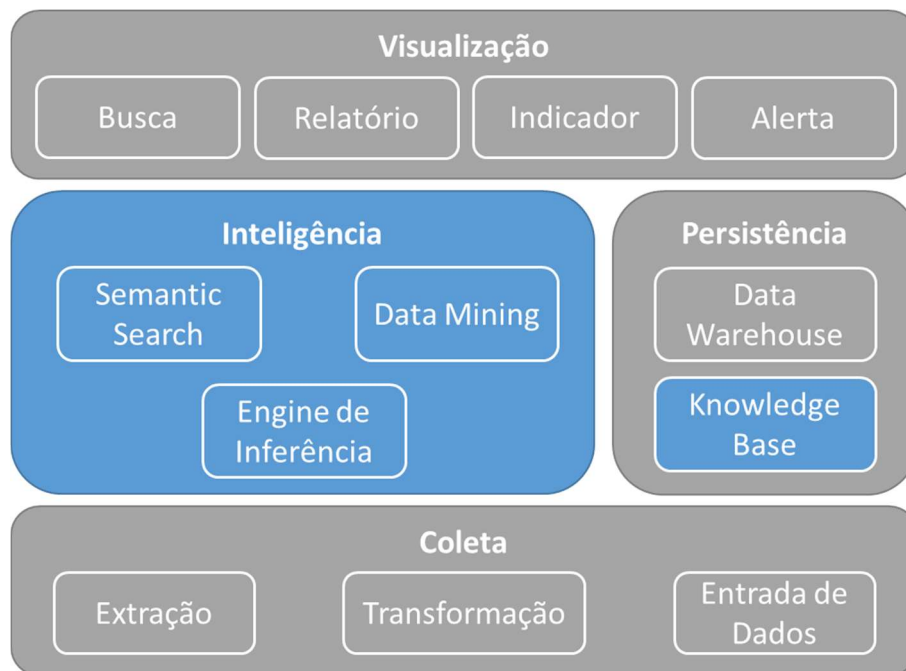


Figura 9. Visão lógica do GISSA framework, com ênfase no framework de GC

- **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 componente: 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* (como descrita na Seção 3) e a Base para as entradas de dados auxiliares.

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

Assim, ao modelo de inteligência do GISSA, através de uma saída extremamente refinada, é incumbida a tarefa de:

- Aplicar de Mineração de Dados e Inferência das Informações
- 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 a capacidade de adaptação
- Possuir algum tipo de feedback, característica imprescindível para um sistema inteligente.

## 5 MODELO INTEGRADO DE INTELIGÊNCIA DO GISSA

### 5.1 Linked Data e Mashup no GISSA

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 GISSA o desenvolvimento de um modelo de Gestão do Conhecimento que faça uso de mecanismos de inferência, datamining 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 o uso de Linked Data e Mashup no GINGA, segue o seguinte estudo de caso:

**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 nestes estudo de caso:

- 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**

A Ontologia de Aplicação especifica formalmente os conceitos do domínio da uma aplicação em foco. Ela é utilizada como o vocabulário comum de integração de dados.

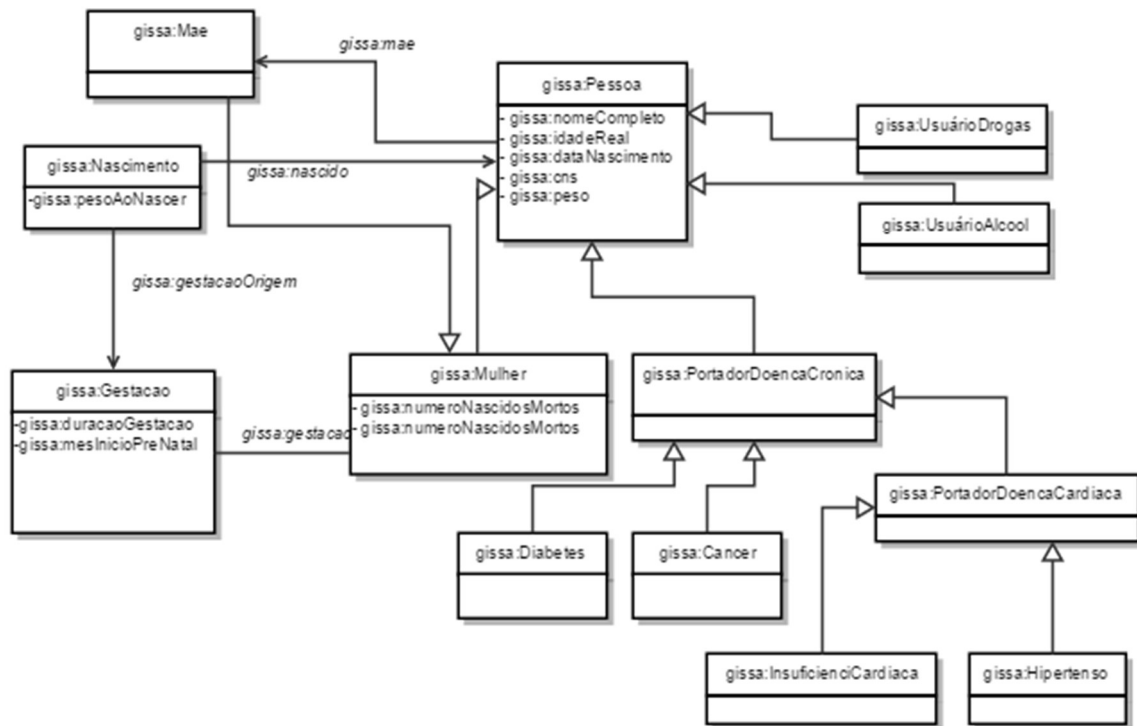


Figura 10 –Ontologia de Aplicação

**Passo 2: Seleção das Fontes de Dados;**

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

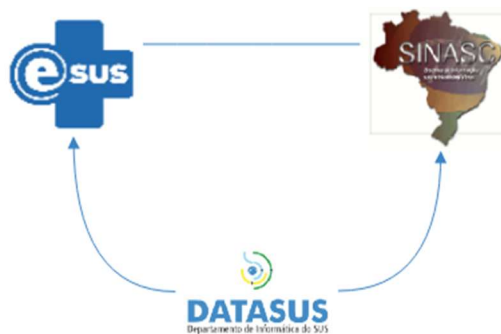


Figura 11 –Conjunto de Dados

**Passo 3: Geração das Ontologias Exportadas;**

As Ontologias Exportadas descrevem formalmente os esquemas das fontes locais em termos da OA. A ontologiaexportada 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. Na Figura 13, temos o mapeamento do esquema relacional, GISSA\_REL, para a Ontologia Exportada, GISSA\_OWL, bem como as Ontologias Exportadas de cada base, E-SUS e SINASC (Figura 14).

