


ORGANIZAÇÃO DE UM DATA WAREHOUSE CLÍNICO

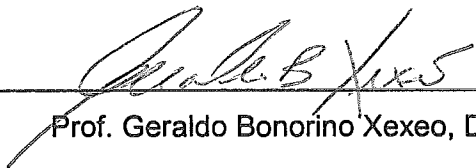
Ana Cláudia Oliveira Garcia dos Santos

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

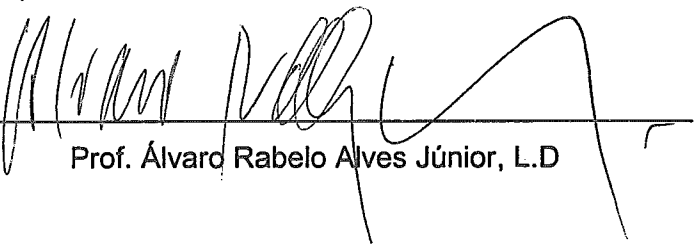
Aprovada por:



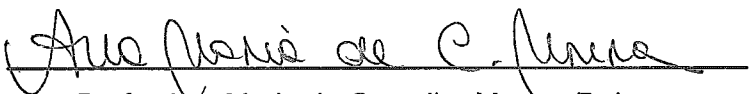
Profa. Ana Regina Cavalcanti da Rocha, D. Sc.



Prof. Geraldo Bonorino Xexeo, D. Sc.



Prof. Álvaro Rabelo Alves Júnior, L.D



Profa. Ana Maria de Carvalho Moura, D. Ing.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2001

SANTOS, ANA CLÁUDIA O. GARCIA DOS

Organização de um *Data Warehouse*
Clínico [Rio de Janeiro] 2001

X, 104 p. 29,7 cm (COPPE/UFRJ,
M.Sc., Banco de Dados, 2001)

Tese - Universidade Federal do Rio de
Janeiro, COPPE

1. *Data Warehouse*

2. Sistemas de Informação Médica

3. Cadeias de Valor

I. COPPE/UFRJ II. Título (série)

À minha mãe

Agradecimentos

Ao meu orientador, Geraldo Xexéo, pelas pertinentes idéias e por me dizer que a perfeição não existe.

À Ana Regina, pelo carinho e preocupação de mãe e essencial participação neste trabalho.

Ao Dr. Álvaro Rabelo, pelo espaço para a realização deste trabalho e pelo exemplo de competência.

Ao Dr. Ximenes, pela sua simpática disponibilidade para a avaliação dos textos médicos.

À Dra. Lísia Rabelo e Dr. Luis Carlos Passos, pela colaboração mesmo nos momentos finais de suas teses.

À todos da Fundação Bahiana de Cardiologia, em especial ao NPqD, pelo apoio e a Ronny pela indispensável participação neste trabalho.

Ao Dr. Maurício Blanche, pelas explicações dos conceitos médicos

Às meninas da COPPE, pela atenção e participação nos passos da burocracia

Ao meu irmão, Beto, pela força nos desesperos dos "deadlines".

À minha irmã, Ana Luíza, pelo entusiasmo da minha conquista.

À Doralina, por nunca esquecer da minha tese.

A todos os amigos que contribuíram perguntando pela minha tese.

À minha amiga Kátia, pelas constantes cobranças e valiosas dicas e a Nicolas pelo incentivo amigo.

Ao Be, por todos os conselhos carinhosos.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

ORGANIZAÇÃO DE UM *DATA WAREHOUSE* CLÍNICO

Ana Cláudia Oliveira Garcia dos Santos

Março/2001

Orientador: Geraldo Bonorino Xéxeo

Programa: Engenharia de Sistemas e Computação

O processamento de informações de uma organização ocorre com o objetivo de agregar valor ao negócio ao qual se propõe. Com isto, em sua situação do dia a dia, um profissional se executa atividades ou toma decisões. Isto pode ocorrer em diversos setores da organização e em diferentes tipos de organização. Ambientes com características de *data warehouse* surgiram para apoiar atividades de tomada de decisão e de avaliação do processamento de informações de um certo domínio. Estes conceitos podem ser aplicados à área médica direcionando os objetivos aos diferentes grupos que atuam numa instituição de saúde.

Este trabalho apresenta uma experiência na organização de um *data warehouse* clínico, acompanhando os diversos níveis de análise de dados que podem ser encontrados em uma instituição de saúde.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

A CONCEPTUAL FRAMEWORK FOR CLINICAL DATA WAREHOUSES

Ana Cláudia Oliveira Garcia dos Santos

March/2001

Advisor: Geraldo Bonorino Xéxeo

Department: Computer Science and System Engineering

The information process of an organization aims to aggregate some value to the business it about to propose. In this sense, a professional is faced with activities that claim decision analysis or also with the need to execute them. This situation can occur at the various departments in the organization and at different kinds of them. Environments based on data warehousing came to support decision-making activities and evaluation of information process of a certain domain. These concepts can be applied at the health care area driving the requirements to the different groups that work in this kind of institution.

This thesis shows an experience in the construction of a conceptual framework for a clinical data warehouse, following the diversity of the level of data analyses that can be found in a healthcare institution.

Índice

Capítulo 1. Introdução

1.1. Motivação.....	1
1.2. Objetivo do trabalho	1
1.3. Organização do trabalho	3

Capítulo 2. Ambientes de Dados Analíticos

2.1. Aspectos gerais	5
Breve histórico	5
Definições	6
Estrutura	7
Metadados	12
2.2. Arquitetura do ambiente	13
2.3. Desenvolvimento	16
O ciclo de vida	18
2.4. Requisitos.....	23
2.5. Tecnologia.....	27
Características incorporadas aos SGBDs.....	28
2.6. Considerações finais	30

Capítulo 3. Ambientes Médicos de Análise

3.1. Aspectos Gerais	31
3.2. Sistemas na área médica	33
3.2.1. Sistemas de Informação	33
3.2.1.1. Recuperação da informação	35
3.2.2. Sistemas de Apoio à Decisão em Medicina	36
3.3. Características de ambientes médicos de análise	38
3.3.1. Repositório de dados clínicos	39

3.3.2. <i>Data warehouse</i> clínico	40
3.3.3. Trabalhos relacionados	42
3.4. A Questão da Segurança	44

Capítulo 4. Processo para a Organização das Informações de *Data Warehouse*

4.1. Análise sobre a construção de <i>data warehouse</i>	47
4.2. Processo para organização das informações de um <i>data warehouse</i> : uma proposta com enfoque nos requisitos.....	49
4.2.1. Considerações do processo.....	61
4.3. Considerações Finais.....	62

Capítulo 5. Experiência na Construção de um *Data Warehouse* para Cardiologia

5.1. Os sistemas de informação da FBC/UCCV	63
5.2. Processo de transformação da informação médica	65
5.3. Experiência na organização das informações de um <i>data warehouse</i> para Cardiologia.....	68
5.3.1. Aplicação do processo	69
5.3.1.1. Definição das fontes de dados	69
5.3.1.2. Preparação para a análise	72
5.3.1.3. Levantamento dos requisitos	73
5.3.1.4. Negociação dos dados.....	75
5.3.1.5. Especificação dos requisitos	76
5.4. Descrição do ambiente	76
5.4.1. Arquitetura do ambiente	76
5.4.2. O <i>FBCDataWare</i>	77
5.5. Considerações Finais	85

Capítulo 6. Conclusão

Referências

Anexo 1

Anexo 2

Índice de tabelas

Capítulo 2

Tabela 2-1 - Modelagem dimensional de dados	9
Tabela 2-2 - Ciclos de vida inversos	18
Tabela 2-3 - Avaliação das abordagens de desenvolvimento do <i>data warehouse</i>	22
Tabela 2-4 - Avaliação dos requisitos para construção de um ambiente de <i>data warehouse</i>	26

Capítulo 4

Tabela 4-1 - Atividades do processo de análise de requisitos.....	60
---	----

Capítulo 5

Tabela 5-1 - Atividades da cadeia de valor	67
Tabela 5-2 - Cadeias de valor nos níveis de atuação	68
Tabela 5-3 - Avaliação dos sistemas disponíveis.....	71
Tabela 5-4 - Definição de áreas de interesse e de negócio	72
Tabela 5-5 – Grupos de negócio e dados de interesses	74

Índice de Figuras

Capítulo 2

Figura 2-1 - Arquitetura genérica de um *data warehouse* 13

Figura 2-2 - Ciclo de vida do desenvolvimento de um *data warehouse* 20

Capítulo 3

Figura 3-1 - Disposição da informação em *data warehouse* clínico 41

Capítulo 4

Figura 4-1 - Fases da etapa de análise de requisitos 50

Figura 4-2 - Intervenção gerencial na negociação 62

Capítulo 5

Figura 5-1 - Processo de transformação da informação médica 66

Figura 5-2 - Cenário da negociação 75

Figura 5-3 - Arquitetura do FBCDataWare 77

Figura 5-4- Modelo para a definição de perfis 78

Figura 5-5 - FBCDataWare - Tela inicial - Acesso aos grupos 79

Figura 5-6 - FBCDataWare - Tela de Login - Autenticação no grupo 79

Figura 5-7 - FBCDataWare - Tela inicial do grupo 79

Figura 5-8 – FBCDataWare - Apresentação consultas definidas para o grupo 80

Figura 5-9 - FBCDataWare - Tela de Definição de consultas 81

Figura 5-10 - FBCDataWare - Apresentação do resultados da consulta em tabela 82

Figura 5-11 - FBCDataWare - Apresentação do resultados da consulta em gráfico ... 83

Figura 5-12 - FBCDataWare - Controle de usuários do grupo 84

Figura 5-13 - FBCDataWare - Administração de Grupos e Perfis 85

Introdução

1.1. Motivação

O processamento das informações de uma organização tem como finalidade atribuir ao negócio um certo valor aos objetivos aos quais se propõe. Este processamento é apoiado pelas diversas atividades existentes na organização, o que leva à necessidade de serem gerenciadas, identificando-se as cadeias de valor que podem ser encontradas e analisando-se os pontos críticos do processo.

A informatização dos processos contribui para o desempenho das atividades e torna, por muitas vezes, mais evidente as necessidades de intervenção gerencial. Com um maior conhecimento das informações que possuem, os usuários criam necessidades diversas relacionadas ao acesso e à manipulação destes dados. Dados armazenados, apenas, com o objetivo de registro das operações do negócio, passam a ser um importante objeto no auxílio às atividades gerenciais.

Os ambientes com enfoque em análise de dados são, em sua maioria, organizados a partir de objetivos genéricos e muito amplos dentro da organização. A motivação deste trabalho está, principalmente, na necessidade de organizar um ambiente com caráter gerencial direcionado aos diversos níveis de análise de dados que podem ser definidos em uma organização.

1.2. Objetivo do trabalho

Este trabalho tem como objetivo a organização de um *data warehouse* (DW) clínico acompanhando os níveis de informação de uma instituição de saúde. Estes níveis de informação são determinados pelas cadeias de valor da instituição e os grupos de negócio que dela fazem parte.

O local de experimentação deste trabalho foi a Fundação Bahiana de Cardiologia (FBC)/ Unidade de Cardiologia e Cirurgia Cardiovascular (UCCV), uma das Unidades do Hospital Universitário Prof. Edgard Santos (HUPES) da Universidade Federal da Bahia (UFBA). Na FBC/UCCV é encontrada uma diversidade de sistemas responsáveis por diversos tipos de informação. Cada sistema possui necessidades operacionais diferentes e, por consequência, disponibiliza as informações em diferentes bases de dados. Por questões gerenciais e de pesquisa, muitas vezes estes dados precisam ser analisados de forma comum, o que leva à necessidade de integração das bases de dados facilitando a recuperação dos dados disponíveis na instituição.

Para atender às expectativas de análise das informações geradas nos ambientes operacionais, os dados precisam ser reorganizados, consolidados e agregados, ou seja, devem ser realizados procedimentos diferentes daqueles para os quais estão preparados. Foram constatadas algumas dificuldades no acesso às informações operacionais já evidenciadas por DAPHNE (1998):

- A recuperação dos dados específicos leva muito tempo e ocorre freqüentemente com um certo retardo;
- Atividades de análise desviam o andamento das operações diárias e interferem no desempenho das transações;
- Os dados recuperados das bases de dados operacionais são irreconhecíveis e listados em linhas 'secas', isto é, não estão em formato facilmente inteligível;
- Os dados armazenados sofrem atualizações constantes e são pouco consistentes para análises a longo prazo;
- A estrutura em que os dados estão dispostos dificulta o processamento de consultas *ad hoc*.

Diante destas características, torna-se evidente a necessidade de um outro ambiente com o objetivo de atender às atividades analíticas, diferente do ambiente operacional voltado para o processamento de transações e operações do cotidiano de uma instituição. Um *data warehouse* é tratado como um ambiente que exige uma organização dos dados destinada a proporcionar um bom desempenho no processamento de consultas, tornando-se um grande repositório de dados extraídos das bases operacionais.

Uma instituição de saúde caminha basicamente em três níveis de informação: (i) o nível do Paciente onde a doença do indivíduo é colocada em evidência para que lhe seja dado um melhor tratamento, também visto como medicina curativa; (ii) o nível da Comunidade analisado através de pesquisas epidemiológicas em casos da instituição com o objetivo de melhorar a qualidade dos tratamentos e avaliar os índices de confiabilidade da instituição também visto como medicina preventiva; e (iii) o nível da Instituição para avaliar gastos da instituição e promover melhoramentos.

Estes níveis de informação são responsáveis por guiar a organização do *data warehouse*. Cada grupo de negócio encontrado na instituição possui os seus interesses de análise definidos referentes a cada um desses níveis. Este cenário possibilita a organização de um *data warehouse* com uma organização de dados específica para cada área do negócio, podendo ser visto, desta forma, como um conjunto de *data marts*.

1.3. Organização do trabalho

A organização de um *data warehouse* deve partir de uma forte determinação para avaliação das atividades, processos e resultados gerados pelos dados que circulam na instituição. Um dos fundamentos deste trabalho foi estabelecer a necessidade de um ambiente para análises dentro do fluxo de informações da instituição, descrevendo o processo de desenvolvimento e as implicações de um ambiente como este na área de saúde.

O capítulo II descreve o surgimento de ambientes desta categoria em áreas diversas, os métodos e processos utilizados para desenvolvê-los e as características incorporadas à tecnologia existente.

O capítulo III trata dos sistemas encontrados na área médica, características restritas aos dados destes sistemas e alguns trabalhos relacionados a ambientes para análise de informações nesta área.

O capítulo IV descreve o fluxo de informações em uma instituição de saúde e o processo utilizado neste trabalho para organização de um *data warehouse* na área médica.

O capítulo V descreve os sistemas encontrados na instituição de trabalho, a experiência de construção de um ambiente para análise de informações na FBC/UCCV, e a versão do produto gerada neste trabalho: o FBCDataWare.

Por fim, no capítulo VI, são descritas as conclusões do trabalho, suas contribuições e as perspectivas para futuros trabalhos.

2. Ambientes de Dados Analíticos

O objetivo deste capítulo é apresentar os principais aspectos que envolvem o desenvolvimento e a tecnologia de um data warehouse, sendo dada uma ênfase maior ao processo de desenvolvimento.

2.1. Aspectos gerais

Breve histórico

Antes do *data warehouse* receber esta denominação e se tornar um ambiente de trabalho, notava-se uma tendência de extrair dados do ambiente operacional e transformá-los em objetos de análise fora do ambiente de produção. Isto acontecia porque o alto nível de detalhe e a estrutura de armazenamento na base de produção dificultava o acesso e análise dos dados. A idéia básica, então, era retirar os dados da linha de produção sem afetar o desempenho das aplicações operacionais com a análise coletiva dos dados e obter maior controle sobre os dados que agora estão fora do ambiente de transações on line livre para agregações e resumos (INMON, 1997).

À medida que esta atividade tornou-se comum, percebeu-se mais claramente uma projeção do ambiente de produção para um ambiente de análise ou um ambiente de apoio à decisão. Os dados extraídos da base de produção eram armazenados em dispositivos diversos e utilizados para um fim definido com o auxílio de outras ferramentas. Era necessário, então, que fossem armazenados em uma base de dados comum e fosse disponibilizada uma aplicação com suporte aos requisitos de análise dos usuários. Surgiu daí a denominação de um grande repositório de dados.

Desta forma, as diversas fontes de dados operacionais de uma empresa podem ser integradas em um só repositório, disponíveis para cruzamento de informações e proporcionando, possivelmente, uma visão geral de todo o negócio.

Definições

Com a materialização deste conceito muitos trabalhos tiveram como objetivo definir e implementar um *data warehouse* adequado aos seus interesses. Mattison (1997), por exemplo, aplica estes princípios na área de telecomunicações e afirma ser possível aplicá-los a qualquer tipo de negócio. Este autor define um *data warehouse* simplesmente como uma coleção de dados copiados de outros sistemas e montados em um local diferente disponível para os propósitos do usuário final em atividades de apoio à tomada de decisão e coleta de informações. Como se pode ver, Mattison não deixa claro o tipo de fonte de onde são transportados os dados para o *data warehouse*.

Uma outra definição, citada por GILL&RAO (1996), trata o *data warehouse* como um processo contínuo que combina dados de múltiplas fontes heterogêneas incluindo dados históricos e agrupados, segundo alguma política, para apoiar as necessidades de consultas estruturadas e *ad hoc*, relatórios analíticos e apoio à tomada de decisão. Esta definição ainda é deficiente quanto à escolha dos dados capturados tendo em vista que nem sempre todos os dados operacionais disponíveis são necessários para uma análise específica.

Finalmente, segundo INMON (1997), considerado o pai do tema, *data warehouse* é um conjunto de dados orientado por assuntos, integrado, não-volátil e variável no tempo, de apoio às decisões gerenciais. Esta é uma definição robusta pois trata da categoria dos dados armazenados e do comportamento do ambiente de *data warehouse*.

Fica claro que as diversas definições se completam e que, de uma certa forma, cada uma dá um enfoque maior em alguns aspectos. Apesar das diferentes visões em torno de uma definição para um *data warehouse*, a idéia central se mantém: dar suporte às atividades decisórias e gerenciais de uma instituição disponibilizando as informações corretas de maneira mais flexível.

As características mencionadas por Inmon, entretanto, são mais objetivas e merecem ser destacadas:

A) orientado por assuntos

A base de dados do *data warehouse* é organizada sob a perspectiva do usuário final. Diferente das bases operacionais que são organizadas em torno das transações da instituição, para o *data warehouse* o usuário elege um tema e as informações extraídas do ambiente operacional giram em torno desse tema.

B) integrado

Como os dados de um *data warehouse* têm sua origem em bases de dados distintas, é necessário manter a consistência e a uniformidade dos dados na base do *data warehouse*. Dado do tipo sexo, por exemplo, pode estar definido em uma fonte como (X, Y) e em outra como (M, F); no *data warehouse* eles devem ser convertidos para uma única apresentação.

C) não-volátil

Os dados em um *data warehouse* não são atualizáveis, eles são apenas carregados e disponibilizados para consultas após serem transformados e integrados. Diferente do ambiente operacional onde os dados são alterados a todo instante requerendo múltiplas transações e operações de controle para manter a integridade e consistência dos dados, estas atividades são dispensadas pelo *data warehouse*.

D) variável no tempo

Cada registro se refere a um dado instante do objeto. Um objeto que tenha seu estado alterado na base operacional, gera um nova entrada no *data warehouse* referente ao instante da transferência. Este comportamento exige, sempre, um atributo de tempo em cada registro e faz com que o *data warehouse* tenha um tempo válido muito maior do que as bases operacionais.

Estrutura

Antes do surgimento dos bancos de dados relacionais, os dados eram armazenados sem nenhum tratamento prévio. Eram relatórios inteiros e formulários, cada um representado por um registro. Além de gerar um grande volume de dados, esta organização promovia a inconsistência e a falta de integridade dos dados devido, principalmente, ao alto grau de repetição das informações e à falta de controle na inserção.

Com o objetivo de eliminar a redundância dos dados armazenados, surgiu o modelo entidade – relacionamento com tabelas separadas e dependentes. Esta técnica trouxe um grande benefício no processamento de transações, que passaram a ser feitas de maneira simples e determinística. Em compensação, os sistemas de recuperação e manipulação de dados tornaram-se complexos devido à necessidade de um maior controle no relacionamento entre as tabelas (KIMBALL, 1996c, MATTISON, 1997).

A deficiência na capacidade de responder questões analíticas típicas, em tempo hábil e de maneira prática, levou à necessidade de se estruturar a base de dados de modo mais específico para atender a este tipo de consulta. A normalização de tabelas, que fora um benefício para as aplicações habituais, passou a ser um empecilho para a praticidade e produtividade das atividades gerenciais, o que fez surgir o conceito de dados multidimensionais.

A modelagem multidimensional dos dados é uma técnica de projeto lógico, mais simples, que procura representar os dados em um padrão intuitivo, descrevendo o negócio de forma mais acessível ao usuário. A idéia de análise multidimensional é induzida pela necessidade de se avaliar a informação através de diferentes perspectivas, simulando um cubo com n dimensões, cada uma representando as diferentes formas sob as quais o usuário deseja visualizar o dado. Estas perspectivas, por sua vez, giram em torno de um tema central denominado fato.

Estes são os componentes básicos de um modelo multidimensional: fatos, dimensões e atributos particulares. Os fatos são representados por tabelas maiores, indexadas por chaves compostas. As dimensões são representadas por um conjunto de tabelas menores, chamadas tabelas dimensão, ligadas a uma tabela de fato. Cada tabela de dimensão corresponde, exatamente, a um dos componentes da chave composta da tabela fato (AGOSTA, 1999, INMON, 1997, KIMBALL, 1996c).

Os atributos das dimensões compõem a tabela dimensão além do código de identificação e enriquecem o detalhamento de uma determinada visão do fato. Através deles, é possível se chegar à hierarquia das dimensões, descrições detalhadas de cada elemento da dimensão e propriedades estendidas. Os atributos dos fatos são os valores tratados para suprir as necessidades de análise através de médias, agregações ou índices. Por esta razão, na maioria dos casos, são chamados de métricas. A natureza destes atributos define a categoria do fatos. Desta forma, os fatos são aditivos se for possível somar os seus atributos a partir de qualquer dimensão escolhida para agrupá-los; são semi-aditivos, se isto só for possível a partir de certas dimensões e, são não-aditivos, se o somatório dos fatos não fizer sentido em dimensão alguma (KIMBALL, 1996c).

Para este tipo de modelo, algumas das regras do modelo relacional devem ser ignoradas, embora outras sejam fundamentais. A tabela a seguir (Tabela 2-1) apresenta as etapas do desenho do modelo e os respectivos produtos diretos ou indiretos.

Tabela 2-1 - Modelagem dimensional de dados

Etapas	Produto
Definir o tema de interesse estratégico ou área de negócio que se deseja medir.	O fato (escopo do modelo)
Avaliar os processos e o contexto que envolve o tema	As dimensões Os componentes da chave da tabela de fato
Definir a granularidade do fato, ou seja, o que cada registro na tabela de fato representa em relação ao tema	O nível de consolidação do fato ou a unidade de tempo que será tratada (a frequência de atualização da base)
Definir a granularidade das dimensões	Os atributos das tabelas de dimensão
Definir os atributos da tabela de fato	Os atributos calculados da tabela de fato (Categoria da tabela de fato)

Num modelo multidimensional, existe pelo menos uma tabela para cada dimensão. O esquema descrito constitui-se no modelo básico e é chamado de estrela (*star schema*) devido à presença de uma tabela dominante (tabela fato), no centro do modelo, interligada às diversas tabelas de dimensão. Existe, também, o esquema flocos de neve (*snow flake schema*) que propõe a normalização das tabelas dimensão, tornando as pontas da estrela do esquema anterior, centros de novas estrelas.

O modelo estrela é o mais utilizado pela simplicidade e pela eficiência satisfatórias. Já o esquema flocos de neve não é aconselhável, devido ao aumento da complexidade do modelo para o usuário final, sem muita relevância no espaço de armazenamento adquirido (KIMBALL, 1997a).

De uma certa forma, pode-se dizer que houve um regresso aos tempos dos modelos pré-relacionais trazendo a redundância dos dados e os campos calculados. Isto seria verdade, se não fosse usada a experiência com o modelo relacional e o processamento de transações, para desnormalizar apenas o necessário, associada ao fato da carga de dados trazida da base operacional ser restrita aos interesses da análise.

Data warehouse departamental

Diante da necessidade de se obter domínio de dados do ambiente operacional para análises específicas, observa-se uma tendência crescente dos diversos departamentos das empresas em organizar sua própria base de dados, limitada aos seus interesses. Estes pequenos ambientes de análise são chamados de *data marts* e são caracterizados por envolver uma área específica do negócio restrito a apenas um grupo de trabalho.

Os modelos de dados de *data marts* podem ser usados independentes, integrados a outro *data mart* ou ligados em conjunto na construção de um *data warehouse* completo. Inmon (1997) considera o *data mart* como nível departamental dentro do seu ambiente projetado, que é o ambiente que envolve todos os dados disponíveis em seus respectivos níveis de informação: operacional, atômico (*data warehouse*), departamental (*data mart*), individual. Desta forma, o nível departamental seria derivado do nível atômico. Sigal in (BONTEMPO&ZAGELow, 1998) defende que os *data marts* são frutos exclusivos de um *data warehouse* e devem ser criados quando forem realmente necessários, devendo-se evitar principalmente a criação de um *data warehouse* virtual a partir de vários *data marts*. Isso se deve ao fato de que é necessário um *data warehouse* real quando se deseja um bom desempenho em funcionalidades cruzadas e análises de tendências.

Muitas empresas optam por começar com um ambiente departamental e, só após obter um retorno de seus usuários, expandi-lo. Esta decisão se deve ao fato de que, quanto maior o escopo, mais valor o *data warehouse* tem para a empresa e mais cara e trabalhosa é a sua criação e manutenção, o que não acontece com os *data marts* por serem menores. Seguindo esta linha, *data marts* também podem ser criados como subconjuntos de *data warehouse*, em busca de autonomia, melhor desempenho e simplicidade de compreensão (KIMBALL, 1998a).

(KIMBALL et al, 1998b) defende a idéia de que os *data marts* devem ser considerados fatias da grande torta que é o *data warehouse*, mas a sua organização não deve estar cega para a tendência de união destes subconjuntos. Ou seja, deve haver uma conformidade na definição dos fatos e dimensões na criação dos *data marts* para que seja possível o surgimento de um *data warehouse*.

Povoamento do *Data Warehouse*

O povoamento do *data warehouse* deve ser feito passo a passo e incrementado a cada dia, assumindo assim um comportamento evolutivo. Para ilustrar este processo, Inmon (1997) descreve o fenômeno dia 1 - dia N que acontece da seguinte forma: no dia 1, há um grupo de sistemas efetuando basicamente processamento operacional; no dia 2, as primeiras tabelas da primeira área de interesse no *data warehouse* são povoadas despertando a curiosidade de alguns para o processamento analítico; no dia 3, mais áreas do *data warehouse* são povoadas e com isso surgem mais usuários; no dia 4, à medida em que mais áreas do *data warehouse* vão sendo povoadas, alguns dos dados que residiam no ambiente operacional são colocados de forma apropriada no *data warehouse*. O *data warehouse* está sendo agora descoberto

como uma fonte integrada para o processamento analítico e a competição para se chegar ao *data warehouse*, torna-se um obstáculo para sua utilização. No dia 5, bases de dados departamentais começam a florescer, devido ao fato observado de que é mais fácil e mais barato efetuar o respectivo processamento trazendo os dados do *data warehouse* para os seus próprios ambientes departamentais de processamento; no dia 6, acontece a corrida aos sistemas departamentais. Os usuários vão gradativamente abandonando os detalhes do *data warehouse* devido aos atrativos para se obter os dados departamentais. Finalmente, no dia N, a arquitetura encontra-se plenamente desenvolvida, o *data warehouse* já adquiriu uma certa força e para os sistemas de produção só resta o processamento operacional. É importante salientar que durante esta evolução, que pode levar anos, o ambiente de apoio à tomada de decisão está em plena atividade.

Granularidade do *Data Warehouse*

Uma questão fundamental no projeto de um *data warehouse* refere-se ao nível de detalhe em que as unidades de dados serão mantidas. Este é um aspecto importante porque afeta, diretamente, o volume de dados armazenados no *data warehouse* e, ao mesmo tempo, o tipo de consulta que pode ser respondida. Quanto maior o nível de detalhes, menor o nível de granularidade. À medida em que o nível de granularidade se eleva, há uma correspondente diminuição da possibilidade de utilização dos dados para responder às consultas, ainda que os dados se tornem muito mais compactos. O uso dos dados com alto nível de granularidade é muito mais eficiente se estes apresentam o detalhamento suficiente. Este ponto de equilíbrio deve ser analisado, cuidadosamente, no início do projeto e construção do *data warehouse*.

Particionamento do *Data Warehouse*

Um outro aspecto importante no projeto de um *data warehouse* diz respeito à representação das dimensões. O particionamento é descrito como a repartição dos dados em unidades físicas separadas, que podem ser tratadas de forma independente. Esta tarefa é implementada através da quebra de uma tabela de dados operacionais em duas ou mais tabelas no *data warehouse* ou da separação das linhas de uma tabela e armazenamento em tabelas separadas, como, por exemplo, a divisão de tabelas por fase de ocorrência (*vendas_1998*, *vendas_1999*). Esta tarefa é representada no *data warehouse* como parte da chave da tabela de fato.

Metadados

Definidos como dados sobre os dados, os metadados são responsáveis pela utilização produtiva do *data warehouse*. É através dos metadados que é possível se saber o que são e onde estão os dados armazenados bem como determinar o padrão das definições dos dados a serem armazenados (INMON, 1997, KIMBALL et al, 1998b, MARCO, 1998).

Basicamente, os metadados mantêm informações sobre (GILL&RAO, 1996, KIMBALL et al, 1998b):

- Definição física do dado armazenado (visão do programador);
- Definição conceitual do dado (visão do analista ou usuário);
- Fonte de origem do dado;
- A transformação sofrida no momento de sua migração do ambiente operacional;
- Estrutura do modelo de dados;
- Relacionamento entre o modelo de dados e o *data warehouse*;
- Histórico de extrações.

Para dados legados e externos, os metadados podem incluir também uma descrição resumida, a localização de armazenamento e parâmetros de acesso e controle do documento que contém os dados (KIMBALL et al, 1998b).

Por vários motivos, os metadados assumem um papel mais importante no *data warehouse* do que no ambiente operacional. Neste, o tratamento dos metadados é deixado para segundo plano no nível da documentação, e este procedimento não prejudica o andamento dos sistemas. Já o *data warehouse* possui um processo de desenvolvimento heurístico e iterativo, que leva à necessidade de se manter os metadados sempre exatos e atualizados para garantir a eficácia do processo.

Outro fator de relevância diz respeito ao mapeamento dos dados entre os dois ambientes. Quando os dados são passados do ambiente operacional para o *data warehouse*, sofrem diversas transformações para adequação ao modelo. O registro destas transformações é de grande importância nas atividades de rastreamento dos dados do *data warehouse* para o sistema operacional. Além disso, existe a importância associada ao tempo em que os dados permanecem armazenados. Não é incomum em um *data warehouse* que as estruturas de dados sejam alteradas com

uma certa regularidade, o que torna extremamente necessário manter o acompanhamento destas alterações.

2.2. Arquitetura do ambiente

A impressão mais importante obtida no estudo do *data warehouse* talvez seja o reconhecimento da existência de dois ambientes de trabalho distintos: o ambiente dos sistemas em operação e o ambiente dos sistemas para informação (ORR, 1997). Até então foram apresentadas as características da base de dados integrada que compõe este último ambiente e que representa o *data warehouse* propriamente dito. Entretanto, esta base está inserida em um contexto mais amplo, que dá o alicerce e os subsídios necessários para a sua criação e manutenção. A abstração desta arquitetura é essencial para descrever os componentes envolvidos.

Para muitos, o *data warehouse* é mais considerado como um processo de desenvolvimento do que como um produto (SEN&JACOB, 1998), partindo-se do princípio de dar a idéia de que o produto estaria finalizado por completo após ter sido implementado o seu primeiro modelo. Mas será visto adiante, com mais detalhes, que o ciclo de desenvolvimento se renova constantemente devido à necessidade de se incorporar novos requisitos. Em conseqüência disto, cada camada da arquitetura também deve ser revista para atender às necessidades do usuário.

Uma arquitetura genérica pode ser definida com base nas características apresentadas e nos requisitos impostos para um ambiente composto por um *data warehouse*. As camadas básicas são bem definidas e descritas a seguir (Figura 2-1).

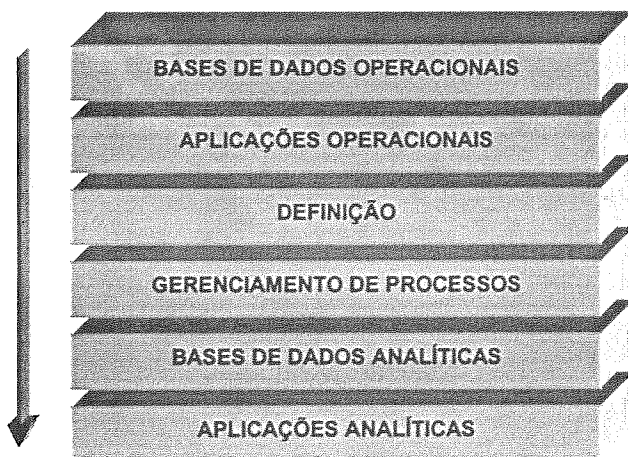


Figura 2-1 Arquitetura genérica de um *data warehouse*

Bases de dados operacionais

Esta camada representa todas as fontes de dados responsáveis pela carga do *data warehouse*. Estas fontes são encontradas nas mais diversas formas, informatizadas ou não, e podem conter dados das seguintes categorias (GILL&RAO, 1996): (i) **dados de produção**, informações das bases de dados operacionais que podem ser encontradas em diferentes plataformas: relacionais, não-relacionais, arquivos etc; (ii) **dados legados**, informações de importância histórica geradas por sistemas que já estão fora de uso, mas que tem seus dados mantidos apenas para estes objetivos (fora das aplicações operacionais, estes dados estão arquivados em dispositivos externos ou em bancos fora de linha); (iii) **sistemas internos de escritório**, dados de formulários não-eletrônicos, relatórios, planilhas, gráficos, documentos, são úteis para apoiar análises através de departamentos; e (iii) **dados externos**, informações que não são do controle, propriedade ou operação da instituição (artigos, planilhas, relatórios utilizados como dados comparativos). É importante ressaltar que todos estes dados só farão sentido no *data warehouse* acompanhados de marcas de tempo.

Aplicações operacionais

Refere-se a todas as aplicações e ferramentas que geram dados para as fontes operacionais, sejam elas os sistemas de gestão da empresa, sistemas legados, planilhas eletrônicas ou documentos. Esta é a camada com a qual o usuário final interage para as atividades do dia a dia e é através dela que toma conhecimento das informações disponíveis na empresa. Em relação à arquitetura, ela existe para compor o cenário mas, neste contexto, interessam os dados que são gerados e não a forma como isto é feito.

Definição

A tarefa de transferir dados de uma base para outra, requer uma série de definições para assegurar que o conteúdo de uma base será consistente com o conteúdo da outra, principalmente quando os dados partem de bases heterogêneas para uma base integrada. Desta forma, esta camada diz respeito a todas as tarefas destinadas à organização do *data warehouse*: definição das dimensões que serão tratadas e as respectivas tabelas e atributos a serem criados, agregações e resumos que devem ser feitos, processos necessários para a transferência dos dados e manutenção do *data warehouse*. Também são definidas as normas de padronização e os procedimentos de gerência dos metadados.

Gerenciamento de Processos

Os processos definidos na camada anterior precisam ser executados tanto para a construção do *data warehouse* como para manutenção da base de dados integrada. O *data warehouse* deve estar sempre sendo atualizado tanto no que se refere ao conteúdo quanto à estrutura, pois as mudanças realizadas na base de dados operacional devem ser refletidas na base que está sendo construída. Esta camada é responsável por manter a seqüência dos processos a serem executados, como também por gerenciar as atividades de transporte e distribuição entre os diversos ambientes, transações e processamento de mensagens.

Base de dados analíticos

Esta camada envolve a base de dados integrada que armazena os dados operacionais da instituição para fins analíticos, o *data warehouse*. A depender da estrutura desenvolvida pode ser criada apenas uma base com todo o conteúdo analítico, pequenas bases específicas ou uma combinação destas. Estas diferentes composições dizem respeito à distribuição física do *data warehouse*, que neste aspecto pode ser classificado da seguinte maneira (INMON, 1997):

- **Virtual:** consiste em simplesmente prover os usuários finais com facilidades adequadas para extração das informações diretamente dos bancos de produção, não havendo assim redundância pois não existe o recarregamento dos dados. Esta opção pode, no entanto, sobrecarregar o ambiente operacional;
- **Centralizado:** consiste de um único banco de dados físico, contendo todos os dados para uma área funcional específica, um departamento ou uma empresa, devendo ser usado onde existe uma necessidade comum de informações. Um *data warehouse* central, normalmente, contém dados oriundos de diversos bancos operacionais, devendo ser carregado e mantido em intervalos regulares;
- **Distribuído:** como o próprio nome indica, possui seus componentes disponíveis em diferentes bancos de dados físicos. Em geral possui um alto grau de redundância e, por conseqüência, procedimentos complexos para carga e manutenção;
- **Em camadas:** esta estrutura considera o armazenamento dos dados em diferentes servidores de acordo com o nível de detalhes. Desta forma são distribuídos os dados altamente resumidos, os dados com um nível de detalhe intermediário e os dados mais detalhados ou atômicos. A maior parte das consultas é feita no nível mais resumido o qual deve ser otimizado para atender uma carga maior de

usuários e um baixo volume de dados, enquanto que para os outros níveis são reservadas as consultas mais pesadas.