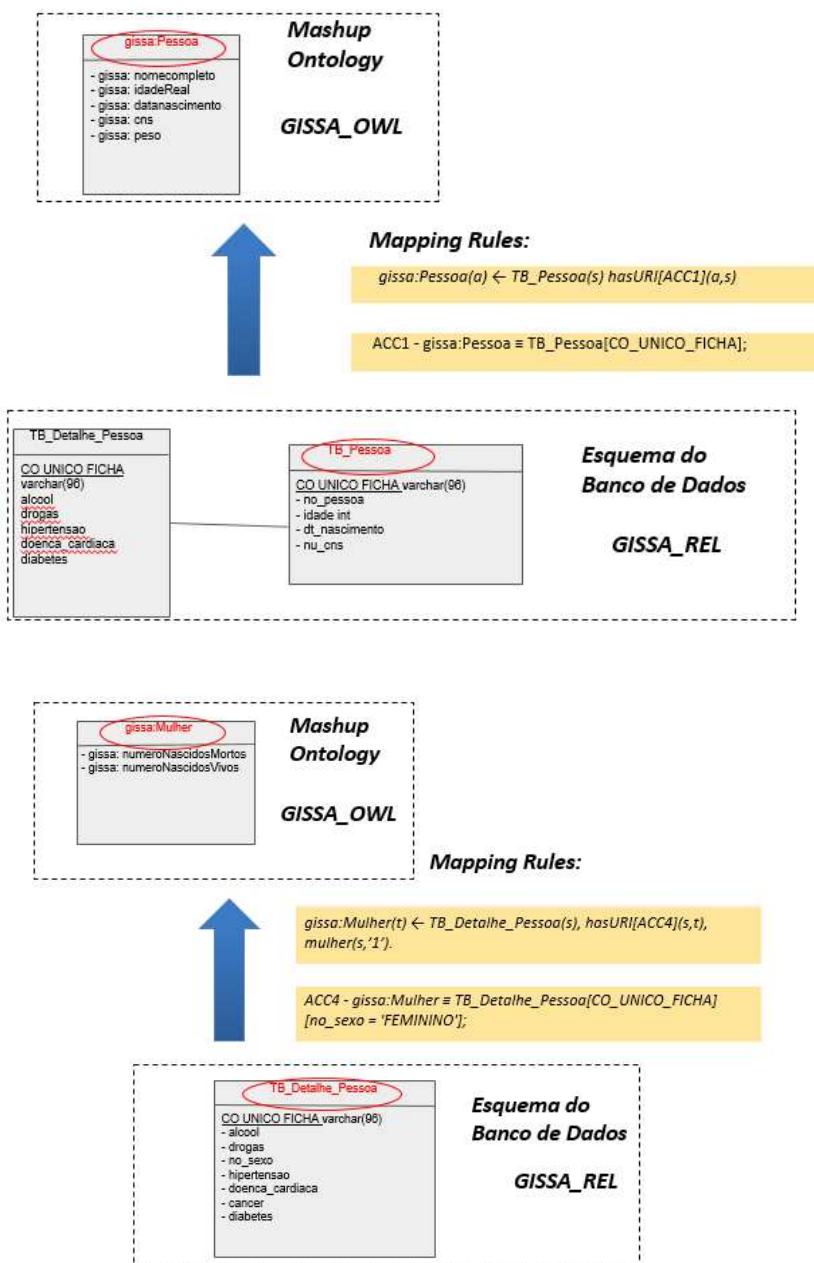


Figura 12 –Mashup Ontology Gissa\_OWL

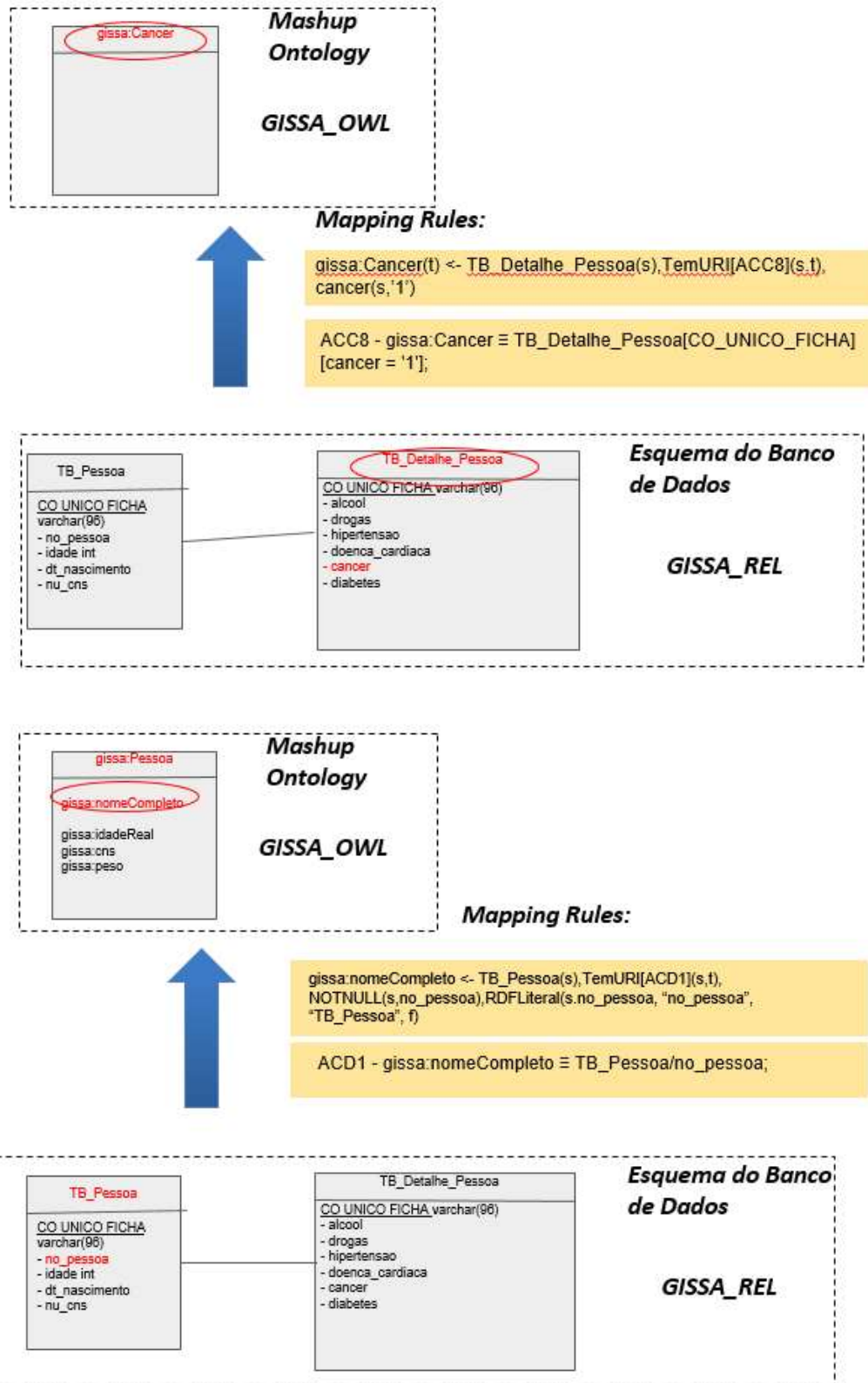


Figura 13 –Mashup Ontology Gissa\_OWL



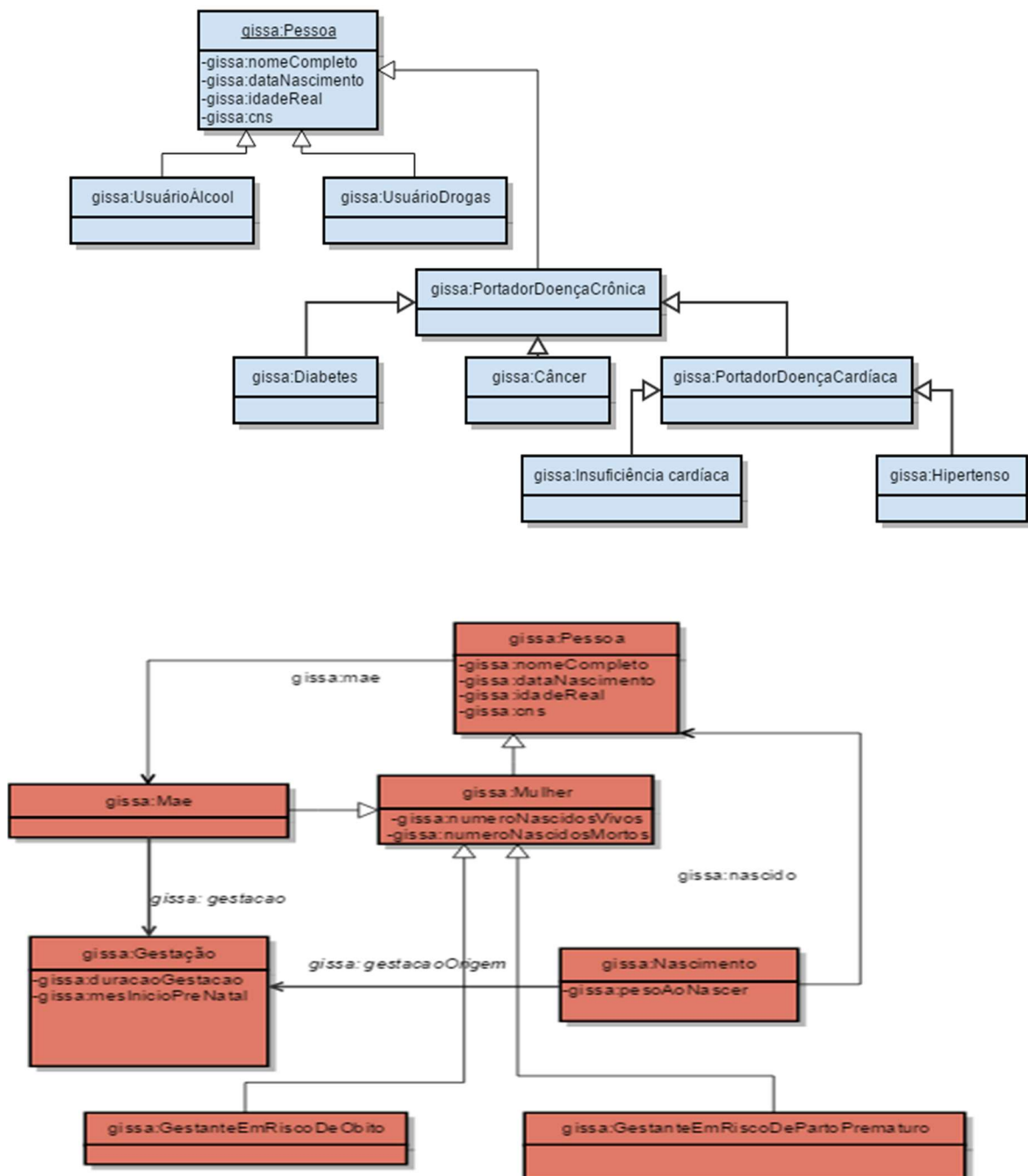


Figura 14 –Ontologias Exportadas (Atenção Básica e Informações de Nascidos Vivos)

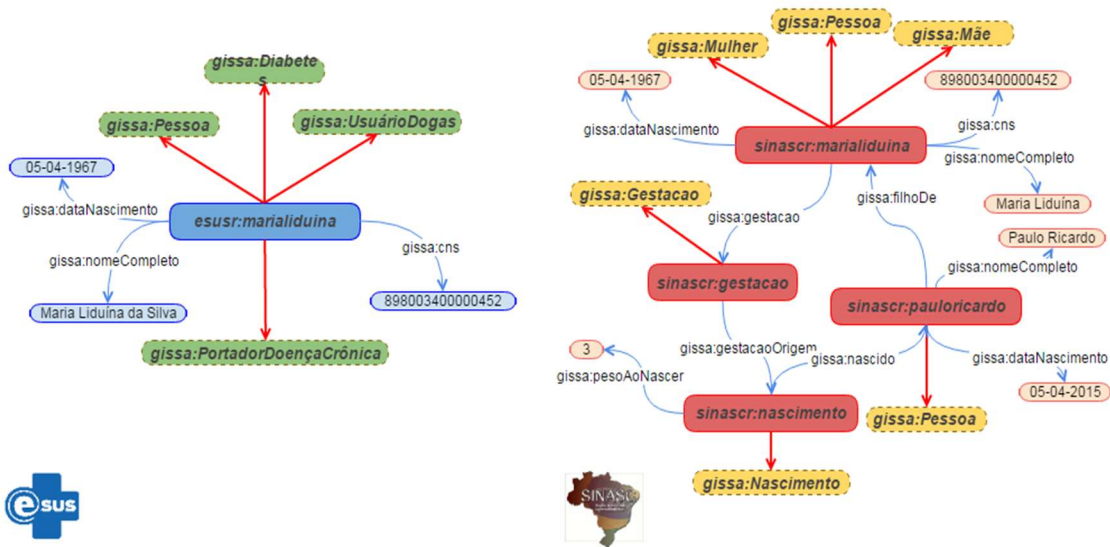


Figura 15 – Exemplo de Instância Exportada

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

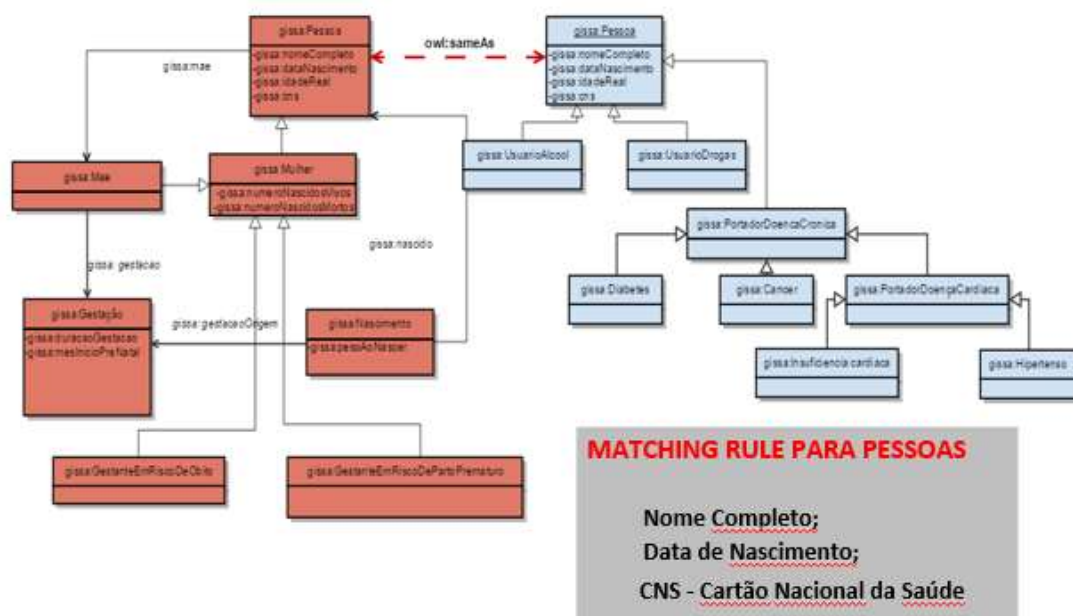


Figura 16 –Descoberta dos Links entre as duas fontes de dados

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**.

Na Figura 16, podemos ver a explicitação do link entre as duas bases. Já na Figura 17, podemos ver instâncias já triplicadas de ambas bases.