Aplicações analíticas

Nesta camada, encontra-se a parte do ambiente de apoio às atividades gerenciais, destinado a tornar as informações analíticas disponíveis para o usuário, através de aplicações adequadas para prover as atividades de consulta e manipulação dos dados armazenados no *data warehouse*. É importante lembrar que esta tecnologia coloca à disposição dos usuários finais, artifícios com os quais eles não estão habituados a trabalhar. Desta forma a interface dos sistemas deve ser de fácil adaptação às atividades a que se propõe.

O Instituto Ken Orr (ORR, 1996) propôs uma arquitetura baseada em camadas que representam toda a estrutura de dados, comunicação, processamento e apresentação. Esta estrutura é utilizada no desenvolvimento dos projetos do instituto.

Gill&Rao (1997), propuseram uma arquitetura de referência que divide os componentes do ambiente de *data warehouse* em blocos que estão relacionados às funcionalidades específicas do *data warehouse*, e camadas que representam o ambiente necessário para implementar os blocos.

Mattison (1997) defende que a construção do sistema em camadas é uma maneira de ganhar eficiência e otimizar os esforços na organização de um *data warehouse*. Desta forma, uma infra-estrutura física e operacional inicial é construída para apoiar a evolução do desenvolvimento. E, com o amadurecimento do ambiente, os custos com o desenvolvimento de novas aplicações são mais baixos.

2.3. Desenvolvimento

A arquitetura sugere os componentes fundamentais na elaboração de um ambiente de *data warehouse*, mas existem dois aspectos que se destacam em sua construção: o projeto da interface com os sistemas operacionais e o projeto do *data warehouse* propriamente dito.

No contexto da criação de um ambiente secundário dentro da instituição que o abriga é importante não pensar em projeto como uma fase única e estática. O desenvolvimento de um *data warehouse* acontece de maneira heurística e cada passo feito será repetido em um outro momento com enfoques diferentes (INMON, 1997).

Uma das alternativas é começar criando *data marts*, seguindo uma estratégia assunto por assunto ou *bottom-up*, e ir evoluindo aos poucos (KIMBALL, 1998a). A outra é construir um protótipo do *data warehouse*, para um grupo de usuários, com pequenas amostras de dados, e aguardar um consenso do grupo quanto aos requisitos e funcionalidades para carregar o *data warehouse* com dados operacionais e dados externos, assumindo uma estratégia *top-down*. Esta estratégia foi bem descrita em referência a Inmon (1997), no início deste capítulo, ao se tratar da questão do povoamento e é justificada pelos custos despendidos com os recursos consumidos e pelo impacto no ambiente operacional relacionado a uma implementação global e completa do *data warehouse*. De acordo com o que foi mostrado na seção sobre *data marts*, estas duas técnicas devem coexistir durante o processo de desenvolvimento para que se obtenha um *data warehouse* robusto e preparado para acompanhar os requisitos.

Como existem diferentes abordagens para o desenvolvimento de *data warehouses*, é importante ter em mente uma metodologia no início de sua construção, devendo-se fazer esta escolha fundamentada em pelo menos três aspectos: escopo do *data warehouse* (corporativo, departamental), grau de redundância dos dados e perfil do usuário final (ORR, 1996).

Os requisitos de um *data warehouse* não estão completamente definidos antecipadamente. O seu escopo pode ser amplo, abrangendo todo o conjunto de informações da instituição, ou restrito a uma base de dados de um único departamento. Inmon (1997) defende a idéia de se criar uma pequena parte do *data warehouse* e disponibilizá-la para utilização do usuário. A partir daí deve-se obter o *feedback* do usuário e ajustar, adequadamente, os níveis de detalhe desejados. (KIMBALL, 1996a) reforça que um conjunto de entrevistas com o usuário é crucial para a preparação do projeto. Nestas reuniões, os usuários expõem suas opiniões sobre o que lhes é mais importante no negócio. Entretanto, o projetista deve ter o cuidado de não questionar diretamente aspectos de projeto que fogem dos interesses do usuário final.

O grau de redundância está diretamente ligado à distribuição física dos dados discutida anteriormente. A forma mais utilizada é o *data warehouse* centralizado, seja porque a visão integrada dos dados é utilizada apenas em um local ou porque o volume de dados não é significativo ou até mesmo pela dificuldade de acesso a informações distribuídas. Uma análise precisa do nível de utilização do ambiente pode ajudar a decidir o tipo de estrutura a ser utilizada.

O perfil do usuário final interfere na escolha do modelo de dados a ser utilizado pelo *data warehouse*. Atividades simples de geração de relatórios e consultas pré-estruturadas não sobrecarregam nem o SGBD e nem o ambiente servidor, podendo ser feitas através de uma estrutura mais complexa, como por exemplo a estrutura multidimensional, altamente desnormalizada para facilitar o acesso, mas que exige reconstrução a cada mudança. Já ambientes dinâmicos, com atividades de análise complexas, típicas de ambientes de apoio à tomada de decisão, exigem mais de todo o ambiente, devendo-se optar neste caso por uma arquitetura simples, de fácil alteração e normalizada, ideal para atender às necessidades de constantes mudanças.

O ciclo de vida

O processo de desenvolvimento do ambiente do *data warehouse* ocorre partindo de um enfoque diferente do que é seguido para os sistemas operacionais clássicos. Neste último, as atenções estão voltadas para os processos envolvidos nas atividades do dia a dia do usuário, enquanto que no ambiente de *data warehouse* os requisitos dos usuários são mais importantes e o processo se inicia nos dados. Em outras palavras, o ciclo de vida para um ambiente operacional corre de maneira inversa ao do ambiente de apoio à decisão. Para o ambiente operacional são definidas as atividades do negócio e os processos necessários para cumpri-las. Parte-se, então, para o levantamento dos dados envolvidos na execução destas atividades. Para o ambiente de apoio à tomada de decisão, são levantadas as áreas de maior interesse e as informações que são geradas por estas áreas. Então, estes dados são trabalhados e, após a base de dados ser montada, parte-se para as definições dos processos exigidos para análise dos dados. Inmon (1997) mostra claramente a diferença entre os dois tipos de ciclo (Tabela 2-2):

Tabela 2-2 - Ciclos de vida inversos

Ciclo de Vida Clássico	Ciclo de Vida do <i>data warehouse</i>
Levantamento de necessidades	Implementar o repositório de dados
Análise	Integrar os dados
Projeto	Procurar as distorções
Programação	Definir os programas para os dados
Teste	Projetar o sistema de apoio à decisão
Integração de módulos	Analisar os resultados
Implementação	Entender as necessidades

O início do desenvolvimento deve-se dar após a definição da estratégia e avaliação dos aspectos discutidos anteriormente, relacionados ao escopo e ao nível de utilização do repositório. Estas decisões feitas antes da implementação do projeto têm um impacto significativo no esforço de execução do mesmo.

O objetivo de um *data warehouse* é prover dados com qualidade. Os requisitos dependem, portanto, das necessidades individuais de informação de seus usuários. Mas, apesar das necessidades por informações específicas mudarem com frequência, os dados associados não mudam. Partindo-se do princípio de que as operações de negócio de uma instituição permanecem relativamente constantes, existe apenas um número finito de objetos e eventos com as quais uma organização está envolvida. Na fase de **análise de requisitos**, os dados do ambiente operacional são identificados em suas tabelas e arquivos e são tratados de forma a mantê-los integrados no *data warehouse*. São definidos, também, que aspectos serão tomados como fatos e quais serão suas dimensões.

A **modelagem** dos dados é feita quando se obtém o domínio dos atributos operacionais. São avaliados os elementos de tempo para as chaves e os dados derivados para as tabelas de fato, as necessidades de agregação e cálculos dos dados, a necessidade de replicação, enfim, todos os aspectos relacionados à organização da base de dados. A escolha da técnica de modelagem deve ser avaliada em torno da complexidade do modelo sob os aspectos de volume de dados estimado e tipos de operações a serem realizadas. É importante ressaltar que nesta fase devem, também, ser modelados os metadados cuja importância já foi discutida anteriormente.

Como pode-se notar, a criação de um *data warehouse* não envolve, simplesmente, a transferência dos dados do ambiente operacional para o ambiente analítico. Após o controle dos dados trabalhados, chega-se à fase de **definição dos processos** necessários para execução desta transferência. Estes processos envolvem atividades de extração dos dados operacionais, refinamento, limpeza e reformatação dos dados seguindo as especificações definidas na fase anterior e são fundamentais para a construção do *data warehouse* propriamente dito.

A fase de **construção** envolve a sincronização dos processos definidos para montagem e povoamento do *data warehouse*. Nesta fase, é necessário o gerenciamento dos metadados no que diz respeito às alterações dos mesmos nas bases operacionais, às definições de consultas e visões, a padronização das definições e monitoração das replicações e atualizações. Estas atividades devem ser

realizadas de maneira que sua execução não atrapalhe o andamento do ambiente operacional e mantenha o *data warehouse* atualizado em relação às alterações encontradas nas fontes de origem.

A **disponibilização** do repositório de dados construído é feita a partir da construção de interfaces que possibilitem o usuário acessá-lo, atendendo às suas necessidades gerenciais. Esta fase é iterativa e incremental, visto que, à medida em que o usuário vai conhecendo os dados que lhe estão disponíveis, também vai percebendo as possibilidades de acesso e vão surgindo novas necessidades em termos de visualização e manipulação dos dados.

O ciclo de vida do desenvolvimento de um *data warehouse* é evolutivo. É notório que além das demandas em relação à interface, com a utilização do *data warehouse* e percepção das possibilidades de acesso aos dados operacionais, vão surgindo novos requisitos do negócio, necessidades de atingir novas áreas e englobá-las ao conhecimento do ambiente gerencial. Após esta constatação reinicia-se o ciclo, desta vez avaliando-se também a possibilidade de se encontrar novos requisitos contemplados em repositórios já implementados na instituição, o que levaria à integração de repositórios sem que sejam necessários acessos diversos ao ambiente operacional para o mesmo propósito (Figura 2.2).

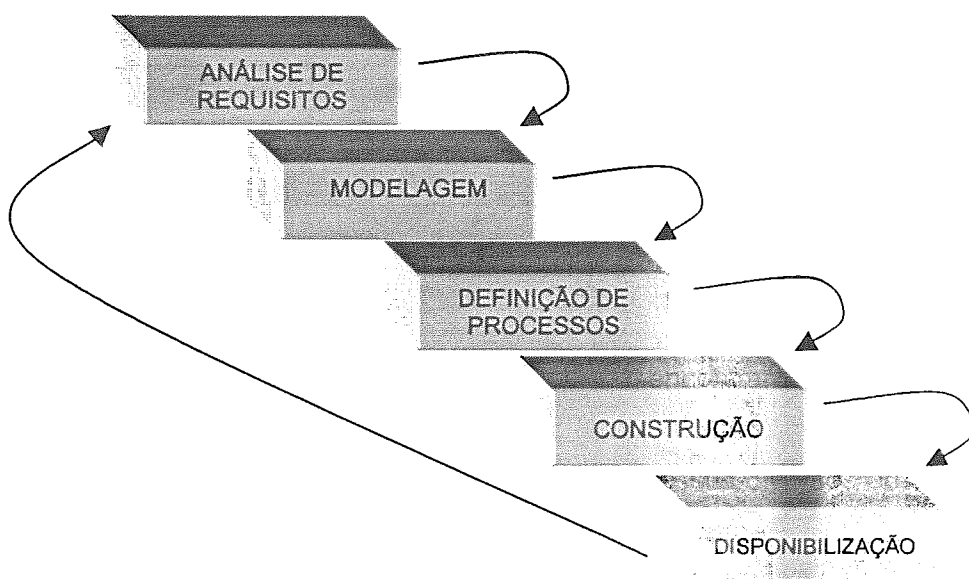


Figura 2-2 - Ciclo de vida do desenvolvimento de um data warehouse

KIMBALL (KIMBALL, 1996a, KIMBALL, 1997b) chama a atenção para a necessidade de se avaliar a duração do repositório de dados no sentido de medir o quanto se deve voltar aos dados históricos para trazê-los ao *data warehouse*, evitando assim que o repositório seja sobrecarregado com dados que não serão utilizados.

Outro ponto importante é a avaliação do uso dos dados, o acompanhamento das lentas mudanças de dimensões que não devem ser ignoradas em nenhum dos acessos.

Porter&Rome (1994) defendem a importância da segurança e privacidade dos dados. A ausência destes fatores pode complicar ou, até mesmo, limitar o desenvolvimento do *data warehouse* ainda que este não permita a atualização de dados. A possibilidade de acessar dados de outras áreas e convertê-los em informações úteis pode ser ameaçadora em muitas instituições. É essencial, também, a validação dos requisitos do usuário com o modelo que está sendo proposto, através da capacidade de responder às questões analíticas levantadas com o usuário.

Para Mattison (1997), um *data warehouse* é uma coleção de proposições de valor e define proposição de valor como sendo qualquer coleção contínua de componentes de aquisição, armazenamento e acesso que, quando utilizadas juntas, disponibilizam determinado valor a um grupo de usuários finais. O processo de desenvolvimento por proposição de valor foi idealizado em três etapas:

Desenvolvimento da visão do problema: nesta etapa são definidos os grupos de negócio e com cada grupo são discutidos os principais problemas de negócio que eles têm necessidade de resolver no dia-a-dia, afim de desenvolver um consenso entre os participantes sobre as informações coletadas. A partir deste ponto são apresentadas as ferramentas de acesso e o benefício de cada uma delas para a solução dos problemas. Feito isto, são definidas com os usuários suas próprias proposições de valor.

Organização e priorização: nesta etapa são avaliadas as proposições de valor, desenvolvidas na etapa anterior, no sentido de determinar a prioridade de cada uma delas no contexto da instituição e organizá-las em uma seqüência lógica de disponibilização.

Desenvolvimento e construção: finalmente, de posse da lista de proposições de valor definidas, é possível construir a infra-estrutura física e operacional baseada nas condições de custos e riscos avaliados na etapa anterior.

Esta última abordagem foi tratada com mais detalhes por constituir a base principal deste trabalho.

A seguir tem-se uma tabela comparativa apresentando três abordagens diferentes com relação ao desenvolvimento do *data warehouse* e à projeção para o processo descrito (Tabela 2-3):

Tabela 2-3 - Avaliação das abordagens de desenvolvimento do *data warehouse*

Projetos	Etapas do desenvolvimento	Mapeamento genérico	Características Diferenciais
Definido por: Ralph Kimball	Escolher a área do negócio do <i>data mart</i>	Análise de requisitos	<ul style="list-style-type: none"> ▪ Políticas de duração do repositório ▪ Avaliação do uso dos dados ▪ Técnicas de modelagem
	Escolher a tabela fato	Análise de requisitos	
	Identificar e ajustar as dimensões	Análise de requisitos	
	Escolher os fatos	Modelagem	
	Armazenar os dados pré-calculados nas tabelas fato	Modelagem	
	Definir os atributos das tabelas dimensão	Modelagem	
	Definir a duração da base de dados		
	Verificar a necessidade de rastrear as lentas mudanças das dimensões		
	Decidir os modos e prioridades das consultas	Definição de processos	
ASU Data Warehouse Definido por: J. D. Porter e J. J. Rome	Identificar o tópico ou assunto principal	Análise de requisitos	<ul style="list-style-type: none"> ▪ Questões de segurança ▪ Validação do modelo de dados
	Adicionar um elemento de tempo às tabelas	Análise de requisitos	
	Nomear campos nas tabelas ou visões adequadamente	Modelagem	
	Adicionar campos derivados, quando necessário	Modelagem	
	Duplicar os dados, se necessário	Modelagem	
	Excluir campos extras dos arquivos operacionais	Definição de processos	
	Criar tabelas ou visões lógicas	Definição de processos	
	Levar em consideração a segurança		
	Avaliar se o modelo de dados responde às questões do negócio		
Definido por: H. Gill e P. Rao	Determinar a estratégia de implementação		Decisões de planejamento
	Determinar a metodologia de desenvolvimento		
	Selecionar a arquitetura do <i>data warehouse</i> apropriada		
	Selecionar o escopo inicial de implementação	Análise de requisitos	
	Desenvolver cenários de uso do negócio	Disponibilização	
	Coletar os metadados	Modelagem	
	Selecionar e especificar os diversos requisitos	Análise de requisitos	
	Modelar os requisitos	Modelagem	
	Implementar o modelo fisicamente	Construção	
Definido por Mattison	Organizar grupos de negócio	Análise de requisitos	Definição de grupos de negócio Acompanhamento da cadeia de valor es
	Apresentar tecnologia	Análise de requisitos	
	Definir proposições de valor	Análise de requisitos	
	Organizar e priorizar as proposições de valor	Análise de requisitos e Modelagem	
	Construir a estrutura física do ambiente	Construção	
	Desenvolver a base operacional do ambiente	Definição de processos	

2.4. Requisitos

Data warehouse tornou-se um importante mecanismo de apoio ao processamento analítico e estratégico. Este fato tem sido evidenciado em diferentes organizações influenciadas por ramos de negócios diversos. A partir da análise de algumas abordagens, foi observada uma tendência à admissão de um ambiente desta categoria, seja por impasses no aproveitamento dos dados disponíveis, como por exemplo, grande volume de dados que precisam de tratamentos específicos, seja pela falta de tecnologia adequada para as atividades gerenciais sobre o ambiente operacional.

A especialização ou a fragmentação do processo de transformação da informação constituem-se artifícios que os negócios utilizam para acompanhar a eficiência em termos de gerenciamento do conhecimento. O que se tem visto é que na maioria das grandes organizações são criados vários grupos de especialistas, cada grupo disponibilizando o seu conhecimento em áreas específicas e dispersas. Esta fragmentação do processo pode criar “valas” de conhecimento que podem manter, proteger e as vezes esconder informações úteis para outras áreas de negócio impossibilitadas de serem acessadas por inúmeras razões (MATTISON, 1997):

- A informação pode estar inacessível através de meios mecânicos ou tecnológicos;
- A informação foi armazenada de maneira que apenas as pessoas que organizaram aquele conhecimento são capazes de entender;
- Alguns grupos podem preferir ocultar detalhes do seu negócio por não querer que outras áreas tomem conhecimento destes.

Algumas empresas têm como produto primário a aplicação do conhecimento para a solução dos problemas das pessoas, como é o caso das instituições de saúde. Estas instituições não geram nenhum produto, mas o domínio do seu negócio envolve a transformação do conhecimento. Neste sentido, Mattisson (1997) seguiu as seguintes etapas na construção de um *data warehouse* na área de telecomunicações, às quais podemos adaptar a outra área em estudo:

- Identificar os tipos de conhecimento que estão dirigindo o processo de transformação;

- Encontrar na organização as correlações diretas entre a estrutura básica organizacional do negócio e a parte essencial operacional dos sistemas que dirigem esta estrutura.

Esta análise permite observar o processo de transformação da informação necessário para a tomada de decisão através de uma cadeia de valor¹. Este conjunto é formado pelas atividades principais, que são aquelas que caracterizam o negócio mantido pela organização, e pelas atividades de apoio, que são aquelas que tornam possíveis as atividades principais.

Os exemplos clássicos deste tipo de ambiente referem-se a empresas cujas atividades estão associadas a comércio do tipo venda de produtos para avaliação de lucros, gerência de investimentos ou análise de mercado. A diversidade de interesses levou o ambiente de dados integrados para organizações cujas áreas de negócios diferem entre si. Cada organização apresenta requisitos próprios específicos à sua área de atuação. Estes requisitos caracterizam alguns cenários que impulsionam o desenvolvimento de um ambiente de *data warehouse* com objetivos também moldados às suas necessidades.

Das abordagens avaliadas é interessante apresentar aquelas cujos objetivos são distintos mas complementam as metas gerais do ambiente.

A primeira abordagem a ser avaliada está ligada ao ramo das telecomunicações e destina-se a prover subsídios para avaliar as formas que a empresa, os sistemas de informação e a infra-estrutura podem ser melhor utilizadas para melhorar a sua lucratividade e competitividade (MATTISON, 1997). A outra é um ambiente de consolidação de dados farmacêuticos aplicados e integração aos dados de registros eletrônicos do paciente para apoiar o tratamento orientado ao paciente, otimizar terapias medicamentosas, auxiliar a continuidade do tratamento e obter retornos positivos do paciente (KELSO&MYERS, 1998). Por fim, um projeto de *data warehouse* desenvolvido na Universidade de Stanford com objetivos genéricos de coleccionar e integrar de maneira eficiente informação de fontes de dados heterogêneas, inclusive sistemas legados (HAMMER *et al.*, 1995).

KELSO&MYERS (1998) ainda chamam a atenção para as diferentes análises que podem ser feitas. São elas: a análise passiva de resultados, baseada em perfis,

¹ Cadeia de valor é um conjunto de atividades que geram subsídios para que a instituição atinja seus objetivos (PORTER, 1997).

história e tendências utilizadas para definir padrões, planos de testes e cursos de ações; e a análise ativa de resultados, baseada nas ocorrências atuais utilizada para intervenção apropriada do tratamento e monitoração da situação. Ambas as definições podem também ser aplicadas a diferentes áreas de negócio.

A avaliação das abordagens é apresentada na tabela a seguir (Tabela 2-4):

Tabela 2-4 - Avaliação dos requisitos para construção de um ambiente de *data warehouse*

Ramo de Atividade	Cenário	Objetivos específicos da organização	Metas do ambiente
Telecomunicação Mattison	Numerosas transações gerando grande volume de dados	Monitorar o desempenho da rede Executar as tarefas de planejamento e otimização da rede	Manipulação de grandes volumes de dados Análise passiva de resultados
	Necessidade de analisar os dados brutos ou abstratos (sessões e conexões)	Medir eficácia da rede	Análise ativa de resultados
	Necessidade de fazer parte do ambiente competitivo analisando as tendências do mercado	Prover ferramentas para auxiliar a: Desenvolver rapidamente novas infra-estruturas baseadas em marketing Efetuar mudanças culturais necessárias para torná-las mais centradas no usuário e menos centrada na tecnologia	Análises estatísticas
	Acompanhamento das mudanças tecnológicas	Desenvolver os recursos para ajustar rapidamente as mudanças Analisar o quão bem estas mudanças serão implementadas	Análises estatísticas
Farmacêutico KELSO	Tradição na análise de dados complexos e uso de inovações no gerenciamento de dados	Manter a tradição	
	Dados distribuídos em forma e lugares diversos	Avaliar atividades dos especialistas Possibilitar a prescrição de tratamentos padronizados Promover a eficácia no estabelecimento de tratamentos com medicações alternativas recomendadas.	Análise passiva de resultados
	Necessidade de acessar o valor do tratamento	Capacidade de comparar custos e resultados de tratamentos alternativos	Reduzir custos
	Necessidade de analisar resultados da aplicação de medicamentos em pacientes distintos para definir tratamentos mais eficientes	Melhorar a qualidade do tratamento	Análise passiva de resultados
	Necessidade de prever resultados através da análise conjunta das queixas do paciente e medicação aplicada	Integrar as diversas formas de tratamentos de saúde	Facilitar a comunicação
	Necessidade de avaliar o progresso do tratamento	Melhorar a continuidade do tratamento	Análise ativa de resultados
Acadêmico WHIPS	Necessidade de coordenar o tratamento e avaliar se os resultados terapêuticos são satisfatórios	Auxiliar a garantia de um curso apropriado do tratamento Atrair a satisfação do paciente	Análise ativa de resultados Facilitar a comunicação
	Grande volume de informação - artigos, relatórios, tutoriais - gerados a todo momento	Colecionar dados científicos	Manipulação de grandes volumes de dados Integração das diversas fontes de dados
	Grande intensidade de dados sendo gerados no dia a dia exigindo o espaço dos dados mais antigos	Acesso a dados históricos	Integração das diversas fontes de dados
	Ineficiência no processamento de consultas das bases de dados habituais	Melhor desempenho nas consultas previamente definidas e a dados integrados	Eficiência em consultas analíticas

2.5. Tecnologia

A principal questão relacionada à coexistência do ambiente dos sistemas operacionais e do ambiente dos sistemas analíticos refere-se à organização física de suas bases de dados. De um lado aplicações baseadas no processamento de transações que requerem alto desempenho; do outro, os dados são organizados de maneira apropriada para atender consultas e relatórios específicos (Codd et al, 1994). Por estes aspectos, já se pode perceber o quanto são incompatíveis as estruturas exigidas por cada ambiente e, conseqüentemente, a necessidade de tecnologias que atendam às duas demandas.

A maioria das ferramentas de acesso a *data warehouses* explora a natureza multidimensional dos dados. Para isto, tornou-se comum estruturar os dados em bancos de dados relacionais tradicionais em esquemas do tipo estrela ou flocos de neve, descritos anteriormente. Estes esquemas utilizam múltiplas tabelas e relacionamentos para simular esta característica.

Em termos de bancos de dados multidimensionais, existem alguns mecanismos que valorizam a sua existência acrescentando atrativos para os usuários finais. Em muitas situações deseja-se, apenas, selecionar um subconjunto de interesse e analisá-lo de forma mais detalhada através de seus atributos. Retornando à analogia entre o modelo multidimensional de dados e um cubo, o mecanismo de *pivoting*, sugere que este cubo seja rotacionado para mostrar uma face particular. Outra operação importante é chamada de *slicing and dicing* que permite a seleção de uma face ou subconjunto do cubo. São importantes, também, as operações de *roll-up* e *drill-down* que se referem à navegação hierárquica entre as dimensões. Muitas vezes, a análise se inicia nos dados agregados e depois parte-se para o detalhamento maior, ou vice-versa (AGRAWAL et al, 1997).

A tecnologia destinada ao processamento analítico (OLAP) surgiu para evitar que fossem usados artifícios na tecnologia existente para o processamento de transações (OLTP). OLTP (*On Line Transaction Processing*) é um conjunto de classes de programas que facilita e gerencia aplicações orientadas ao processo de transações, tipicamente para entrada e recuperação de dados. OLAP (*On Line Analytical Processing*) é a tecnologia que habilita os usuários a acessar dados de maneira multidimensional, rápida, interativa e fácil de usar, desempenhando computações métricas avançadas do tipo comparações, variações de porcentagem,

classificações e taxas multidimensionais nos dados no *data warehouse* (DAPHNE, 1998).

A principal diferença entre as aplicações OLTP e o *data warehouse* é a capacidade deste para descrever precisamente o passado. Isto ocorre devido ao fato destes sistemas estarem sempre atualizando os dados com novas entradas, sem sobrepor dados antigos (KIMBALL, 1996b). A habilidade de apresentar relatórios multidimensionais e prover análises dos dados também diferencia OLAP de qualquer outra ferramenta de consulta e relatório disponível (DAPHNE, 1998).

Existem três importantes categorias de OLAP, que diferem entre si na forma como os dados são armazenados fisicamente mas que mantém a mesma interface e habilidade analítica:

- **OLAP Multidimensional (MOLAP)**, determina que os dados e agregações devem ser armazenados em uma estrutura multidimensional, ganhando em desempenho no processamento dos cubos.
- **OLAP Relacional (ROLAP)**, determina que os dados devem ser armazenados em uma estrutura relacional ou apoiado por consultas em SQL, permitindo usufruir das vantagens da tecnologia relacional e do gerenciamento de dados mais comum, mas o processamento dos cubos deve ser feito na utilização.
- **OLAP Híbrido (HOLAP)**, determina que os dados devem ser armazenados em bancos de dados relacionais e as agregações em estruturas multidimensionais. Esta categoria oferece as vantagens do MOLAP para as agregações sem duplicar o armazenamento dos dados.

Os produtos que seguem a linha do ROLAP oferecem mecanismos de servidores analíticos, ferramentas de análise e relatórios, a partir de uma visão relacional através da lógica do sistema.

Características incorporadas aos SGBDs

Para atender a estes e outros requisitos, fornecedores de SGBD relacionais têm adicionado funcionalidades a seus produtos ou incluído extensões às estruturas de armazenamento e aos operadores relacionais, como no caso do RedBrick, e esquemas de indexação especializados, como oferecido pelo Sybase IQ. Estas funcionalidades podem melhorar o desempenho para recuperações por conteúdo através de pré-junções de tabelas usando índices ou pelo uso de listas de índices totalmente invertidas. Alguns utilizam um esquema de mapeamento de bits para pré-processar as relações e até mesmo, algumas vezes, outras relações armazenadas de

outra forma que não por linhas. Este caso tem a vantagem de ser uma extensão do mecanismo de indexação para um banco relacional, complementando o espaço de dados. Esta abordagem, entretanto, requer tempo e espaços adicionais para os índices (WELDON, 1997) (AGRAWAL et al, 1997). Outros produtos disponíveis, que seguem estes princípios ou equivalentes são: DSS Server da MicroStrategy, o Axsys da Information Advantage Inc., MetaCube da Informix Software Inc. (FIRESTONE, 1997).

Os fabricantes que optaram por bancos multidimensionais, o fizeram mantendo os dados organizados como um cubo n-dimensional baseado numa estrutura de armazenamento especializada não relacional, implementada utilizando um *array* esparso para armazenar as agregações pré-computadas, como por exemplo Essbase da Arbor Software e o Gentium da Planning Sciences. A opção de manipular diretamente objetos multidimensionais, é recomendada para ambientes mais estáveis no que diz respeito a mudanças de requisitos. Nos bancos de dados multidimensionais, as dimensões são identificadas ao criar a estrutura do banco. Por esta razão, muitos deles exigem uma recarga completa quando ocorre uma reestruturação, de forma que adicionar uma nova dimensão pode ser trabalhoso. Entretanto, têm a vantagem de ser uma representação compacta e eficiente, se o volume de dados for pequeno (WELDON, 1997) (AGRAWAL et al, 1997).

Por outro lado, os dados de um *data warehouse* podem estar distribuídos por múltiplos bancos de dados, inclusive sob diferentes sistemas de gerenciamento, como já foi visto nos aspectos de distribuição de dados. Pelo fato dos bancos de dados multidimensionais fornecerem uma visão específica dos dados da empresa, cada área pode assumir uma organização dos dados para a sua visão do negócio, atendendo às suas necessidades. No entanto, é pouco provável que a mesma estrutura atenda, igualmente bem, às questões de tomada de decisão das diversas áreas da empresa. Neste caso, um sistema de banco de dados relacional é usualmente mais adequado para gerenciar um banco de dados integrado, provendo uma estrutura mais descompromissada em relação aos princípios analíticos de acordo com as necessidades de cada área.

Uma solução freqüentemente encontrada é a separação do gerenciamento dos dados entre o *data warehouse* relacional integrado da empresa e os seus *data marts* multidimensionais. Esta alternativa introduz a necessidade de uma estratégia de distribuição de dados que coordene a alimentação de novos dados aos bancos multidimensionais. Uma solução semelhante é adotada no caso em que o *data warehouse* possui diferentes níveis de detalhe: a camada atômica, de maior nível de

detalhe, é mantida em formato relacional, enquanto a camada contendo dados resumidos, pode ser mantida em formato multidimensional. Gerenciadores como Acumate da Kenan Technology Inc., Express da Oracle Corp. e Lightship da Pilot Software Inc., começaram como ferramentas de apoio à tomada de decisão, com formatos de armazenamento proprietários otimizados para processamento do tipo OLAP. Estes produtos tiveram sua aplicabilidade estendida pela adição de acesso relacional, sem que abandonassem o formato multidimensional (WELDON, 1997).

2.6. Considerações finais

De acordo com o que foi visto no início deste capítulo, a análise multidimensional dos dados não é algo novo. Por diversas razões, os dados já eram extraídos da forma que havia disponível no ambiente operacional existente e transformados em planilhas, relatórios, gráficos, enfim, em algum formato do qual se o pudesse extrair informações para análises.

A renovação foi encontrada na parte de engenharia de sistemas e tecnologias que pudessem apoiar através de métodos e ferramentas todas as necessidades de extração, transformação e agregação dos dados e, também, na apresentação das informações. Uma visão detalhada desta evolução pode ser encontrada em (PENDSE, 2000).

3. Ambientes Médicos de Análise

Neste capítulo são apresentados alguns sistemas de informação desenvolvidos para a área médica e exemplos de sistemas em ambientes integrados, assim como alguns aspectos importantes neste contexto.

3.1. Aspectos Gerais

A revolução da tecnologia da informação, provocada pela crescente demanda nos diversos setores industriais como bancos, empresas aéreas e setores de manufaturas e de serviços, atingiu a área médica após mais de 10 anos de estabelecida nas outras áreas. Os sistemas que utilizam informação médica aproveitam as oportunidades dispostas pela tecnologia para promover serviços de saúde reduzindo custos e melhorando a qualidade.

Esta revolução na área de saúde envolve a implementação de registros médicos eletrônicos, atualização dos sistemas de informação hospitalar, utilização de *intranets* para compartilhar informações entre os participantes relacionados, utilização de redes públicas, tipo Internet, para distribuir informações relacionadas à saúde e promover diagnósticos remotos via telemedicina (RAGHUPATHI, 1998). Sistemas de informação projetados especificamente para registrar ou transmitir informações clínicas detalhadas de um paciente específico para uso direto do especialista são raros e muitas vezes são limitados a reportar resultados de exames laboratoriais e radiológicos (KAHN, 1998).

Muitos sistemas têm sido desenvolvidos com diferentes objetivos para atingir diferentes grupos de interesse. A principal característica dos sistemas na área médica é o tipo de dado manipulado. STEAD & HAMMOND in (RAMIREZ *et al.*, 1994) apresentaram o seguinte resumo dos tipos de dados que podem ser encontrados num sistema de registro médico:

- **Dados demográficos:** são dados de identificação e dados sócio-econômicos do paciente;
- **Dados genéricos e independentes do tempo:** são os protocolos de pesquisa, como informações de um determinado tumor ou um registro das necessidades especiais do paciente;
- **Problemas ou diagnósticos:** incluem fatores associados que podem indicar a origem de um problema, se o mesmo está ativo ou inativo, e as datas dos tratamentos associados;
- **Dados temporais:** dados que não são gerados por uma solicitação ou prescrição, mas são analisados segundo sua variação no tempo, como por exemplo as queixas do paciente, resultados do exame físico e avaliação do progresso da doença;
- **Dados gerados a partir de uma prescrição,** resultados de testes laboratoriais e radiológicos;
- **Terapias,** incluindo dietas, vacinas, prescrições etc.;
- **Dados de admissão/ consulta:** incluem datas, lugares e dados gerados durante uma consulta.

Ainda pode-se citar os dados de informação genérica sobre saúde utilizados em sistemas de educação e tutoriais.

Em se falando de sistemas hospitalares são, também, importantes os dados internos financeiros mantidos pela instituição. Por muitos anos, os sistemas de informação hospitalar mantinham apenas dados administrativos (ANDERSON, 1998). Recentemente estes dados têm sido cada vez mais utilizados no sentido de otimizar o atendimento e melhorar a qualidade dos resultados.

Um dos principais problemas encontrados no desenvolvimento de sistemas que envolvem registros eletrônicos de pacientes é a resistência dos especialistas em modificar a sua forma de tratar os dados do paciente. Este problema se dá principalmente pela dificuldade de estruturar a informação médica. Grande parte do atendimento é subjetiva e muitas vezes própria do especialista, o que leva à preferência por dados dispersos e não definidos, como campos em texto livre, e à resistência de muitos em compartilhar o método individual de tratamento. Os dados estruturados exigem dos especialistas um vocabulário comum predefinido, tornando mais fácil o uso da informação para pesquisas clínicas e garantia da qualidade, para troca de informações entre outras instituições de saúde, para sistemas integrados de apoio à decisão, para acesso *online* à literatura biomédica e muitos outros benefícios

(van MULLIGEN, 1998). Por outro lado, a informação em texto livre pode dar maior liberdade de expressão ao especialista e o aproxima mais do registro em papel, embora muitas vezes dificulte a comunicação entre os colegas devido a problemas como a forma de organização do texto.

Mesmo com o desenvolvimento dos sistemas de informação de laboratórios e radiologia, os métodos usados para registrar e compartilhar informação não foram alterados desde a concepção do registro médico formal mantida há anos. Embora muitos estudos tenham mostrado sérias inadequações nos registros médicos manuscritos, a conveniência em utilizar o papel dificulta a aceitação do registro eletrônico (ANDERSON, 1998, KAHN, 1998).

3.2. Sistemas na área médica

Muitos sistemas têm sido desenvolvidos nesta área, ainda que se considere os problemas apresentados na sessão anterior em relação à organização da informação médica. Aliam-se, para isto, interesses dos especialistas, em melhorar a qualidade e o custo do trabalho, e o interesse dos desenvolvedores de sistemas em expandir o conhecimento e investir em uma área com grande potencial.

Os sistemas podem variar em escopo, sendo genéricos, para toda a área médica, específicos para uma especialidade - neurologia, cardiologia, oncologia etc. – ou para uma categoria, abrangendo sistemas administrativos simples, sistemas educacionais, prontuário médico eletrônico, teleconsultas entre outros. Uma característica desses sistemas é tornar cada vez mais próxima a informática do domínio médico, ou seja, desenvolver sistemas voltados para as necessidades reais de médicos, pacientes e, até mesmo, estudantes em formação.