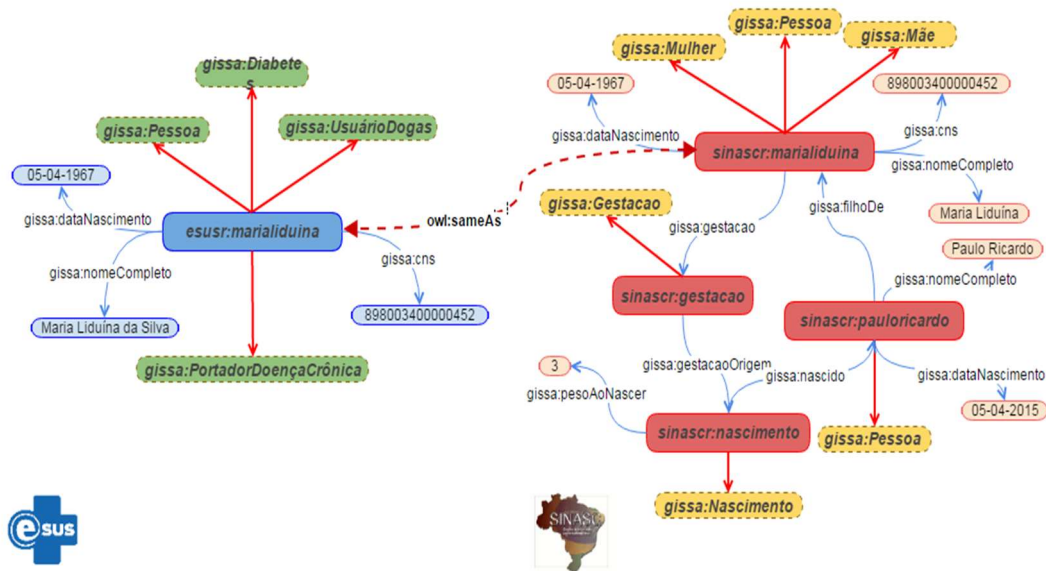


Figura 17 –Instância Exportada com ligação

Passo 5: Limpeza dos Dados.

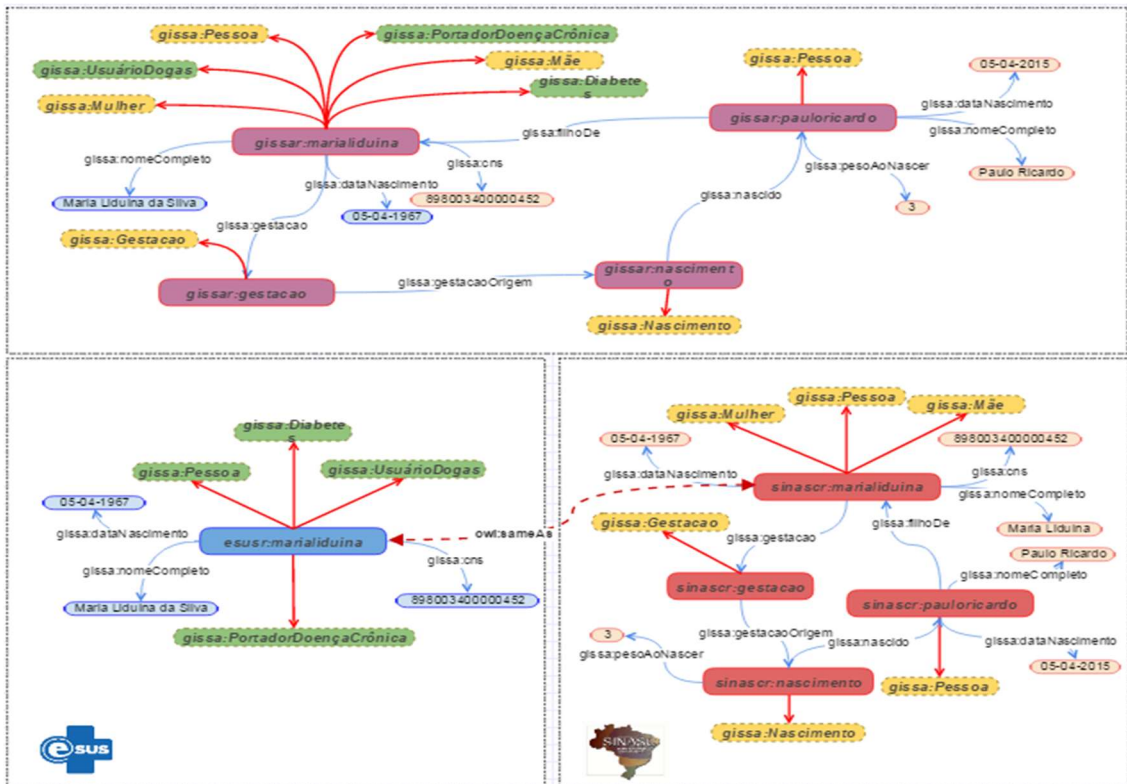


Figura 18 –Ontologia

Passo 6: Especificação das Regras de Inferências

Com as Ontologias populadas e as regras definidas, assim como a ligação entre as fontes de dados estabelecida, é possível 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.

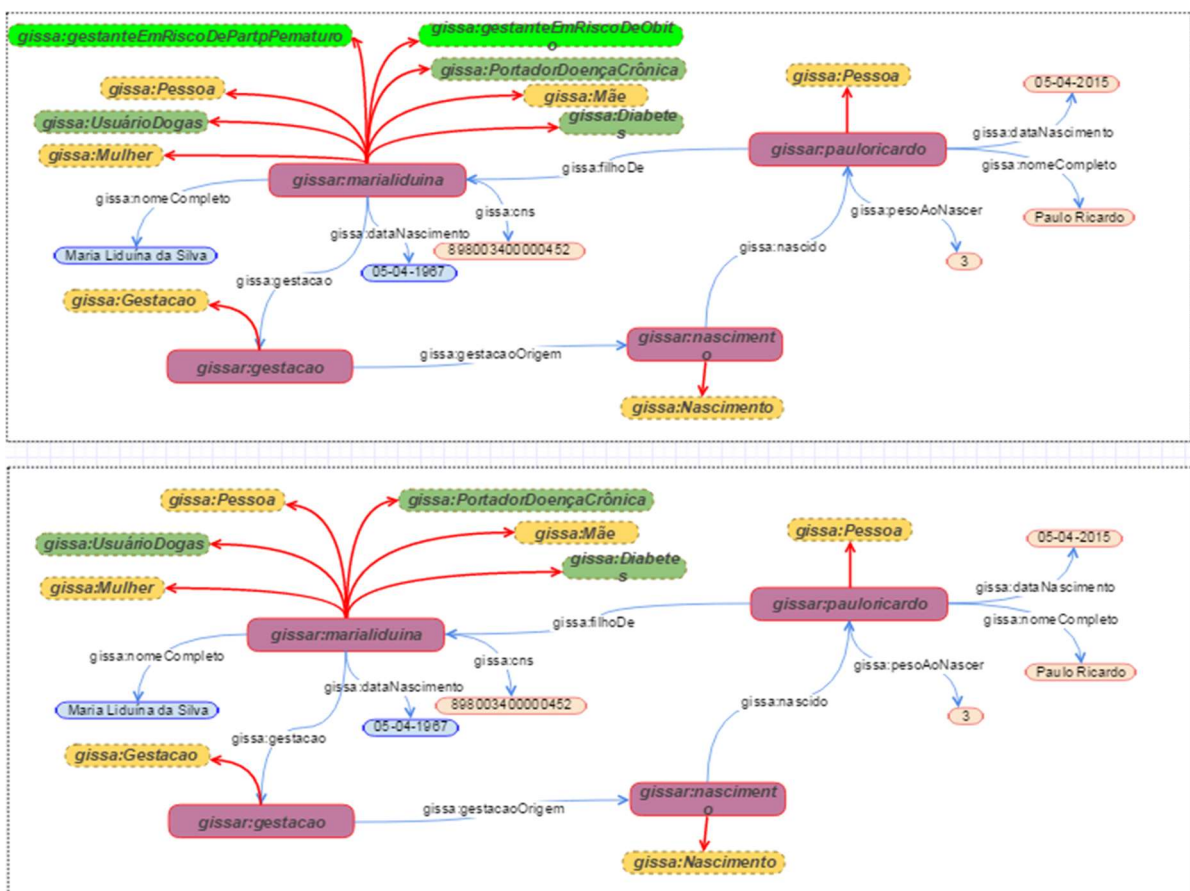


Figura 19 –Inferência entre as fontes de dados

### 5.3 Modelo de Inteligência do GISSA

A figura 20 mostra o Modelo de Inteligência de Mineração de Dados do GISSA (figura 7 do relatório de META FÍSICA 03 – Atividade 3.1), criado a partir de uma metodologia desenvolvida pelo Grupo de Aprendizagem de Máquina do Laboratório Centauro da UFC.