3.2.1. Sistemas de Informação

Os sistemas de informação compreendem toda a tecnologia necessária para aquisição, armazenamento e recuperação de qualquer forma e tipo de informação. Os avanços tecnológicos em hardware possibilitaram a diversificação na forma de apresentação da informação e levaram à evolução do software no sentido de acompanhar estas inovações e possibilitar melhor utilização da informação.

Os sistemas de informação médica envolvem o armazenamento e recuperação de informações relacionadas ao paciente. Estas informações são desde dados administrativos e orçamentais do paciente até os dados gerados no seu atendimento

como sua história clínica, dados do exame físico, prescrição médica, diagnóstico, tratamento adotado etc. Podemos listar alguns dos objetivos destes dados relacionados à melhoria do trabalho dos especialistas, à maior utilização dos dados disponíveis e à satisfação geral do paciente:

- Organização e codificação de conhecimentos e dados médicos;
- Aquisição e representação de conhecimentos de avaliação adquiridos através da experiência prática dos especialistas;
- Comunicação entre profissionais de saúde através de sistemas informatizados e facilidade de acesso a especialistas mais experientes;
- Padronização da informação médica para facilitar a comunicação e o acesso aos diversos níveis de profissionais da área de saúde.

Os estudos relacionados a registros de pacientes têm direcionado os interesses nas questões de aquisição e representação da informação como também na avaliação da qualidade da informação e aceitação do processo (RECTOR, 1996). Os sistemas que manipulam informações médicas devem estar preparados para trabalhar com grandes bases de dados com diferentes níveis de estrutura, bem como prover um alto nível de acesso não mais baseado em critérios precisos e tradicionais de consulta. Assim devem ser abrangentes e flexíveis, capazes de extrair informações em diferentes bases de dados e em diferentes formatos (LIMA *et al.*, 1997).

A manipulação de dados médicos exige alguns critérios que podem ser avaliados para qualquer tipo de dado que esteja sendo tratado. Estes critérios podem ser enumerados como (RECTOR, 1996): abrangência, precisão, disponibilidade, rapidez, aceitação pelo usuário (profissional médico) e aceitação pelo paciente.

A abrangência deve ser avaliada tanto em relação ao paciente, no sentido de que registros médicos devem conter todos os dados médicos e eventos terapêuticos ocorridos em toda sua vida para seus cuidados adequados, bem como suprir as necessidades do clínico na observação e avaliação dos registros. Deve possibilitar a manipulação das diferentes formas, mas também oferecer uma ligação coerente entre eles. A visualização dos dados deve ser feita de maneira conveniente e precisa, de modo que os usuários devem ser capazes de acessar essas informações com a mesma naturalidade e confiabilidade com que usam o papel. Isto pode estar relacionado ao fato de que o uso do computador, no acesso aos dados, não deve perturbar o paciente durante o atendimento (RATIB, 1997).

O rápido desenvolvimento dos sistemas de informação computadorizados e da informática médica tem disponibilizado uma variedade de documentos em diferentes formas. Grande quantidade desses dados tem sido gerada em forma de imagens. Isso se deve ao fato dos especialistas basearem-se, cada vez mais, nas diversas modalidades de imagens para formular seus diagnósticos. Por esta razão, os sistemas de informação que utilizam este tipo de dado devem estar aptos a visualizar e comparar as imagens em suas modalidades, correlacionando o que foi observado com o restante das informações disponíveis em outra forma (RATIB, 1997). As novas tecnologias de computação são responsáveis pelos avanços em imagens médicas relacionados à sua aquisição e apresentação. Ao lado disso está a importância da visualização do problema no momento da definição do tratamento, principalmente em casos de cirurgia ou casos em que há uma evolução da patologia e esta evolução precisa ser avaliada nos diversos estágios.

Os sistemas de informação com imagens, normalmente, estão integrados a outros sistemas de informação. Isto ocorre com mais frequência no caso de sistemas de informação com imagens médicas pela necessidade de integrar, por exemplo, o sistema de informação hospitalar ao sistema de informação radiológica. Esta integração favorece à aplicabilidade e à utilidade desses sistemas em geral (CHANG *et al.*, 1992).

3.2.1.1. Recuperação da informação

Nos sistemas de informações, em geral, mais importante do que o armazenamento da informação é a forma com que esta será acessada. A recuperação da informação é uma das áreas onde mais se concentram as atenções nos sistemas de gerenciamento de banco de dados. A ela pode-se associar a eficiência no tempo de resposta e a consistência do armazenamento.

O crescimento explosivo da quantidade de informação disponível, nas suas diversas formas, tem propiciado a busca de ferramentas eficientes de recuperação da informação. A eficiência na busca e recuperação da informação está intimamente ligada à sua representação interna. Muitas vezes grandes quantidades de dados são ignoradas, ou até mesmo dados deixam de existir sem nunca terem sido acessados, por não haver uma maneira efetiva de recuperá-los.

Diversos trabalhos de pesquisas sobre busca e recuperação da informação têm sido utilizados para atender aos mais variados propósitos. Recuperação de textos, através da indexação de palavras chaves e métodos de busca por similaridade da informação disponível e relevância (SALTON, 1995) (HUIBERS *et al.*, 1996).

Recuperação de imagens por indexação de objetos (DE MARSICOI et al, 1996), comparação de cores (FLUM, 1995), modelagem semântica (CHU et al, 1994). Estes métodos utilizam técnicas variadas como a utilização de artifícios do tipo redes neurais, lógica fuzzy, cálculos estatísticos.

3.2.2. Sistemas de Apoio à Decisão em Medicina

Os sistemas de apoio à decisão em medicina são programas destinados a melhorar as atividades e os resultados no processo de tomada de decisão dos especialistas. A principal motivação para a construção deste tipo de sistema nesta área é a habilidade de processar em tempo e precisão satisfatórios diversas informações, em virtude da constante preocupação com a satisfação do paciente e necessidade de um processo ótimo de tomada de decisão. Um fator importante neste contexto é o aumento progressivo do volume de dados, diversidade de informações geradas e inovação do conhecimento médico que devem ser utilizados neste processo.

O dia a dia do especialista da área médica gira em torno de tomadas de decisão, seja pela escolha dos dados necessários para elucidar uma questão ou pela decisão em si. Desta forma, o questionamento feito ao paciente, os exames requisitados, a determinação da patologia sugerida através da análise dos dados (diagnóstico), a definição do tratamento terapêutico que deve ser adotado, o acompanhamento do tratamento seguido, a determinação da necessidade de intervenção cirúrgica, são alguns dos tipos de decisão freqüentes na atividade dos profissionais da área médica a que podem ser apoiadas por estes sistemas.

Os sistemas inteligentes de apoio à decisão utilizam técnicas de inteligência artificial para apoiar o especialista no processo de tomada de decisão através da sugestão de diagnósticos e terapias. Estes sistemas são importantes quando as informações são incertas ou incompletas e podem ser do tipo passivo, que o especialista procura somente quando necessita de auxílio em seu processo de tomada de decisão; ou pode ser ativo se o sistema é preparado para emitir alertas no acontecimento de certos eventos.

Os sistemas inteligentes de apoio à decisão, também conhecidos como sistemas especialistas, possuem uma base de conhecimento tida como o componente de apoio às decisões que contém o conhecimento do sistema e uma máquina de inferência preparada para utilizar esta base de conhecimento de forma correta e no momento apropriado. Esta base é alimentada com o conhecimento de especialistas da área em questão, tornando o sistema útil no apoio à decisão de não especialistas.

A questão da origem do conhecimento no processo de tomada de decisão é muito importante para diferenciar as categorias deste tipo de sistema. Os ambientes de apoio à tomada de decisão discutidos no capítulo anterior disponibilizam recursos para auxiliar o processo de tomada de decisão através das informações geradas nas diversas fontes da instituição em conjunto com o conhecimento do usuário para guiar este processo. Desta forma, não existe nesta categoria um conhecimento embutido no sistema. Este é induzido pelo especialista através da formulação de consultas que podem ser vistas como perguntas ao sistema em relação aos dados do assunto em questão.

Em qualquer tipo de sistema de apoio à decisão alguns aspectos possuem fundamental importância para garantir a qualidade na realização das tarefas às quais estes sistemas se propõem. Alguns deles são listados a seguir, chamando-se atenção para as características dos ambientes de dados para análise:

- Utilização de dados precisos. Todo o processo decisório baseia-se na diversidade de dados disponíveis e depende diretamente da qualidade deles. Desta forma, é importante que sejam coletados de maneira adequada e que sejam suficientes possibilitando o mínimo de informações necessário para apoiar o processo. É importante ressaltar que a existência de dados em excesso pode prejudicar o desempenho do processo de análise. Como foi visto no capítulo anterior, é necessário se tratar os dados extraídos das fontes dispersas e definir o nível de detalhes que será tratado no *data warehouse*.
- Usabilidade da interface. Uma preocupação presente na área da informática médica é quebrar o bloqueio dos médicos à utilização de computadores para apoiar o seu trabalho. Daí a importância da interface do sistema que se comunica com o usuário, neste caso, o profissional da área médica. Sendo assim, a sua utilização não pode ser um desafio impactando no curso habitual do trabalho, distraindo a atenção do paciente que está sendo atendido ou retardando o trabalho de análise.
- Conhecimento confiável. Tanto para o desenvolvimento de uma base de conhecimento embutida, no caso dos sistemas inteligentes, como para a elucidação de problemas através da análise de dados, nos ambientes de dados integrados, é fundamental que o responsável por estas tarefas seja especialista na área que está sendo tratada; ou seja, não adianta oferecer os subsídios para construção ou utilização de dados da área de cardiologia se o indivíduo não possuir o conhecimento pertinente.

- **Capacidade de utilização apropriada do conhecimento.** No contexto dos sistemas inteligentes, isto é o equivalente à organização de um bom mecanismo de inferência, responsável pela utilização adequada da base de conhecimento. Já nos ambientes de dados para análise, esta capacidade está diretamente relacionada à habilidade do especialista em selecionar adequadamente os dados em conjunto com a utilização do seu conhecimento específico e a experiência adquirida na área.

Como se pode perceber, os níveis de tomada de decisão de cada um são diferentes, o que pode justificar a utilização do termo não ser adequada muitas vezes aos sistemas de ambientes de dados para análise, mesmo porque existem outros objetivos além de apoio às decisões gerenciais.

3.3. Características de ambientes médicos de análise

Com o advento das reformas nos tratamentos clínicos e a rápida introdução do tratamento gerenciado no ambiente de saúde, a necessidade do acesso mais eficiente das informações médicas tem crescido dramaticamente para administradores de hospitais, especialistas, enfermeiros e pesquisadores. Para que as instituições de saúde sejam mantidas neste ambiente altamente competitivo, elas devem ser capazes de processar, gerenciar e analisar rapidamente grandes quantidades de informação. Da mesma forma, pesquisadores da área se esforçam em aumentar a qualidade dos tratamentos de saúde, baseando-se em resultados clínicos e estudos mais extensos que envolvem regimes de tratamento, procedimentos e práticas médicas. (SCULLY et al, 1997).

Os ambientes que atendem a estas necessidades têm como uma das principais características a integração dos dados disponíveis na instituição gerados para os diferentes propósitos das atividades operacionais e analíticas. Por esta razão, foram descritos em detalhes apenas os sistemas de informação e os sistemas especialistas, pois estes apresentam potencial para que tenham seus dados integrados em uma base comum, uma vez que possuem entrada de dados completa e armazenamento em diferentes ambientes da instituição.

No contexto de aplicações e dados da área médica, pode-se encontrar duas denominações para ambientes de análise, cada uma com objetivos bem definidos e organização própria. Uma delas é o conceito de repositório de dados clínico que representa a integração das bases de dados operacionais dispersas e a outra é o *data*

warehouse clínico que, como já vimos, é a organização de uma base de dados para fins analíticos e, neste contexto, avaliação de experiências e ensaios médicos.

3.3.1. Repositório de dados clínicos

A informatização de uma instituição normalmente ocorre de forma não integrada. Cada área descobre uma necessidade em termos de automatização da informação e implementa esta necessidade, sem se preocupar com a integração com as outras bases de dados ou sem apresentar interesse nos outros dados.

Numa instituição de saúde é comum haver sistemas para exames de laboratório, sistemas para controle farmacêutico e sistemas para exames radiológicos. Os sistemas administrativos, muitas vezes, existem para atender às contas médicas e controlar custos do paciente, e muitas vezes possui dados demográficos e algumas informações a respeito da visita do paciente à instituição. Com este exemplo, percebe-se que é possível aprofundar-se no acompanhamento de um paciente, através dos dados disponíveis a partir dos sistemas de informação da instituição. A dificuldade está no acesso rápido e fácil destas informações.

Não é impossível que cada uma das aplicações disponíveis esteja alimentando bases de dados em diferentes SGBDs. O acesso fácil a estas informações que estão armazenadas em múltiplas bases é problemático devido à natureza do fabricante, aos formatos e à variedade de diferentes protocolos de comunicação. Este aspecto importante chama a atenção para a possibilidade de haver sistemas legados. Não é difícil imaginar o quanto, muitas vezes, estes dados são ignorados por estarem disponíveis em plataformas pouco comuns. Com os avanços da microcomputação e as facilidades que esta arquitetura dispõe em termos de interface e conectividade, os adventos em informatização são sempre afastados do *mainframe*, deixando, desta forma, para os sistemas legados, o histórico ou apenas o que ainda não foi re-implementado com a nova tecnologia (SCULLY et al, 1997).

Com o conhecimento das possibilidades de informações disponíveis e maior participação de especialistas, enfermeiras, pesquisadores médicos e administradores em atividades que requerem a avaliação comum de informação das diversas áreas, surgiu o conceito de repositório de dados clínicos.

Um repositório de dados clínicos é uma base de dados central que reúne os dados gerados por todas as aplicações operacionais da instituição, inclusive os sistemas médicos legados, para fins clínicos. Este repositório tem como principal objetivo otimizar o acesso e recuperação de informações dispersas, na instituição, em

relação a um paciente. Sendo assim, mantém uma estrutura relacional sem compromisso com os aspectos de dimensionalidade.

O NIH¹ dispõe de um repositório de dados clínicos que coleta informações armazenadas desde 1975, quando foram instalados os sistemas médicos, até os dias atuais. O repositório é atualizado continuamente com os dados correntes e constitui um único lugar para informações clínicas administrativas e de pesquisa do Instituto, com informações demográficas dos pacientes e de exames de laboratório.

Na University of Virginia Health System, foi projetado um repositório de dados clínicos para prover uma visão retrospectiva rápida, detalhada e flexível dos tratamentos de saúde realizados em pacientes de algumas instituições. Este projeto está sendo desenvolvido pela ITC-Academic Computing Health Sciences em cooperação com o Departamento de Ciências de Avaliação de Saúde (DHES), Centro Médico de Computação (MCC), Fundação de Serviços de Saúde (HSF)² e vários outros departamentos clínicos. O repositório utiliza um banco de dados relacional (Sybase) para manter uma base de dados integrada, derivada de várias bases de dados de saúde mantidas pelo MCC, HSF, departamentos individuais e algumas coleções de dados estaduais e nacionais.

Quando estes objetivos são ampliados na busca de informações consolidadas, envolvendo grupos de pacientes, para pesquisas mais complexas, é necessária a estrutura de um *data warehouse* que facilita as diversas possibilidades de cruzamento de dados.

3.3.2. *Data warehouse* clínico

No capítulo anterior, foi discutida a organização de um *data warehouse* e suas potencialidades. No contexto da área médica é importante estar atento para a natureza dos dados e as possíveis dimensões.

O escopo de um *data warehouse* clínico pode ser amplo, abrangendo desde os dados pessoais dos pacientes até dados financeiros da instituição; como também pode ser restrito a uma especialidade e composto de todos os dados inerentes a esta especialidade.

¹ NIH – National Institute of Health - <http://www.cc.nih.gov/isd/cdr/index.html>

² DHES - Department of Health Evaluation Sciences, MCC - Medical Center Computing, HSF - Health Services Foundation

Pedersen e Jensen (1998) lançaram a idéia de que um *data warehouse* clínico deve apoiar diversos níveis de análise de dados (Figura 3-1). O primeiro nível é o nível do paciente, composto por dados pessoais do paciente, voltado para prover o melhor tratamento possível. O próximo nível é o nível de grupo onde são analisados dados de um grupo de pacientes com alguma particularidade comum. Este nível é importante para analisar os tratamentos adotados e os respectivos resultados de modo a avaliar o processo, sendo também importante para a pesquisa e atendimentos médicos. O último nível é o nível da instituição de saúde, onde dados clínicos, financeiros e pessoais do paciente são combinados para que se possa analisar a qualidade dos serviços oferecidos.

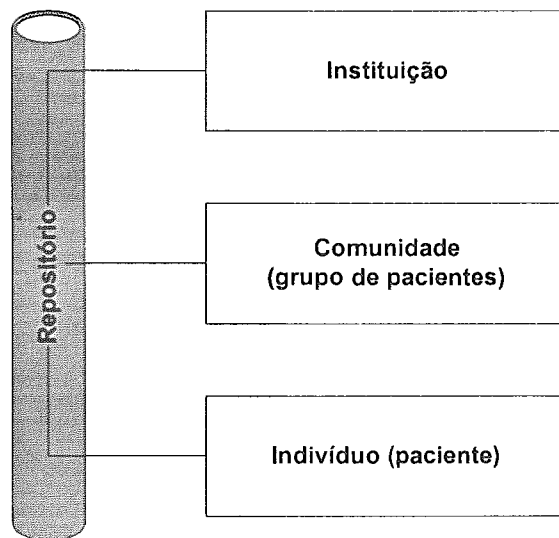


Figura 3-1 - Disposição da informação em *data warehouse* clínico

(BERNDT&HEVNER, 1998) desenvolveram uma metodologia sistemática para analisar estados de saúde de comunidades. O CATCH/IT³ é definido como um completo avaliador para permitir o rastreamento do estado de saúde da comunidade. Esta avaliação era feita manualmente, e passou a utilizar a tecnologia da informação para automatizar a metodologia, através de um *data warehouse* de larga escala, relatórios e formulários acessíveis e técnicas de *data mining* e análises estendidas. O CATCH conta com 226 indicadores de estados de saúde de múltiplas bases de dados e utiliza uma estrutura de comparação inovadora e critérios de avaliação calibrados para produzir uma lista ordenada dos problemas de saúde da comunidade. Este

³ CATCH/IT – Comprehensive Assessment for Tracking Community Health/ Information Technology

sistema tem sido usado em muitas cidades da Flórida, com atenção para as questões de saúde de maior prioridade e medindo o impacto das despesas com saúde nos retornos da comunidade.