Modelo de Inteligência GISSA - Data Mining

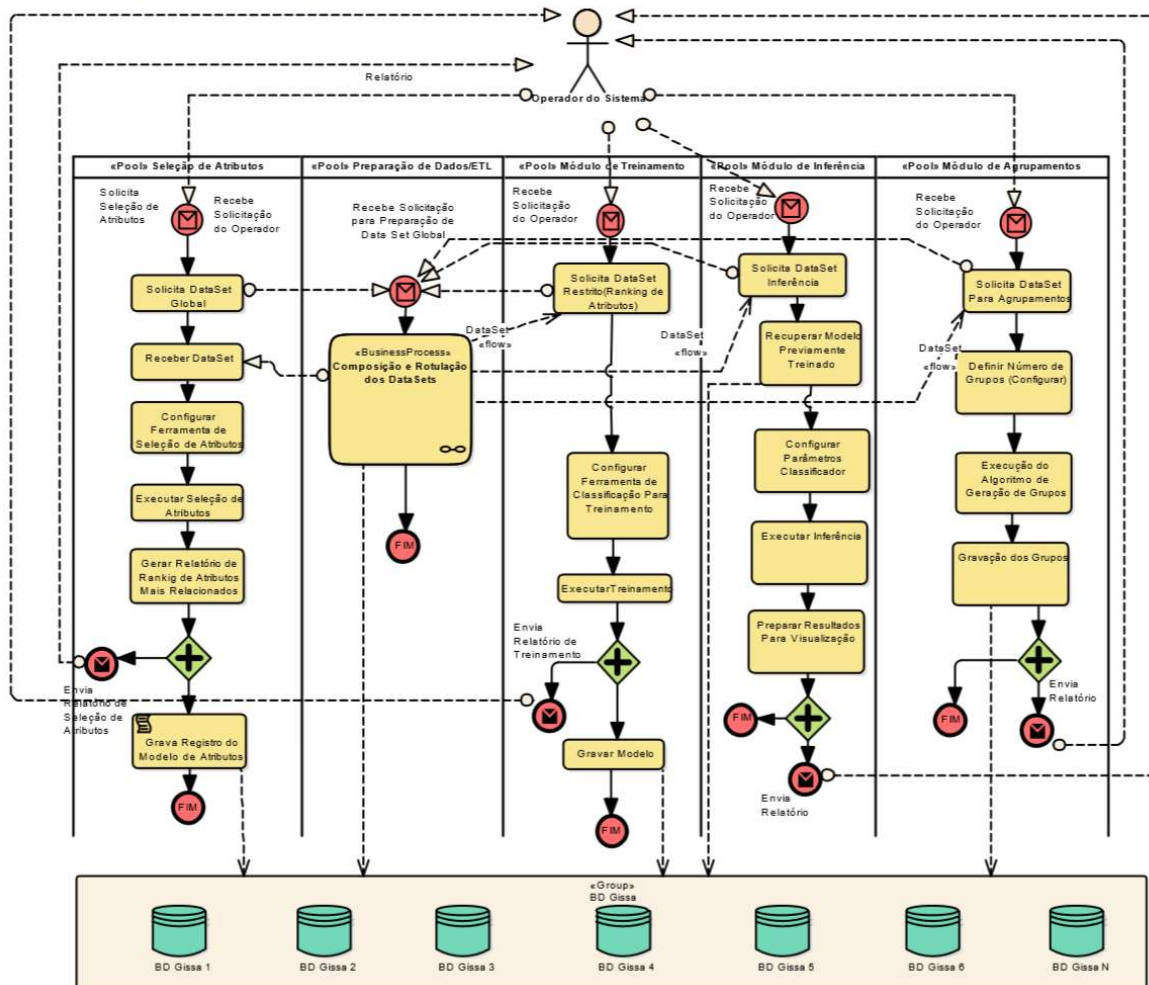


Figura 20 – Modelo de Inteligência do GISSA

Para a utilização do modelo será realizada a análise para aquisição, formatação e estruturação dos dados para a base de dados de data mining. Este passo possibilitará a construção dos modelos de inteligência computacional e a realização dos experimentos de data mining. Nesta atividade, também serão definidos as categorias a que cada padrão poderá pertencer.

O desempenho de métodos de data mining depende, além das diferenças entre os algoritmos e os parâmetros utilizados, da seleção dos atributos a serem utilizados pelo classificador.

A seleção dos atributos será realizada a partir de experimentos e análises de avaliação de desempenho de algoritmos de seleção de atributos a fim de obter um conjunto de características que melhor descrevem os dados do modelo.

O passo seguinte compreende a análise comparativa dos algoritmos de reconhecimento de padrões da literatura. Compreende-se nesta etapa a definição dos modelos de aprendizagem de máquina que serão utilizados bem como as abordagens de aprendizagem a serem adotadas (supervisionada, não supervisionada, semisupervisionada, etc). Além disso, será analisado se e como o sistema deverá ou não aprender após a inclusão de novas informações.

O framework open source Weka (FRANK; HALL; TRIGG, 2000) contém uma extensiva coleção de algoritmos de aprendizagem de máquina e será inicialmente utilizado para a avaliação e testes dos algoritmos de inteligência computacional. Para a realização de previsões/classificações, é necessário que o tempo de resposta seja viável e em tempo real.

Será realizada a avaliação das técnicas de ML selecionadas quanto a viabilidade de classificação em tempo real. Em seguida, será realizado o estudo de viabilidade de tecnologias de machine learning para incorporação no GI2S (frameworks e bibliotecas).

Nesta etapa, será verificado a possibilidade de reuso de ferramentas de data mining / inteligência computacional ou utilização de implementações específicas de algoritmos de Inteligência computacional.



Figura 21. Metodologia de Reconhecimentos de Padrões desenvolvida na UFC

Os entregáveis deste projeto são a lista de atributos mais relevantes e os algoritmos de aprendizagem de máquina que deverão ser utilizados para cada indicador/alerta/relatório do GISSA.

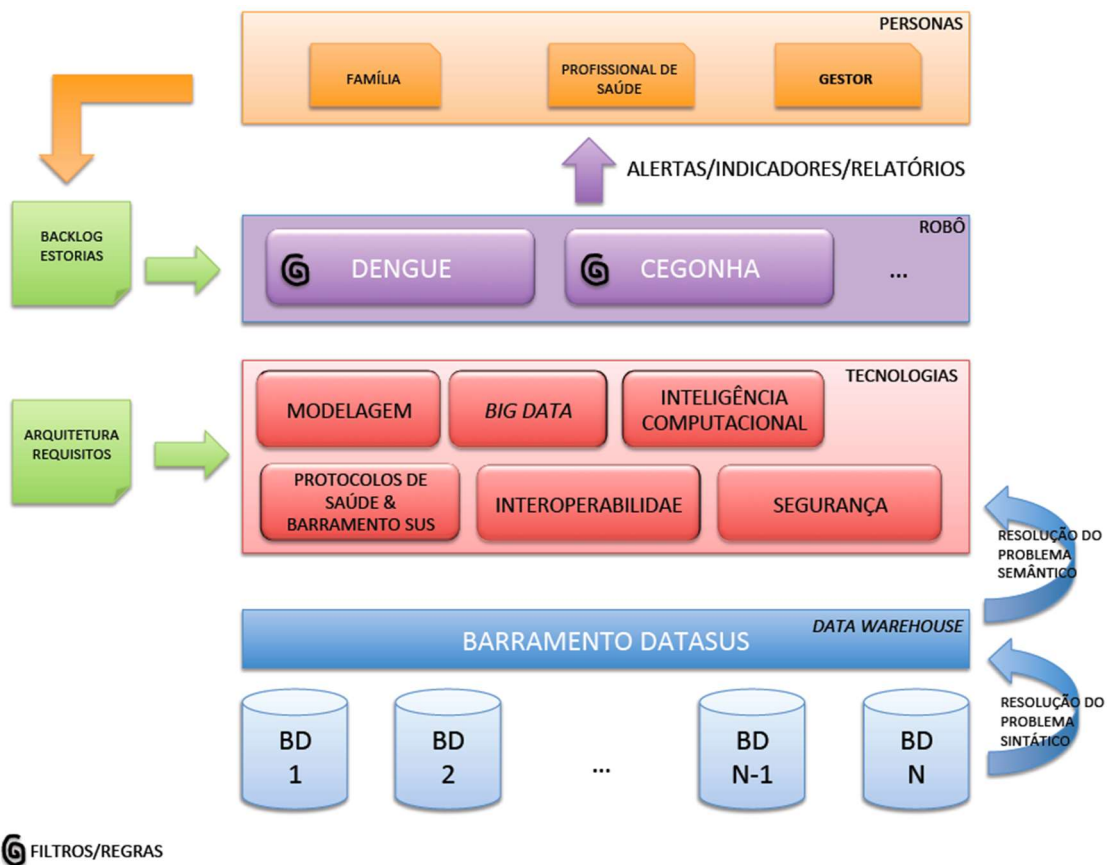
A figura 9 mostra as diversas etapas da metodologia utilizada no desenvolvimento da arquitetura de inteligência do GISSA mostrada na figura 21 (detalhado no item 5.3 do relatório de META FÍSICA 03 – Atividade 3.1).

1. NIVELAMENTO: Nesta fase é realizado um treinamento com todos os envolvidos. Trata-se de um treinamento básico sobre reconhecimento de padrões e aprendizagem de máquina. Ao fim do treinamento toda a equipe estará falando o mesmo “idioma”.
2. PROSPECCAO INICIAL: Os representantes da empresa (cliente) apresentam claramente o problema. Esta apresentação deve ser repetida até que todos estejam de acordo sobre os desafios.
3. SELECAO INICIAL: Nesta fase são selecionados ferramentas e algoritmos a serem testados. Podem ser selecionados classificadores elementares, redes neutras, comitês de classificadores, etc.
4. PREPARAÇÃO DOS DADOS: Dependendo como os dados estão disponíveis é possível que se tenha que voltar ao passo 3.
5. TRANSFORMAÇÃO DOS DADOS: Modificação nos dados para submissão à ferramentas de análise.
6. EXPERIMENTOS: Primeiros experimentos com as ferramentas selecionadas;
7. GANHADORA: Identificação do algoritmo de Machine Learning que comporá a tecnologia solução.
8. CONSTRUÇÃO: Nesta fase é construída uma ferramenta de software contendo uma solução fechada para o problema proposto. Aqui usa-se Engenharia de Software convencional.
9. EXPERIMENTOS: Realização de experimentos com a ferramenta desenvolvida.
10. TESTES: Realização de testes de integração e testes unitários
11. ENTREGA DO PRODUTO DE SOFTWARE: O módulo Bloco Mineração de Dados estará pronto para ser incorporado ao componente de Gestão do Conhecimento do framework GISSA.

### 5.4 Arquitetura do Modelo Integrado de Inteligência do GISSA

A figura 22 mostra um primeiro draft do diagrama de blocos funcionais do GISSA, criada a partir da visão lógica do GISSA framework (figura e do relatório META FÍSICA 3 – Atividade 3.1). Nela merecem destaque:

- Resolução do problema sintático: o bloco Datawarehouse extrai informações dos diversos bancos de dados existentes no sistema de saúde, de estruturas das mais diversas (heterogêneas).
- Resolução do problema semântico: esta questão é delegada ao bloco Tecnologias onde encontram-se mecanismos de modelagem, interoperabilidade e inteligência computacional.



FILTROS/REGRAS

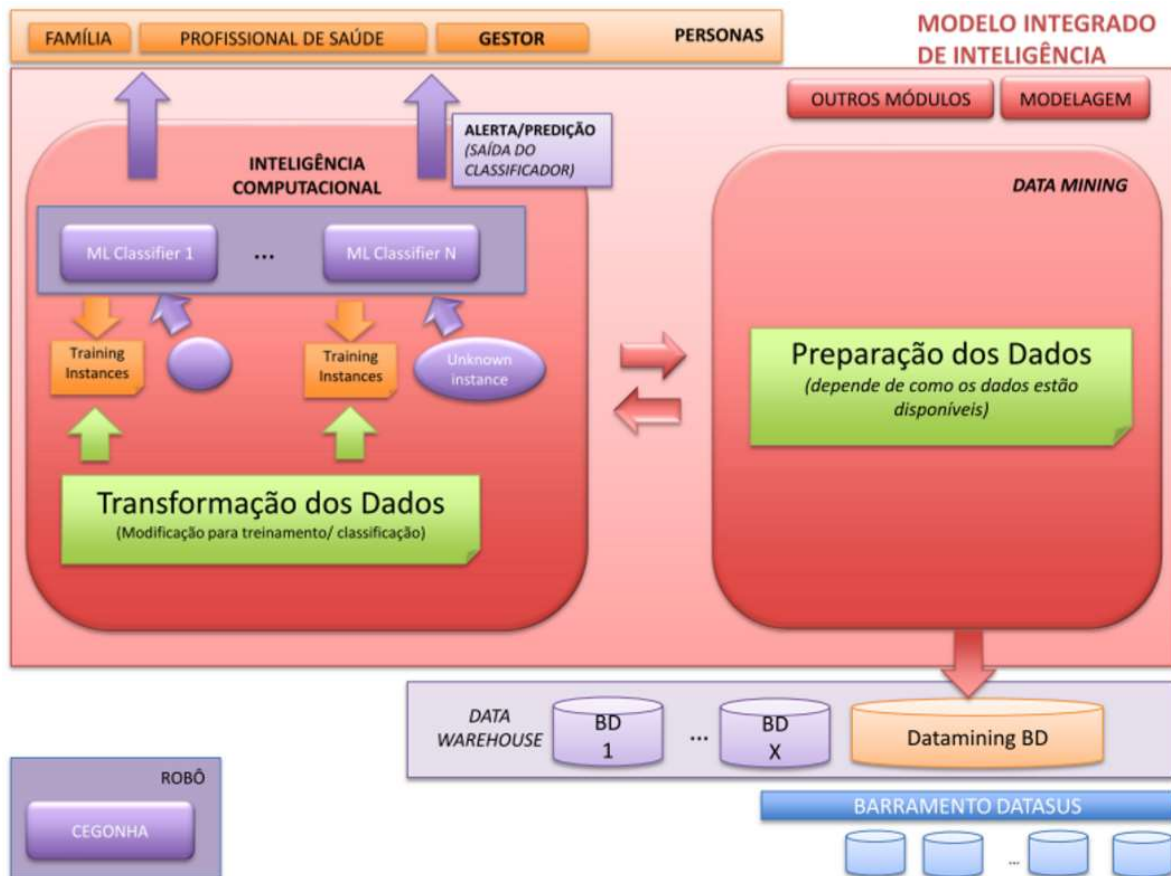
**Figura 22** – Diagrama de blocos funcionais do GISSA.

Os dois outros blocos da figura 2 são o “robô”, que deverá engendrar informações capazes de alimentar aplicações para contextos como a Rede Cegonha Dengue, etc, e o bloco Personas que representa os atores envolvidos no escopo do projeto.