NOVIUS.ihn *data warehouse* provê a base da infraestrutura e capacidade analíticas necessárias para entender e gerenciar custo, qualidade e desempenho. É composto por um *data warehouse* que padroniza, armazena e gerencia dados clínicos demográficos e financeiros através de uma rede integrada de saúde (IHN⁴). NOVIUS auxilia os usuários na importação, integração e gerenciamento de dados de uma variedade de fontes de dados, incluindo sistemas médicos compartilhados, sem atentar para a origem dos dados. Esta solução única melhora tanto a qualidade das decisões de negócio como a habilidade de gerenciar mudanças e desempenho operacional (NOVIUS.IHN, 1999).

3.3.3. Trabalhos relacionados

Muitos trabalhos têm sido desenvolvidos neste sentido. Algumas instituições têm trabalhado apenas em repositório de dados clínicos (NIH, UVaHS), outras apenas em *data warehouse* clínico (NOVIUS, (BERNDT&HEVNER, 1998)). Na maioria das vezes, os projetos começam como repositório de dados clínicos e são ampliados para *data warehouse*, ou possuem as duas abordagens associadas no mesmo ambiente. São discutidos alguns exemplos a seguir.

O sistema de apoio à decisão clínica desenvolvido pela MEDai⁵ possui poderosas capacidades analíticas permitindo a análise de resultados e provendo facilidades para o desenvolvimento de protocolos. MEDai é uma companhia especializada em utilizar técnicas de inteligência artificial para análise de dados médicos. Neste sistema estas técnicas são utilizadas para ajustes de riscos para os pacientes. Para a exploração dos dados são disponibilizadas técnicas de *data mining* e *drill down*. Este sistema tem características de *data warehouse* e repositório de dados clínicos para o qual são extraídos dados de um sistema hospitalar existente para compor dados clínicos e financeiros (MEDAI, 1999).

⁴ IHN – Integrated Health Network

⁵ MEDai – Medical Artificial Intelligence

O departamento de informática médica do Columbia-Presbyterian Medical Center⁶, têm desenvolvido um complexo Sistema Clínico de Informação do qual faz parte um repositório de dados clínicos e um *data warehouse* clínico. O repositório de dados contém cerca de 2 milhões de pacientes armazenados durante 10 anos. Cada registro de paciente, no repositório, consiste de uma coleção de eventos clínicos que estão relacionados à saúde do paciente. Estes eventos são informações genéricas sobre a atividade realizada na ocasião – o que foi feito, quando, por quem e onde – ou alguma informação específica sobre o evento. O *data warehouse*, cuja iniciativa foi do departamento de Ensaios Clínicos, é implementado em um banco de dados relacional (Sybase) e alimentado desde 1994 a partir de fontes de dados clínicos e financeiros, gerando uma coleção integrada de 1,5 milhões de pacientes. Além de estar conectado aos diversos departamentos clínicos do próprio centro, em 1998 o *data warehouse* passou a receber dados da base do New York Hospital, resultando em mais de 5 milhões de registros de pacientes. O *data warehouse* serve a especialistas, pesquisadores e administradores. Para manter a confidencialidade dos pacientes, os pesquisadores devem obter autorização do Institutional Review Board (IRB) para acesso aos dados.

O CareWeb⁷, desenvolvido pela University of Michigan Health System (UMHS), consiste de três componentes principais: um repositório de dados, uma interface WEB e uma coleção de rotinas para carregar e validar os dados. O repositório também está em um banco de dados relacional (Sybase), e os dados são identificados através do número de identificação do paciente e de seu número de identificação da admissão do paciente. O CareWeb oferece as facilidades de um repositório de dados clínicos no rastreamento do tratamento de um paciente e de *data warehouse* através do relacionamento de dados de vários pacientes e rotinas de *drill down*. Todos os principais elementos de dados de interesse possuem referências cruzadas que podem, facilmente, gerar visões agregadas de tratamentos de saúde envolvendo um só paciente ou todos os pacientes disponíveis. Todo ano o UMHS recebe cerca de 1 milhão de pacientes externos e o hospital da Universidade trata mais de 33.000 pacientes internados, centenas de exames de laboratório e radiológicos. Relatórios clínicos e outras informações são criados para estes pacientes em diversos sistemas espalhados na instituição.

⁶ CPMC - <http://www.cpmc.columbia.edu/edu/topics/repository.html>

⁷ CareWeb - <http://www.med.umich.edu/mcit/careweb/>

3.4. A Questão da Segurança

Segundo (JAJODIA et al, 1999), a utilidade dos sistemas de informação parte de três propriedades básicas: disponibilidade, integridade e segurança. Disponibilidade no sentido de que os recursos do sistema precisam ser funcionais e acessíveis o suficiente para atingir os objetivos propostos. Integridade para garantir um sistema internamente consistente e um modelo válido do mundo real. E por fim, a segurança, para que os usuários estejam confiantes de que não existem áreas comprometidas no sistema e que são seguidas políticas mandatórias e criteriosas.

A recente incorporação da informatização e a existência de bancos de dados trazem inúmeras possibilidades de avaliação epidemiológica a partir da utilização de registros hospitalares, ambulatoriais e certificados de óbito. A implantação de sistemas de informação suscita discussões sobre possíveis "desvios" na utilização de informação e quebra de confidencialidade de pacientes e indivíduos participantes de pesquisas. Em outras situações, grupos de interesse podem se opor ao uso de informações de saúde quando a autoria dos procedimentos pode ser identificada e utilizada para uma auditoria médica ou de outra natureza. Grupos industriais e comerciais podem considerar não adequados sistemas de informação que permitam a identificação de riscos ocupacionais e ambientais por pessoas que não fazem parte na organização.

O compartilhamento de registros de pacientes e maior facilidade de acesso às informações fazem parte dos objetivos primários de sistemas médicos. Estas necessidades chamam a atenção para um aspecto importante na concepção de sistemas médicos que é a segurança dos seus dados. Este aspecto está ligado a diversos fatores na manipulação dos dados e a ausência dele pode gerar diversos problemas econômicos, psicológicos e sociais (BARROWS, 1996). Embora se possa argumentar que os registros em papel são frágeis e despreocupados deste aspecto, a capacidade de alteração, violação e divulgação das informações com os registros eletrônicos é muito maior, devido às possibilidades de cópias de arquivos, transmissão de dados via redes, divulgação de informações a áreas de acesso público e modificação dos dados com menos possibilidade de deixar vestígios.

A natureza da confidencialidade em dados médicos relacionados ao paciente requer maior atenção aos aspectos de segurança no desenvolvimento e administração do projeto. Autenticação através da identificação do usuário e senhas, criptografia de

dados entre o cliente e o servidor e controle de acesso aos dados na rede são artifícios que podem garantir a segurança no ambiente.

A privacidade dos dados médicos disponíveis pode ser concebida em diversos sentidos dentro da instituição, entre indivíduos externos e especialistas. Cada um desses pode levantar questões importantes do contexto de segurança de dados e exigir diferentes formas de implementação. O primeiro deles é impedir o acesso aos dados de pessoas não autorizadas ou externas à instituição, evitando o conhecimento e a divulgação indevida dos dados. O segundo diz respeito à propriedade dos dados coletados e determina impedir o acesso a dados particulares por pessoas não autorizadas. E finalmente, impedir a alteração dos dados por pessoas não autorizadas internas ou externas, garantindo a integridade dos dados, um aspecto muito importante em se tratando de dados médicos, onde é primordial se tratar da vida e do bem-estar de pessoas.

Diante das categorias apresentadas, chama-se a atenção para a necessidade de se definir grupos de privilégios e propriedades e a categoria dos dados entre os grupos. Em outras palavras, é necessário definir que tipo de dado é acessível e para quem é permitido o acesso, bem como definir os recursos disponíveis aos grupos uma vez que os dados tenham sido acessados. Estas questões são importantes para se determinar uma política de segurança dentro da instituição. Muitas das tentativas de se implementar segurança em bases de dados falham pela falta de uma política adequada aos requisitos internos (BARROWS, 1996).

ASARO et al. (1999) definiram os cenários de violação de privacidade. Tomando os cenários como acessos inapropriados, são identificados indicadores motivacionais e comportamentais que apontam a violação da confidencialidade. Os indicadores são descritos em termos das características dos pacientes, do relacionamento entre o usuário do sistema e o paciente e características das sessões aplicadas. O conjunto destes fatos que compõem os indicadores, contribuem para melhorar a detecção de violação do privacidade e deve ser atualizado de acordo com as variações evidenciadas na instituição.

Em uma pesquisa feita no Canadá (KESHAVJEE, 1999), foi avaliado o comportamento de especialistas e pacientes em termos da privacidade e confidencialidade diante de um projeto chamado COMPETE⁸ que estuda como os

⁸COMPETE - Computerization of Medical Practices for the Enhancement of Therapeutic Effectiveness

medicamentos são prescritos. Os especialistas observaram que os pacientes não tinham muito interesse em saber mais sobre o projeto e estavam satisfeitos com a confidencialidade apresentadas a eles. Os pacientes que haviam utilizado o sistema pelo menos uma vez, mostraram-se preocupados com a falta de controle que poderia ocorrer com a informação em computador e apresentaram diversas restrições quanto à segurança da informação sobre sua saúde, mostrando-se interessados em saber o que acontecia com a informação mesmo que fosse feita de maneira anônima. Os especialistas não tinham noção dos pontos de vista dos pacientes e ambos os grupos sentiram necessidade de mais detalhes sobre o assunto para que pudessem julgar as vantagens e as desvantagens.

Um dos grandes desafios no desenvolvimento de *data warehouse* médico é proteger a privacidade do especialista e a confidencialidade dos dados do paciente ao mesmo tempo em que se mantém a flexibilidade das consultas.

Em repositórios de dados clínicos a maior preocupação está na confidencialidade dos dados do paciente, já que são feitas consultas diretas às informações do paciente. Em muitos casos, as consultas a informações do paciente não requerem dados específicos como nome e endereço do paciente ou nome do especialista responsável. Em um *data warehouse* os dados necessários são sobre amostras de um grupo ou resultados de um procedimento particular ou uma doença. Sendo assim, a maior preocupação é com a privacidade dos dados pois estes podem ser usados em trabalhos paralelos de pesquisa e as informações específicas podem ser vetadas. Em ambas as situações é essencial a atenção aos aspectos de segurança.

O repositório de CPMC, discutido neste capítulo, aplica artifícios de segurança baseados na autenticação do usuário, omissão de dados específicos e criptografia de dados. Desta forma, toda consulta ao repositório é preparada para remover qualquer identificador do paciente e também suas informações específicas. Dados do tipo número do registro médico, número da conta e identificador do especialista são criptografados. Com isso tem-se o objetivo de prover um nível adicional de segurança para proteger a confidencialidade e também um algoritmo que relacione números de conta criptografados ao número do registro médico atual, permitindo que os dados possam ter referências cruzadas através dos identificadores criptografados, preservando a confidencialidade do paciente e garantindo a flexibilidade nas consultas. Todo o acesso aos dados do paciente, para qualquer indivíduo, deve ser registrado e examinado quando necessário.

4. Processo para a Organização das Informações de *Data Warehouse*

Data warehouses têm sido desenvolvidos de diferentes formas e, como acontece no desenvolvimento de qualquer sistema, cada projeto tem um enfoque especial de acordo com sua área de atuação, usuários finais ou tecnologia. Nos capítulos anteriores, foi apresentado o processo de desenvolvimento genérico de um data warehouse e os diversos aspectos que envolvem os sistemas na área médica. Neste capítulo, será apresentada uma proposta de desenvolvimento de data warehouse com enfoque na aquisição dos requisitos e na organização da informação.

4.1. Análise sobre a construção de *data warehouse*

Data warehouses, conforme definido nos capítulos anteriores, atuam como um ambiente de integração de informações de uma organização independente de haver uma variedade de fontes de dados e se concentram nas diferentes visões de análise do negócio para os diferentes grupos da organização.

A construção de um *data warehouse* não é uma tarefa trivial. No capítulo II foram apresentadas diversas abordagens para a construção de *data warehouses*, cada uma com enfoque em diferentes aspectos no que diz respeito a desenvolvimento de sistemas. Foram descritas abordagens clássicas que muito contribuíram para a forma de organização do modelo de dados e construção destes ambientes, como também para a organização da arquitetura, através de Kimball (1997) e Gil&Rao (1997) por exemplo. Uma outra abordagem importante e detalhada naquele capítulo, dada através de Mattison (1998), enfatiza a aquisição dos requisitos no desenvolvimento do

data warehouse através do estudo do processo de informação que será apoiado dentro da organização. Um aspecto importante é que o autor introduz no processo a definição de grupos de negócios admitindo, assim, a coexistência de interesses distintos, os quais muitas vezes são generalizados.

Essas abordagens estão, de forma geral, muito voltadas para implementação da tecnologia, sendo assumido que o negócio e as necessidades de armazenamento (*warehousing*) são claras e previamente conhecidas. Como a característica mais marcante é a forma de modelagem dos dados, os pioneiros nesta área enfatizaram as técnicas de organização multidimensional dos dados como aqueles citados por Inmon (1997) e Kimball. Outros enfatizaram a forma de distribuição da tecnologia através de aspectos da arquitetura do ambiente como Gil&Rao. Mattison, no entanto, descreveu o desenvolvimento de um *data warehouse* através do processo de informação da organização, para qual está sendo construído o ambiente, enfatizando o processo no tipo de informação que deve ser modelada utilizando a forma, as técnicas de modelagem e tecnologia existentes.

A descrição proposta por Mattison mostra uma evolução na construção de *data warehouse*, primeiro com base tecnológica e depois com base na informação. No entanto, sua proposta de desenvolvimento é feita de forma não detalhada, sendo que muitas atividades não estão explicitamente identificadas nem definidas.

Analisando essas abordagens de construção de um *data warehouse* com as características de desenvolvimento de qualquer produto de software, percebe-se que pouca atenção tem sido dada à importância da definição de requisitos no desenvolvimento de um produto. É amplamente reconhecido no desenvolvimento de sistemas que o sucesso ou fracasso de um sistema pode estar, diretamente, relacionado à definição dos seus requisitos (PRESSMAN, 1997). Esta fase é importante para reduzir as incertezas e os riscos do desenvolvimento. Os requisitos do sistema são definidos como as necessidades, funcionais ou não, que devem ser implementadas. Os requisitos funcionais são as funcionalidades que o sistema deve prover e as informações que devem estar disponíveis. Os requisitos não funcionais se referem às

restrições físicas do sistema como a interface, desempenho, e as formas de entrada e saída.

As abordagens de construção de *data warehouse*, no entanto, pouco discutem sobre a análise de requisitos, necessária para a sua construção. Uma premissa no desenvolvimento de um *data warehouse* é o conhecimento prévio do negócio e da área que será abordada. Contudo, somente o conhecimento de como as informações circulam e do que as pessoas precisam para desenvolver suas tarefas de forma efetiva é capaz de prover as diretrizes necessárias para o armazenamento de boas decisões. Nesta tarefa, é necessário determinar as áreas de negócio dentro da instituição e os grupos de interesses específicos. Existe ainda a possibilidade de interesses cruzados entre os grupos previamente definidos que deve ser tratada e analisada (AGOSTA, 1999). A construção de um *data warehouse* requer, portanto, uma intensa e detalhada análise de requisitos permitindo, dessa forma, construir ambientes de dados que possam resolver problemas específicos do negócio.

Nas seções a seguir, é apresentado um processo de organização de *data warehouse* com foco na necessidade de explicitar como identificar e analisar requisitos para a construção de um *data warehouse*. Esse processo é definido levando em consideração as diferentes abordagens propostas, mais particularmente a de Mattison, como uma forma complementar para o enfoque de definição de requisitos.

4.2. Processo para organização das informações de um *data warehouse*: uma proposta com enfoque nos requisitos

No capítulo 2, deste trabalho, foi apresentada uma generalização do processo de desenvolvimento de um *data warehouse*. Neste processo, foi escolhida a fase inicial de análise de requisitos, para ser detalhada, adaptando-se os princípios utilizados por Mattison. Esta fase é de grande importância para promover a qualidade da organização dos dados armazenados no *data warehouse*.

O objeto de estudo utilizado neste trabalho foi o domínio médico. Embora o domínio do negócio não tenha uma influência direta nas atividades definidas, ou seja,

o processo pode ser adaptado a qualquer domínio, o processamento das informações e as diferentes áreas de interesse que já foram apresentadas em seção anterior, fornecem importantes elementos de adaptação.

Para esta fase do processo foram definidas as seguintes atividades (Figura 4-1):

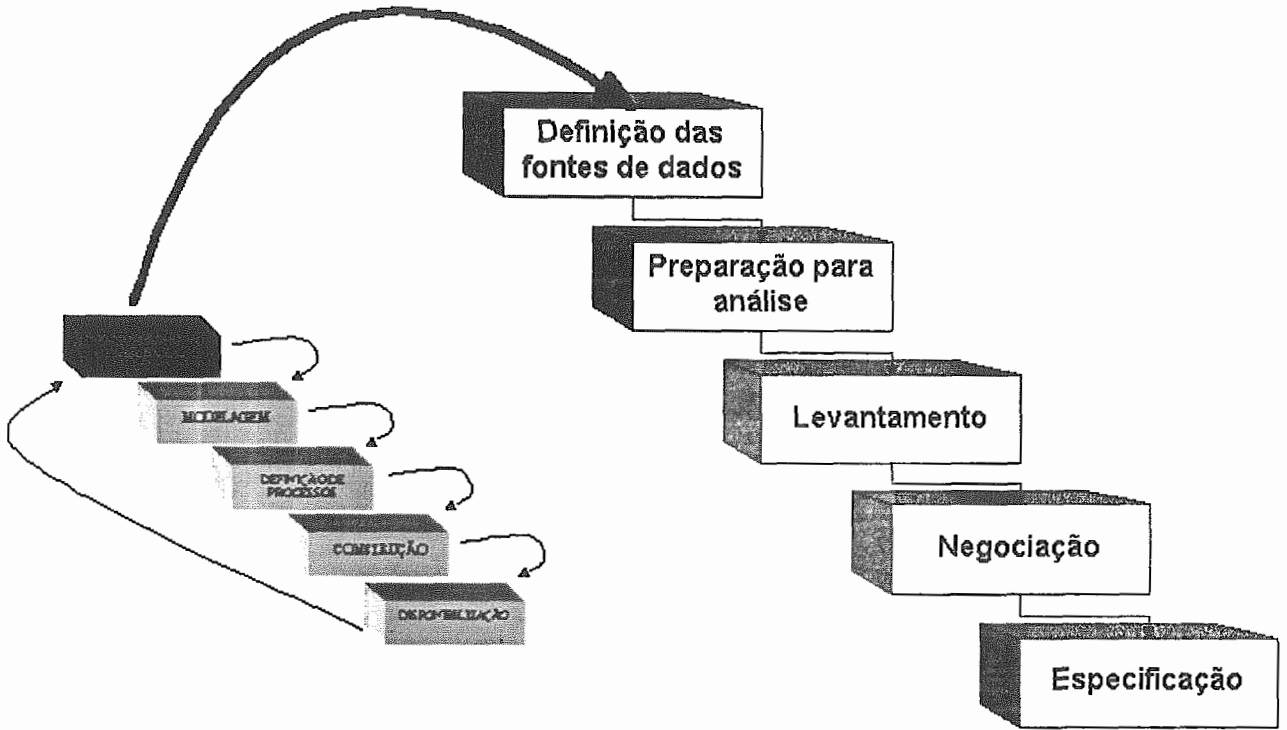


Figura 4-1 - Fases da etapa de análise de requisitos

Definição das fontes de dados

O objetivo desta atividade é avaliar as fontes de dados disponíveis, verificar a qualidade dos dados armazenados e do modelo de dados associado. Este estudo deve dar subsídios suficientes para a escolha das fontes de dados que participarão do *data warehouse*.