A arquitetura do Sistema Integrado Inteligente de Saúde, objeto deste relatório, compreende a adequação de uma engine de inteligência para inserção como um dos módulos do framework GISSA, como anunciado na figura 22.



A arquitetura básica do modelo integrado de inteligência do GISSA é apresentada na Figura 23. Percebe-se a relação desta arquitetura com o diagrama de blocos funcionais do GISSA apresentado na figura 22, onde os componentes Inteligência computacional e Datamining do bloco tecnologia são expandidos.



**Figura 23** – Arquitetura Básica do Modelo Integrado de Inteligência

Esta arquitetura proposta deverá especificar claramente como os dados serão preparados e formatados para utilização pela Engine de Inteligência. Em outras palavras, a etapa de Data Mining (Mineração de dados) será especificamente detalhada, com entradas e saídas bem definidas.

Para realizar os experimentos de Data Mining, é necessário a aquisição, formatação e estruturação dos dados de maneira lógica. A partir destes dados, a avaliação de desempenho dos algoritmos de seleção de atributos e classificação (machine learning / reconhecimento de padrões) poderá ser realizada (ver “Framework de Gestão do Conhecimento, item 5 do relatório da META 3 – Atividade 3.1: Modelo de Inteligência de Gestão na Saúde).

A fase desafiadora na arquitetura é a "caça aos dados". Inicialmente, será criado um grafo que mostre os atributos dos bancos de dados e onde estes estão. Esta etapa tem uma interface com o grupo de Ontologias.

A partir do Modelo Integrado proposto, serão gerados alertas que a engine de inteligência irá fornecer aos diferentes personas através das predições obtidas a partir dos modelos de classificação as predições. A priorização no desenvolvimento dos alertas terá como base as necessidades das personas. A grande relevância destes alertas é prever antecipadamente situações indesejadas, como óbito materno ou infantil, e poder tomar decisões que venham impedir tais situações.

## 6 O GISSA COMO UM SISTEMA INTEGRADO DE INTELIGÊNCIA

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, enquanto uma especificação formal (entendida pelo computador) de uma conceitualização (modelo abstrato de algo real) explícita (pode ser compartilhada) sobre um domínio de interesse, se apresenta neste contexto como um mecanismo eficiente na estruturação do conhecimento relativo a estes dados a serem tratados pelas ferramentas de Data Mining.

A figura 24, extraída de [26], mostra o Data Mining na melhoria do processo de armazenamento e integração de dados do Data Warehouse onde é proposto que Ontologia seja usado 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 [27].

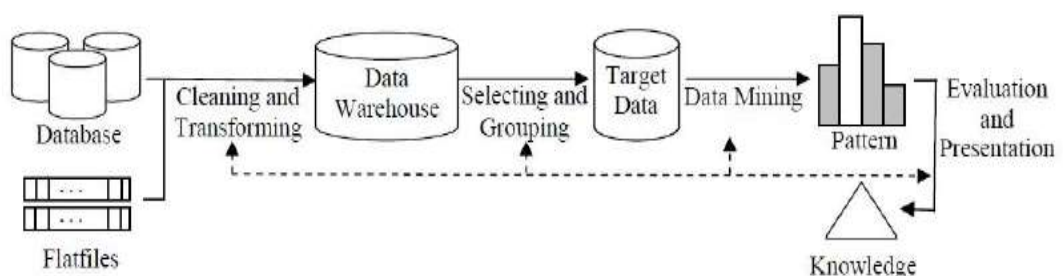


Fig 24: Processo inteligente de Dataming apoiado por Ontologias

O resultado é um Data Warehouse inteligente centrado na ontologia do usuário, o que pode ajudá-lo a contruir modelos que evitem a geração de padrões ineficientes na tarefa de predição do Data Mining. Como mostra a figura 24, este novo processo apoiado por ontologias pode estender novas regras produzindo um mecanismo ativo para a redescoberta de conhecimentos.

Em [26] 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 metodologia a ser utilizado no GISSA na construção deste bloco já foi devidamente apresentada no Relatório META FÍSICA 3 – Atividade 3.1 (item X) e rediscutida neste relatório (item Y).
- A construção do bloco Ontologia que terá como entrada o conhecimento extraído do bloco Data Mining, como mostrado na figura 25, extraída de [26].

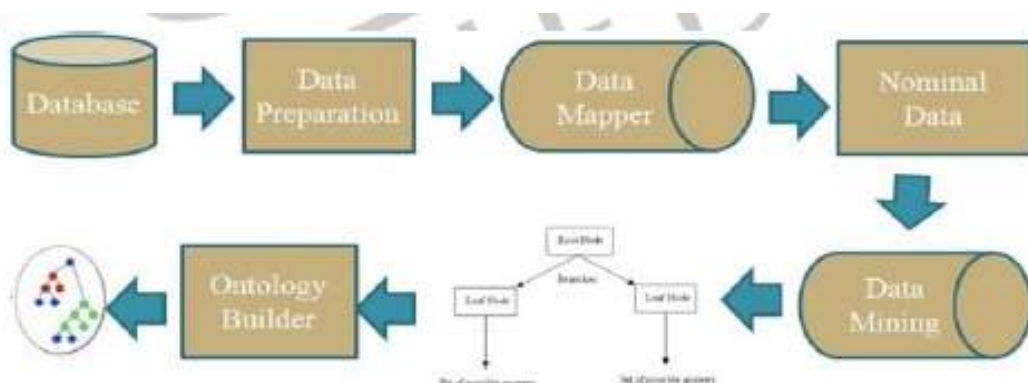


Fig 25: Construção da Ontologia a partir do Data Mining

## 7 CONCLUSÃO

O projeto GISSA encontra-se com 35% de seu tempo de execução. Foi implementado um protótipo pela equipe de desenvolvimento que será utilizado como PoC na cidade de Tauá (Ce), conforme previsto no cronograma do projeto.

Este protótipo contém os blocos funcionais Data Warehouse e Business Intelligence relativo as “estórias” que foram construídas pela equipe de saúde do GISSA, relativas ao universo de abrangência da Rede Cegonha. Foram elaboradas as ontologias associadas aos diversos cenários deste universo.

Neste íterim, a equipe de pesquisa projetou uma arquitetura de Data Mining do GISSA (detalhada nos relatórios de atividade 3.1 e neste relatório). Também foram realizados estudos sobre a integração de dados (ver item 5.3). Estes estudos são baseados em dados reais extraídos de bases do eSUS e SINASC.

Assim, uma primeira avaliação sobre o andamento do projeto mostra uma certa sinergia entre as equipes de saúde, desenvolvimento e de pesquisa que formam o time GISSA.

Os próximos passos no que diz respeito à definição do modelo integrado de inteligência do GISSA, objeto deste relatório (Atividade 3.8) são os seguintes.

1. Definição de qual modelo deverá ser adotado na combinação entre Ontologia e Data Mining no GISSA, dentre os dois modelos apresentados no item 6 deste relatório.
2. Como harmonizar a proposta de integração de dados baseada em Linked Data (ver item 5) na definição do modelo de inteligência do GISSA.