Um aspecto importante desta atividade é a avaliação das possibilidades de integração das fontes de dados disponíveis. Sabemos que para a construção de um *data warehouse* basta uma base de dados e interesses analíticos, mas se a instituição possui bases de dados dispersas, este ambiente pode proporcionar diversos benefícios às atividades tanto operacionais como analíticas através da integração dos dados.

Esta atividade consiste em:

- i) Identificar as bases de dados disponíveis na instituição.

O ponto de partida desta tarefa deve ser o estudo dos sistemas de informação que são utilizados para cumprir as atividades operacionais da instituição ou até mesmo as atividades complementares. Estes, normalmente, estão disponíveis em aplicativos não convencionais que exigem uma avaliação rigorosa quanto à consistência e redundância dos dados. Estes sistemas devem prover os dados essenciais para a análise dos processos que estão sendo desenvolvidos e o conhecimento dos alvos de interesse da instituição. Estes aspectos acontecem porque os sistemas são organizados, de um modo geral, para apoiar as atividades da cadeia de valor de uma organização.

Uma outra fonte importante são aquelas dos sistemas legados. Os dados provenientes destas bases ajudam a compor o volume de dados históricos muito importante em análises estatísticas, tendências e outros mecanismos gerenciais. Para esta categoria, é importante o estudo da época em que o sistema estava ativo para que seja levado em conta a validade dos dados disponíveis e quanto são relevantes em relação aos atuais. Isto se aplica a qualquer área, desde a área financeira, devido às alterações da economia, até à área médica devido às mudanças de protocolos.

Esta tarefa pode abranger até mesmo sistemas em implantação, mesmo que estes não sejam utilizados numa primeira versão. Logo nesta etapa são, também, eliminados os sistemas que não participarão do *data warehouse* em momento algum do desenvolvimento, isto é, os que não armazenam informação.

Aqueles sistemas que não possuem uma alimentação sistemática podem ser utilizados para análises isoladas, ou seja, para *data marts* muito específicos, mas se forem agrupados aos dados institucionais podem gerar desvios analíticos.

ii) Coletar informações a respeito das bases de dados.

Para os sistemas que não foram eliminados na tarefa anterior alguns aspectos genéricos devem ser conhecidos. Estes dados são:

- Tipo de usuário que acessa o sistema. Este aspecto é importante para que se tenha uma idéia da qualidade dos dados que são armazenados na base, devido à experiência e nível de atuação do usuário. Devemos pensar também que pode não haver uma diversidade de sistemas na instituição, ou seja, todos os negócios são tratados por uma mesma fonte de dados que é atualizada por diferentes grupos. A identificação dos usuários será um ponto de partida para definição dos grupos de interesse, bem como a indicação dos responsáveis pelo levantamento de requisitos;
- Categoria do sistema que dispõe da base de dados. A categoria indica basicamente a importância que o sistema possui na cadeia de valor;
- Modelo de dados utilizado para organização da base de dados. Este aspecto é importante por indicar o grau de complexidade que será tratado na manipulação da base de dados. Isto pode determinar condições de projeto como cronograma e priorização de alguns grupos de dados;
- Forma de entrada dos dados. Este aspecto pode determinar o volume de dados da base e a frequência de atualização das mesmas. Isto pode indicar escalabilidade do *data warehouse*;

iii) Avaliar os dados disponíveis em cada base.

Dos sistemas acima avaliados devem ser extraídas as seguintes características de suas bases de dados, para que mais uma vez seja avaliada a sua possibilidade de utilização no *data warehouse*.:

- Tipo de informação que é armazenada na base de dados. Este aspecto irá determinar a importância da utilização dos dados da base na organização do *data warehouse*,
 - Identificadores das entidades principais das bases de dados. Esta característica é de grande importância quando existem fontes de dados diferentes com entidades comuns e passíveis de integração. Os identificadores representam uma forte restrição à possibilidade de integração;
 - Formas de armazenamento dos dados. Este aspecto diz respeito basicamente ao tipo de dado utilizado para armazená-lo. Dados armazenados em texto livre podem ser inúteis em termos de análise;
 - Domínio de validade do dado. Este aspecto demonstra as diferenças na abordagem de dados comuns para fontes de dados diferentes. É também de grande importância quando existe mais de uma fonte e os aspectos de integração podem ser discutidos durante o levantamento;
- iv) Escolher as bases adequadas a serem utilizadas no repositório de dados.

Os diversos aspectos acima estudados a respeito das bases de dados disponíveis devem ser comparados e avaliados em termos de complexidade e viabilidade para participação no *data warehouse*, pelo menos numa primeira versão. Aquelas que não possuem razões técnicas que impossibilitem a sua utilização devem ser levadas para avaliação com o usuário já na primeira apresentação.

- v) Organizar o esquema genérico integrando as bases escolhidas, quando possível

Este esquema ajuda a entender melhor os objetos do negócio que serão tratados e fornecem uma visão ampla da malha de dados que será manipulada. É importante, também, quanto à visualização dos aspectos de integração.

Esta avaliação das fontes de dados também pode auxiliar na determinação de versões do ambiente, a partir de critérios como volume de dados e estabilidade de alimentação da base. É importante salientar que a eliminação de uma fonte de dados, a menos que seja por alguma razão óbvia, volume de dados inexpressivo ou inexistência de identificadores comuns, deve ser feita de comum acordo com os usuários após, pelo menos, a apresentação da proposta de desenvolvimento.

Preparação para a análise

Com a definição das bases de dados que serão utilizadas e conhecimento genérico da área de negócio que é tratada em cada uma delas, é possível definir grupos de interesse distintos dentro da instituição. Os grupos são definidos de acordo com os usuários responsáveis pelos dados das bases estudadas, tanto aqueles que controlam como os que utilizam a informação.

Esta atividade consiste em:

i) Definir grupos de negócio.

Identificar os grupos e as áreas de interesse de acordo com as bases de dados escolhidas. Estes grupos ou áreas são determinados por funções ou cargos dentro da empresa. Esta definição parte dos tipos de usuários dos sistemas identificados na atividade anterior. Os grupos aqui definidos serão os participantes iniciais da etapa de levantamento. Na verdade, durante a primeira apresentação da proposta, os grupos de interesse podem ser distribuídos em diferentes grupos de negócio;

ii) Avaliar a tecnologia viável em termos de cronograma pessoal e desembolso financeiro.

Nesta atividade também é necessário avaliar os diferentes tipos de ferramentas de acesso que devem ser apresentadas ao usuário, para que este escolha a melhor forma de apresentação. É importante avaliar a viabilidade de utilização de cada uma delas e definir que ferramentas serão disponibilizadas para escolha. Algumas ferramentas são bastante utilizadas e esta avaliação pode ser orientada por esta lista:

- Consultas e relatórios tradicionais, os mais simples;

- Gráficos - apresentação de relacionamentos complexos de dados e distribuição de valores;
- Análises estatísticas - utilização de pacotes específicos com cálculos estatísticos;
- Ferramentas OLAP - ferramentas para análises multidimensionais que agilizam o processamento das combinações dos dados;
- Ferramentas da WEB - *softwares* que desempenham buscas, consultas e funções de agentes no ambiente WWW;
- Agentes - rotinas programadas para avaliar eventos e buscar dados para o usuário.

Pode-se decidir apresentar ao usuário toda a avaliação ou apenas as ferramentas definidas como viáveis para implementação.

iii) Planejar o levantamento de requisitos.

Este planejamento resume-se na preparação da apresentação da proposta. Esta proposta deve conter as fontes de dados que foram avaliadas, aquelas que foram eliminadas diretamente e as razões para isto, as restrições encontradas na utilização de algumas fontes de dados seguindo aqueles critérios discutidos na primeira atividade, e apresentar os critérios que deverão ser discutidos para viabilizar a integração das fontes. A apresentação deve ser importante para esclarecer aspectos da tecnologia que ainda são desconhecidos e mostrar a importância de se ter um *data warehouse* na instituição.

Ao final desta apresentação deve-se definir os grupos de negócio que possuem processamento de informações e objetivos definidos no negócio. Para cada um deles deve-se determinar o representante que será o gestor das informações durante o levantamento de requisitos. As reuniões de levantamento devem ser individuais para cada grupo. Se possível, deve-se tentar agendar os encontros nesta primeira apresentação.

Com base nos dados avaliados e no que foi discutido durante a apresentação, pode-se estabelecer um roteiro para a atividade de levantamento de requisitos a seguir.

Levantamento dos requisitos

Esta fase deve ser realizada para cada grupo definido na fase anterior. Tem o objetivo de identificar os interesses de cada grupo, a partir da visão da sua área de negócio, e a partir daí definir e avaliar as proposições de valor pertinentes a cada um deles.

Esta atividade consiste em:

- i) Identificar requisitos, de acordo com a técnica escolhida.

Os requisitos de grupo referem-se às suas necessidades de análise dentro de suas atividades habituais. Desta forma, a identificação de requisitos deve ser conduzida no sentido de se conhecer as atividades que são realizadas pelo grupo e os dados que estão envolvidos nesta atuação.

- ii) Apresentar os dados disponíveis nas bases de dados escolhidas.

Muitas vezes o usuário final não tem um conhecimento amplo dos dados que estão disponíveis. Isto se dá, principalmente, pelo fato de que um sistema na maioria das vezes, não é construído seguindo os requisitos de apenas um usuário ou de seu grupo. Portanto, é necessário que o usuário conheça os dados que podem complementar as suas análises.

- iii) Preparação do protótipo para validação dos requisitos

Um *data warehouse* é uma aplicação não convencional para o usuário comum. Desta forma, é importante que, de posse dos dados escolhidos, se inicie a construção de um protótipo para apresentar uma idéia do que poderá ser a recuperação das informações no *data warehouse*. É importante que seja escolhido um tema simples de ser modelado mas que, ao mesmo tempo, seja de grande importância para o negócio. O objetivo, neste momento, é motivar o usuário para a utilização do ambiente e não o de lhe apresentar um produto final.

- iv) Definir as proposições de valor a partir do protótipo.

As proposições de valor são seqüências de estruturas de análise que acompanham uma certa cadeia de valor dentro do negócio. Com uma melhor visão do ambiente, o usuário se sente à vontade para desenvolver estas proposições.

- v) Avaliar as proposições de valor definidas, analisando prioridades e custos.

Estas estruturas de análise são consultas, seguindo os critérios definidos quanto ao tema abordado, os dados que serão combinados e a forma de apresentação. A seqüência destas proposições é de grande importância para o valor que proporciona. Esta avaliação é feita através da análise das seguintes questões para cada uma das proposições:

- Que valor tem cada proposição
- Qual a dificuldade em identificar e obter o valor requerido
- Qual a infra-estrutura necessária para disponibilizar o valor requerido
- Qual a complexidade no desenvolvimento da proposição
- Qual o nível de valor para as outras proposições (esta poderá ser utilizada pelas proposições seguintes?)

Esta atividade é feita com cada grupo de negócio separadamente, dos quais deve-se extrair o máximo de informação sobre suas atividades e seus objetivos profissionais. Kimball *et al* (1998) definiram conjuntos de perguntas que devem ser feitas para o usuário nas entrevistas para levantamento dos requisitos, embora sejam direcionadas à área de negócios tratando de departamentos, medidas de sucesso, impactos financeiros, enfim, são válidas para se perceber como se deve conduzir o levantamento dos requisitos.

É importante ressaltar que os grupos definidos como possíveis grupos de interesse podem não ser os mesmos que vão gerir a informação. Em outras palavras, os grupos de negócio podem ser compostos por indivíduos de grupos de interesse distintos, que trabalham em função de uma mesma cadeia de valor. Se não foi possível absorver esta diferença apenas com a avaliação da primeira atividade, deverá ser percebida no primeiro encontro com os usuários com os mesmos para promoção do projeto.

Negociação dos dados

O objetivo desta fase é otimizar a distribuição das informações entre os grupos, ressaltando a importância da colaboração entre eles. Esta fase é

necessária, principalmente, quando existem bases de dados particulares de grupos de negócio.

A negociação dos dados pode ser feita seguindo os princípios do modelo *Win-Win* (BOEHM *et al*, 1998) através do qual são negociadas especificações satisfatórias do sistema. São identificadas as condições vantajosas (*win conditions*) para cada grupo e apresentadas aos grupos de negócio.

Esta atividade consiste em:

- i) Definir vantagens para os grupos da troca de informações entre as bases de dados particulares.

Estas vantagens podem ser baseadas nos valores que serão proporcionados a cada grupo com a troca de informações entre eles.

- ii) Estabelecer a estratégia de negociação.

As estratégias devem levar em consideração os termos da negociação que vão ajudar a conduzir a atividade. São eles: os tipos de dados que podem ser públicos, o tempo em que os dados serão proprietários dos grupos, os grupos para os quais serão 'liberadas' as informações (alguns dados proprietários podem se tornar públicos apenas para a gerência).

A apresentação dos dados disponíveis deve ser cuidadosa no momento da negociação. Um grupo não pode tomar como disponível um dado de uma fonte de outro grupo. Antes de haver a negociação dos dados, cada grupo deve tomar conhecimento apenas dos seus dados disponíveis e os dados públicos da instituição. O desenvolvedor pode ser levado a apresentar todos os dados por conhecer todas as fontes.

Especificação dos requisitos

Finalmente, após definir e validar as proposições de valor, estas devem ser especificadas seguindo a metodologia de casos de uso. A definição desta fase também foi necessária para determinar uma forma de representação dos requisitos levantados, que não foi descrita explicitamente no processo por proposição de valor.

Esta atividade consiste em:

- i) Descrever os requisitos definidos em casos de uso.

Uma característica da descrição dos requisitos em casos de uso é utilizar uma linguagem simples para que qualquer um seja capaz de entender o que será feito pelo sistema. Da mesma forma que os processos definidos para aplicações desta categoria são pouco complexos comparados àqueles para aplicações de transações. Sendo assim, os casos de uso são um instrumento importante na validação dos requisitos definidos baseados em proposições de valor.

ii) Concluir o protótipo para validação da especificação.

Esta conclusão já engloba a utilização das proposições definidas em termos de dados e formas de apresentação. Por esta razão, pode-se dizer que temos aqui uma primeira versão do ambiente, que ainda não está preparada para atualizações.

A especificação dos requisitos em casos de uso chama a atenção para alguns aspectos comuns entre processo de desenvolvimento por proposição de valor e aquele proposto pela UML¹ (SCHNEIDER&WINTERS, 1998). A natureza do desenvolvimento de ambientes baseados em repositórios de dados impõe que o levantamento de requisitos seja voltado para as necessidades de um usuário e não do sistema como um todo. Esta é a principal característica do levantamento de dados utilizando casos de uso: a preocupação com as necessidades do grupo e não do sistema. Com isto, vê-se a necessidade de definição de grupos de interesses comuns que possam conduzir a elicitação.

Cada atividade do processo trata de um objetivo específico e gera um produto que deve ser útil na próxima etapa. É importante ressaltar que, a cada iteração do ciclo de desenvolvimento, esta etapa de análise de requisitos deve ser reavaliada, levando-se também em consideração os resultados obtidos nas iterações anteriores. As avaliações feitas para as fontes de dados e ferramentas, os grupos e áreas de interesses definidos podem ser aproveitados e atualizados.

Uma descrição mais detalhada de cada atividade é encontrada na Tabela 4-1. Cada produto pode ser refinado à medida em que as etapas são amadurecidas.

¹ UML – Unified Modeling Language

Tabela 4-1 - Atividades do processo de análise de requisitos

Atividades	Procedimentos	Produtos
Definição das fontes de dados	<p><u>Identificar as bases</u> disponíveis na instituição;</p> <p><u>Coletar informações</u> a respeito das bases: tipo de usuários, modelo de dados, sistema, forma de entrada;</p> <p><u>Avaliar os dados</u> disponíveis em cada base: tipo de informação, identificadores, formas de apresentação, domínios;</p> <p><u>Escolher as bases</u> adequadas para serem utilizadas no repositório de dados;</p> <p><u>Organizar o esquema genérico</u> integrando as bases escolhidas quando possível.</p>	<ul style="list-style-type: none"> • Avaliação das bases de dados estudadas • Resultado da escolha das bases • Listagem dos dados disponíveis
Preparação para a análise	<p><u>Definir grupos de negócio</u>, ou seja, identificar os grupos e as áreas de interesse de acordo com as bases de dados escolhidas;</p> <p><u>Avaliar a tecnologia viável</u> em termos de cronograma, recursos financeiros e pessoal.</p> <p><u>Planejar o levantamento de requisitos</u>: definir os representantes do grupo, agendar encontros, estabelecer roteiro;</p>	<ul style="list-style-type: none"> • Definição dos grupos e suas respectivas áreas de interesse • Definição das possibilidades de apresentação do dados
Levantamento dos requisitos	<p><u>Apresentar os dados disponíveis</u> nas bases de dados escolhidas;</p> <p><u>Elicitar requisitos</u>, de acordo com a técnica de elicitação escolhida;</p> <p><u>Preparação de protótipo</u> para validação dos requisitos;</p> <p><u>Definir as proposições de valor</u>, a partir do protótipo preparado;</p> <p><u>Avaliar as proposições de valor</u> definidas, analisando prioridades e custos.</p>	<ul style="list-style-type: none"> • Definição das proposições de valor de cada grupo • Definição das prioridades de cada proposição
Negociação dos dados	<p><u>Definir vantagens</u> para os grupos na troca de informações entre as bases de dados proprietárias;</p> <p><u>Estabelecer a estratégia</u> de negociação.</p>	<ul style="list-style-type: none"> • Definição de grupos e entidades de interesse (tabela entidade X grupo)
Especificação dos requisitos	<p><u>Descrever as proposições</u> definidas em casos de uso;</p> <p><u>Concluir o protótipo</u> para validação da especificação.</p>	<ul style="list-style-type: none"> • Descrição de casos de uso • 1ª versão da aplicação

4.2.1. Considerações do processo

A tentativa de se gerar um modelo integrado das bases participantes deve ser feita para auxiliar a visualização dos aspectos de integração das diversas fontes de dados escolhidas. A princípio, o modelo não deve conter características de um esquema multidimensional pois nesta fase ainda não foram definidos os fatos e dimensões e o estudo preliminar não foi suficiente para se formar o modelo desta categoria.