## REFERÊNCIAS

- [1] G. Fischer and J. Ostwald, "Knowledge Management: Problems, Promises, Realities, and Challenges," *Intell. Syst.*, vol. 16, no. 1, pp. 60–72, 2001.
- [2] M. Alavi and D. E. Leidner, "Knowledge Management Systems: Issues, Challenges, and Benefits," *Commun. Assoc. Inf. Syst.*, vol. 1, no. 2es, 1999.
- [3] I. Nonaka and H. Takeuchi, *The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York, USA: Oxford University Press, 1995.
- [4] J. Q. Chen, T. E. Lee, R. Zhang, and Y. J. Zhang, "Actualizing Organizational Memory with Information Systems," *Commun. ACM*, vol. 46, no. 12, pp. 73–78, 2003.
- [5] R. S. S. Guizzardi, "Agent-oriented Constructivist Knowledge Management," University of Twente, 2006.
- [6] C. A. Ellis, S. J. Gibbs, and G. L. Rein, "Groupware: Some issues and experiences," *Commun. ACM*, vol. 34, no. 1, pp. 39–58, 1991.
- [7] B. E. Dixon, R. E. Gamache, and S. J. Grannis, "Towards public health decision support: a systematic review of bidirectional communication approaches," *J. Am. Med. Inform. Assoc.*, vol. 20, no. 3, pp. 577–583, Mar. 2013.
- [8] C. A. M. Bastos, L. Rezende, M. Caldas, A. S. Garcia, S. Mecena Filho, M. L. Sanchez, and J. de L. P. Castro Junior, "Building up a Model for Management Information and Knowledge: The Case-Study for a Brazilian Regulatory Agency," in *International Workshop on Software Knowledge (SKY)*, 2011.
- [9] C. A. M. Bastos, A. C. M. Bruno, A. Garcia, L. Rezende, M. A. F. Caldas, M. L. Sanchez, and S. J. Mecena Filho, "Managing Information and Knowledge: a proposal methodology for building an integrated model based on information assets identification," in *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval and the International Conference on Knowledge Management and Information Sharing*, 2013, pp. 520–525.
- [10] 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.
- [11] G. Guizzardi, "Ontological Foundations for Structural Conceptual Models," Centre for Telematics and Information Technology, University of Twente, Enschede, The Netherlands, 2005.
- [12] G. Guizzardi, A. P. das Graças, R. S. S. Guizzardi, and A. Pinheiro, "Design Patterns and Inductive Modeling Rules to Support the Construction of Ontologically Well-Founded Conceptual Models in OntoUML," in *International Workshop on Ontology-Driven Information Systems (ODISE)*, 2011, vol. 83, pp. 402–413.
- [13] 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.
- [14] P. P. F. Barcelos, V. A. dos Santos, F. B. Silva, M. Monteiro, and A. S. Garcia, "An Automated Transformation from OntoUML to OWL and SWRL," in

- Seminário de Pesquisa em Ontologias no Brasil (ONTOBRAS)*, 2013, pp. 130–141.
- [15] OMG, “Model Driven Architecture (MDA): MDA Guide Rev. 2.0,” 2014.
  - [16] W3C, “RDF/XML Syntax Specification,” 2014.
  - [17] W3C, “OWL 2 Web Ontology Language Document Overview,” 2012.
  - [18] W3C, “SWRL: Semantic Web Rule Language - Combining OWL and RuleML,” 2004.
  - [19] W3C, “SPARQL 1.1 Overview,” 2013.
  - [20] F. Carolo and L. Burlamaqui, “Improving Web Content Management with Semantic Technologies,” in *SemTech*, 2011.
  - [21] Y. Raimond, T. Scott, S. Oliver, P. Sinclair, and M. Smethurst, “Use of Semantic Web technologies on the BBC Web Sites,” in *Linking Enterprise Data SE - 13*, D. Wood, Ed. Springer US, 2010, pp. 263–283.
  - [22] A. Swartz, “MusicBrainz: a semantic Web service,” *Intelligent Systems, IEEE*, vol. 17, no. 1. pp. 76–77, 2002.
  - [23] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, “DBpedia: A Nucleus for a Web of Open Data,” in *The Semantic Web SE - 52*, vol. 4825, K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, and P. Cudré-Mauroux, Eds. Springer Berlin Heidelberg, 2007, pp. 722–735.
  - [24] J. Hendler, “US Government Linked Open Data: Semantic.data.gov,” *IEEE Intell. Syst.*, vol. 27, no. 3, pp. 25–31, Apr. 2012.
  - [25] GISSA, “User Story - Exibir Detalhes de Risco de Óbito Materno,” 2015.
  - [26] Atiya Kazi, 2Prof. D.T. Kurian, "An Ontology Based Approach to Data Mining". IJEDR | Volume 2, Issue 4 | ISSN: 2321-9939. 2014.
  - [27] Natalya F. Noy, “Semantic Integration: A Survey of Ontology-Based Approaches”, *ACM SIGMOD Record*, vol 33, issue 4, pp 65-70, 2004.

## ANEXO 1: ESPECIFICAÇÕES DO MODELO DE INTELIGÊNCIA DO GISSA (Data Mining)

DOCUMENTOS RELACIONADOS	SEÇÃO/SUBSEÇÃO
ESPECIFICAÇÃO DE REQUISITOS NÃO FUNCIONAIS (GIS_RNF)	4.2.1, 4.3.1, 5.1.1
ESPECIFICAÇÃO DA ARQUITETURA (GIS_EAR)	4.3, 6.1, 6.3

### 1.1 ESPECIFICAÇÃO DE REQUISITOS NÃO FUNCIONAIS (GIS\_RNF)

- ✓ Subseção 4.2.1: **Acurácia da informação**
  - **Comentários:** Para a eficácia dos algoritmos de *machine learning* (aprendizagem de máquina) os dados devem ser desprovidos de erros. Os procedimentos para garantir a total acurácia da informação deverão ser analisados para verificação de corretude das mesmas.
- ✓ Subseção 4.3.1: **Tempo de Resposta Web**
  - **Comentários:** Para a realização de predições/classificações, é necessário que o tempo de resposta seja viável e em tempo real. Deverão ser investigadas abordagens de *machine learning* que atendam estes quesitos para os dados utilizados no GISS.
- ✓ Subseção 5.1.1: **Acurácia dos Dados Externos**
  - **Comentários:** Embora seja assumido que os dados estão devidamente filtrados e livres de anomalias, os procedimentos que levaram a esta suposição deverão ser devidamente analisados.

### 1.2 ESPECIFICAÇÃO DA ARQUITETURA (GIS\_EAR)

- ✓ Seção 4.3: **Análise de opções de tecnologias e arquitetura - Tecnologia/Arquitetura: Machine Learning / Data Mining**
  - **Comentários:**
    - As ferramentas citadas (Linguagem R, Weka, etc.) devem ser avaliadas para a adequação à tarefa de classificação para o GISS.
    - O tempo de resposta e precisão média dos algoritmos de Inteligência Computacional devem ser avaliados para o conjunto de dados do GISS. Além disso, o modelo de aprendizado de máquina deve possibilitar a classificação/predição em tempo real.
    - Ainda não foi descrito como o modelo de classificação/predição, construído a partir de um algoritmo de machine learning, será atualizado (a partir de novos dados) e reavaliado.
    - É necessário um estudo preliminar para a seleção de técnicas de classificação baseadas em machine learning bem como algoritmos de seleção de atributos para os dados a serem utilizados no projeto GISS.
    - É necessário o conhecimento da estrutura das outras bases de dados para poder compor a estrutura de dados da base de dados de *datamining*.
- ✓ Seção 6.1: **Item 4: "aplica mineração de dados e inferência das informações"**
- ✓ Seção 6.3: **Data Warehouse - datamining**

- o **Comentário:** *Com que frequência esta base será atualizada com novos dados? Os dados desta base serão utilizados para construir diretamente os modelos de Inteligência Computacional ou necessitarão de alguma preparação?*
- ✓ Seção 6.4: " O Componente de **Inteligência** deverá suportar diversas algoritmos de mineração de dados, ontologia, estatística e outras mecanismos de inferência."

## 2. REQUISITOS DO BACKLOG (ASSOCIADOS À INTELIGÊNCIA COMPUTACIONAL)

INFRA	Montagem de Datawarehouse
ETL	Extração de dados <u>DataSUS</u>
ETL	Extração de dados de Sistemas Locais
ETL	Extração de dados de contexto externo
ETL	Motor de Transformação de dados

**Observação:** Não há no backlog um requisito que esteja diretamente vinculado à incorporação dos algoritmos de inteligência computacional.