No contexto do processo através de proposição de valor, é importante deixar claro para o usuário a tecnologia disponível e os benefícios em relação às suas necessidades. Ou seja, são levantados nesta fase os requisitos não funcionais do sistema. Por esta razão, a avaliação da tecnologia disponível é fundamental para verificar a viabilidade de cada uma delas para que o usuário tome conhecimento apenas daquelas que são possíveis de serem implementadas. A viabilidade é avaliada em termos de recursos financeiros, da necessidade de aquisição de ferramentas e custo de suporte, do cronograma, da disponibilidade de tempo hábil para aquisição, aprendizado e implementação da ferramenta, de pessoal, da necessidade de recursos especializados para implementação e necessidade de suporte. Esta avaliação pode ser negociada com o usuário final, de acordo com os investimentos planejados.

Um protótipo é sempre importante para o usuário visualizar a aplicação e desenvolver novas idéias, principalmente, quando os usuários não estão habituados com os elementos da informática, como pode acontecer no ambiente médico. O protótipo elaborado durante a etapa de levantamento auxilia a elicitação das proposições de valor a partir dos grupos de dados definidos.

No contexto de ambientes com *data warehouse*, pode-se perceber que o pior caso na tentativa de integração de bases de dados particulares e públicas é quando cada grupo de negócio só tem interesse e disposição para integração dos dados de sua fonte com os dados públicos, ou na elaboração de um repositório de pesquisa da área (*data mart*). Sendo assim, pode ser necessária uma intervenção gerencial nesta fase, no sentido de se promover uma renegociação (Figura 4-2).

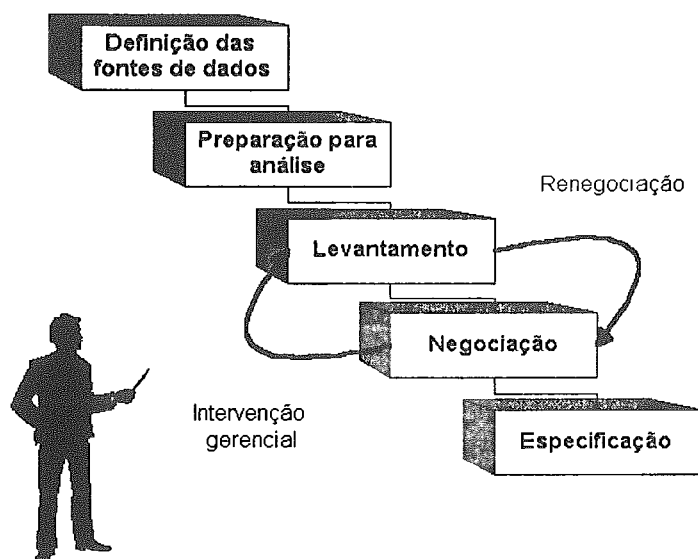


Figura 4-2 - Intervenção gerencial na negociação

Após a validação dos casos de uso através da especificação e protótipo concluídos pode-se considerar que uma primeira versão do ambiente já está disponível, uma vez que os dados que podem ser manipulados compõem um subconjunto ou talvez toda a parte dos dados que se deseja. O ambiente estará pronto para refinamentos em novos dados e definição de consultas. O desenvolvedor pode agora estar envolvido com atividades de construção e manutenção do ambiente.

4.3. Considerações Finais

A idéia principal deste trabalho é experimentar este processo na construção de um *data warehouse* clínico, com algumas adaptações. Em uma instituição de saúde, os valores a serem atingidos são mais amplos do que as atividades que fazem parte do processo de transformação. Cada atividade do processo pode ser dirigida por diferentes grupos especializados, a depender das necessidades do paciente, os verdadeiros beneficiados na cadeia de valores dessas instituições.

5. Experiência na Construção de um *Data Warehouse* para Cardiologia

Este capítulo apresenta o ambiente utilizado para o estudo e a experiência na organização de um ambiente para análise de informações no domínio de cardiologia.

5.1. Os sistemas de informação da UCCV/FBC

A Unidade de Cardiologia e Cirurgia Cardiovascular/Fundação Bahiana de Cardiologia (UCCV/FBC) é um dos principais centros cardiológicos do estado da Bahia e de enorme importância na formação de cardiologistas. Este destaque pode ser atribuído aos constantes investimentos em equipamentos e inovações tecnológicas e à qualidade do seu corpo médico. Esta instituição já contribuiu com diversos projetos envolvendo sistemas da área médica de diferentes categorias, desde 1994, através do convênio COPPE/FBC e diversos trabalhos de teses de mestrado e doutorado. A instituição também mantém um grupo de trabalho para a desenvolvimento de novas áreas de pesquisa e manutenção dos sistemas existentes, o Núcleo de Pesquisas e Desenvolvimento de Software Médico (NPqD).

Dos sistemas desenvolvidos neste contexto, o primeiro deles foi o SEC (OLIVEIRA, 1995), um Sistema Especialista de Cardiologia. Este sistema avalia, a partir dos dados clínicos da visita e da interpretação do eletrocardiograma do paciente, a possibilidade deste estar sofrendo um evento coronariano agudo. Este sistema é destinado a apoiar médicos não especialistas em cardiologia no diagnóstico de eventos coronarianos em locais onde não se dispõe de um cardiologista.

A UCCV/FBC, como uma unidade do Hospital Universitário da UFBA, possui como um de seus objetivos a formação de profissionais médicos desde a graduação a pós-graduação em nível de residência médica, mestrado e doutorado. É prática regular na instituição, uma sessão clínica semanal. Nesta sessão, são discutidos

casos clínicos importantes ocorridos durante a semana. Esta discussão, coordenada pelo chefe da unidade, com a presença de internos, residentes, cirurgiões, cardiologistas clínicos e especialistas nos diferentes tipos de exames cardiológicos, busca identificar a conduta mais adequada ao paciente, que pode ser tratamento clínico, intervenção hemodinâmica ou cirurgia. Para apoiar a preparação e a realização desta sessão, foi desenvolvido o HiperClínica (GAMA, 1996), um sistema cooperativo hipermédia. Neste sistema são registrados os casos de pacientes a serem discutidos, com os dados e exames envolvidos.

O reconhecimento da importância deste sistema para a assistência e o ensino motivou o desenvolvimento do projeto CardioEducar, apoiado pelo PROTEM-CC do CNPq, que inclui a construção de novos ambientes de apoio a sessões da UCCV/FBC. Entre estes já está operacional o ambiente CardioCirurgia para apoiar as sessões de planejamento e acompanhamento cirúrgico e que contém uma base de dados de pacientes submetidos à cirurgia cardíaca.

Outra aplicação importante desenvolvida na instituição, é uma aplicação de teleconsulta em cardiologia, o sistema TeleCardio (LIMA, 1999), destinado a promover a cooperação entre especialistas de cardiologia da UCCV/FBC e outros especialistas de Salvador e/ou do interior do estado. Esta cooperação tem o objetivo de auxiliar os especialistas na elaboração dos planos terapêuticos de seus pacientes, na definição de diagnósticos e no acompanhamento remoto de pacientes da instituição que retornaram a seus domicílios em outras cidades.

A instituição também conta com um sistema de acesso público para informação de pacientes sobre o exame de cintilografia miocárdica (VALLE, 1997). Este não é um exame comum e o seu procedimento não é trivial, o que implica na existência de restrições quanto a alimentação, vestimenta e outros, que devem ser de conhecimento do paciente, antes do seu comparecimento para a realização do exame. Usando um quiosque multimídia o paciente pode ter fácil acesso a estas informações e aos procedimentos envolvidos no exame.

A UCCV/FBC dispõe também de um sistema de informações hospitalares, o SIGAH, atualmente em sua 2ª versão. Este sistema contém vários módulos que apóiam, entre outras funcionalidades, a emissão de laudos dos exames realizados na unidades e o acompanhamento dos pacientes internados através do prontuário eletrônico.

Também foram desenvolvidas bases de dados particulares para pesquisas epidemiológicas a partir do trabalho assistencial realizado nos ambulatórios do

Hospital Universitário. Cada base de dados foi definida por um grupo distinto dentro da UCCV/FBC e seus dados são alimentados a partir de formulários especificados e preenchidos pelos especialistas envolvidos. Estas bases são:

- **Miocardiopatias:** base de dados de pacientes do ambulatório de miocardiopatias;
- **EPICOR:** base de dados com a evolução de pacientes que estiveram internados na Unidade Coronariana da UCCV/FBC com infarto ou angina;
- **ACID:** bases de dados do ambulatório dislipidemia e cardiopatia isquêmica;
- **HiperCardio:** base de dados do ambulatório de hipertensão arterial;
- **Valvulopatias:** base de dados do ambulatórios de valvulopatias.

Diante desta apresentação não é difícil perceber o potencial da UCCV/FBC para a construção de um ambiente de dados. Dos sistemas citados, apenas o sistema de acesso público não está disponível para um repositório de dados por não armazenar informação dos pacientes.

5.2. Processo de transformação da informação médica

Cada instituição de saúde tem um perfil de acordo com os serviços que são fornecidos à população. Esta característica não envolve a especialidade à qual a instituição se dedica. Em lugar disso, este perfil conduz as atividades que agregam valor para a instituição aos objetivos que esta deseja atingir no seu campo de negócio. Alguns destes objetivos podem ser:

- **Cobertura da comunidade:** os recursos existentes são suficientes para atender às necessidades da comunidade local;
- **Qualidade do corpo clínico:** preocupação com a formação do corpo clínico, investindo em treinamentos e avaliações científicas periódicas;
- **Modernização:** são feitos grandes investimentos em tecnologia, provendo serviços com equipamentos modernos;
- **Bem-estar do paciente:** atenção para as instalações da instituição, serviços de hotelaria e atendimento ao paciente pelo pessoal administrativo e de apoio

É importante observar que as características dos propósitos acima descritos, para instituições de saúde, estão na relevância que é dada às atividades principais ou atividades de apoio, da mesma forma que se pode admitir que todas instituições

buscam atingir esses objetivos em comum, e o que faz a diferença é o interesse e quantidade de recursos que possui para os investimentos em cada setor.

Na área médica, o processo de decisão trata não só da definição do diagnóstico como também de toda a propedêutica utilizada para a solução dos problemas do paciente. Os dados obtidos na anamnese do paciente, no seu exame físico e nos exames complementares que realizou, compõem uma massa de informações que devem ser combinadas com o conhecimento e a experiência do especialista. Neste processo, espera-se alcançar o diagnóstico mais preciso e, a partir daí, definir a melhor conduta terapêutica, ou seja, aquela capaz de solucionar o problema do paciente da forma mais eficaz (Figura 5-1).

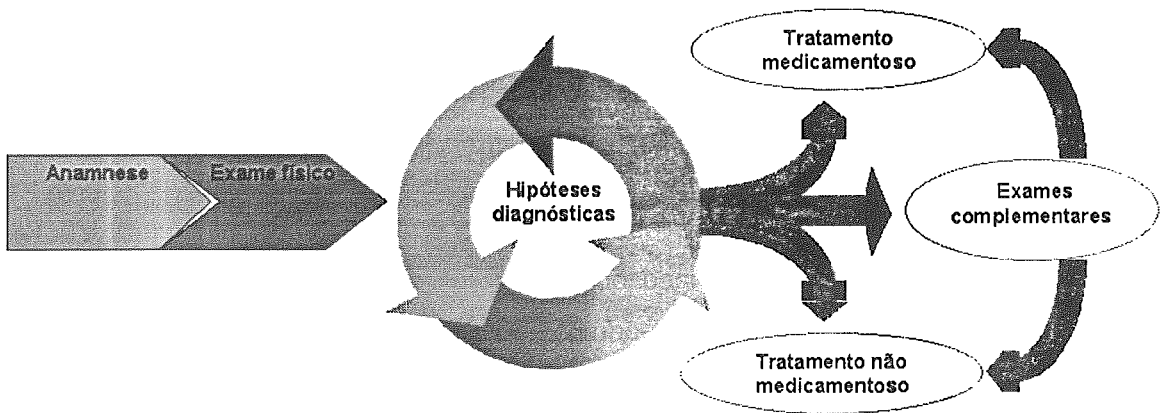


Figura 5-1 - Processo de transformação da informação médica

Este é um cenário básico relacionando os dados envolvidos nas decisões médicas para a assistência de um paciente, e que pode ocorrer em diferentes situações e locais dentro da instituição.

Neste ponto, é preciso determinar cadeias de valores para os níveis de atuação da instituição:

- nível do paciente, caracterizado pelas atividades da medicina curativa, voltadas para a cura e bem-estar do paciente;
- nível da comunidade, caracterizado pelo levantamento da frequência da doença na população e avaliação dos resultados de diferentes alternativas terapêuticas. Este nível é sustentado, basicamente, por pesquisas epidemiológicas, envolvendo, portanto, a medicina preventiva;
- nível da instituição, caracterizado por tratar os aspectos administrativos visando melhorar a qualidade das atividades clínicas e adequar custos.

É necessário definir as cadeias de atividades que fazem parte de cada um desses níveis. A idéia da análise da cadeia de valor é determinar um núcleo de habilidades e objetivos que regem os diversos níveis de atuação da organização. Neste sentido, as principais atividades destas cadeias são basicamente (Tabela 5-1):

- atendimento médico – admissão do paciente e avaliação clínica;
- serviços auxiliares de diagnóstico e tratamento (SADT) - conjunto de ações destinadas ao desempenho de algum procedimento;
- administração clínica - desenvolvimento, aquisição e alocação de recursos para sustentação da instituição; e
- pesquisas epidemiológicas – investigações de casos para determinar a distribuição e fatores determinantes de um problema.

Tabela 5-1 - Atividades da cadeia de valor

Atendimento Médico	Ambulatório		Emergência		Enfermaria	UTI
	SADT	Exames Complementares	Procedimentos terapêuticos	Fisioterapia		
Administração clínica	Técnica		Administrativa e financeira			
	Corpo Clínico	Departamentos	Recepção	Serviços gerais	Controle de estoque	Faturamento
Pesquisas Epidemiológicas						

Uma instituição de saúde existe para atender pacientes e se estabelece porque fatura. Desta forma, não é difícil notar que as atividades principais são aquelas que envolvem diretamente o paciente como o atendimento médico e atividades de SADT e as atividades de apoio (aquelas que envolvem a qualidade e o desempenho dos serviços prestados ao paciente).

A partir da definição das atividades gerais, pode-se caracterizar cada nível de atuação da instituição com uma cadeia de valor (Tabela 5-2):

Tabela 5-2 - Cadeias de valores nos níveis de atuação

Nível	Cadeias de valores
Paciente	Atendimento Clínico → SADT → Faturamento
Comunidade	Atendimento Clínico → SADT → Pesquisas Epidemiológicas
Instituição	Administração Clínica

É importante ressaltar que o *data warehouse* clínico atua mais na área da medicina preventiva, onde são realizadas pesquisas epidemiológicas, e na análise de custos e serviços prestados. Estas investigações envolvem análise de condições de vida e saúde, estudos de mortalidade, inquéritos populacionais, ensaios clínicos, estudos etiológicos, avaliação de impacto de ações e atividades de saúde. Sendo assim, têm mais valor quanto maior for o volume e a diversidade dos casos de estudo.

Embora o *data warehouse* esteja sendo usado há muitos anos, as instituições de saúde ainda estão começando a entendê-lo e a perceber os seus benefícios. Em recente relatório da SYBASE sobre *data warehouse* como a grande arma para competição na área de saúde, foi constatado que o mercado de *data warehouse* cresceu de \$4 bilhões em 1998 para \$6.5 bilhões em 2000 em função desta categoria. Este crescimento se deve aos valores que são proporcionados com a utilização do *data warehouse*, como: gerenciar um melhor relacionamento com o paciente, permitindo um acompanhamento completo das informações do paciente, inclusive alertas de aniversários e datas previstas de retorno; gerenciamento de melhorias nos serviços através da análise de custos e qualidade dos serviços prestados; redução de fraudes, com facilidades para sua detecção e recursos solução; utilização de relatórios como recursos preventivos para que através deles sejam identificadas as áreas que necessitam de melhorias; gerenciamento da doença, monitorando e analisando as condições de cada paciente para agilizar e efetivar decisões no tratamento.

5.3. Experiência na organização das informações de um *data warehouse* para Cardiologia

Com a diversidade de aplicações médicas que a UCCV/FBC dispõe, fica muito difícil, para um membro da instituição, ter acesso a todas as informações disponíveis com agilidade e em tempo hábil. Um dos incentivos para a construção de um *data warehouse*, nesta instituição, é permitir que seus membros, participantes dos diversos

grupos internos, tenham acesso fácil e direcionado a seus interesses, em um só dispositivo de informação.

Com um ambiente integrado, será possível que os usuários de todos os outros sistemas/bases tenham acesso a seus dados, gerados no momento de sua rotina de trabalho, em conjunto com dados de outras áreas, que serão utilizados em sua rotina analítica.

5.3.1. Aplicação do processo

O desenvolvimento do ambiente de dados integrados na UCCV/FBC foi baseado no processo de análise de requisitos apresentado no capítulo anterior. As etapas foram refinadas, à medida em que iam sendo experimentadas.

Um ponto crítico no início do desenvolvimento é partir para o estudo de todas as bases disponíveis e não apenas daquelas de interesse do grupo responsável pela iniciativa. Isto evita que o projeto não deixe de ser apresentado a todos os membros de interesse, mas não implica que serão incorporados ao ambiente na primeira fase. O processo passa por uma fase de definição de prioridades que estuda, exatamente, a seqüência utilização dos requisitos levantados e as possibilidades de cada grupo.

5.3.1.1. Definição das fontes de dados

Nesta etapa foram avaliadas todas as aplicações disponíveis na instituição, as quais foram apresentadas na primeira seção deste capítulo. Esta avaliação utilizou os seguintes critérios (Tabela 5-3):

- categoria do sistema: para avaliar o tipo de informação que pode ser armazenada pela aplicação e qual o tipo de usuário que utiliza o sistema;
- base de dados: para avaliar se a aplicação gera uma base de dados, qual o tipo – arquivos texto, bancos de dados (tipo do SGBD) – e verificar a possibilidade de sua utilização;
- identificador: para avaliar quais os principais identificadores da base de dados e verificar a possibilidade de comunicação com as outras bases;
- características do modelo: para avaliar a complexidade do modelo de dados que determina como os dados estão armazenados. Esta informação será útil na definição da prioridade de utilização das bases;
- estabilidade: para avaliar o volume de dados armazenados que podem ser utilizados.

Com este estudo percebeu-se que conceitualmente os modelos de dados eram muito parecidos. Alguns possuíam algumas tendências especiais derivadas do objetivo da aplicação existente, mas no que se refere ao domínio médico mantinham entidades comuns. Esta observação ajudou na preparação do modelo genérico, reponsável pela carga dos dados iniciais do ambiente.

Tabela 5-3 - Avaliação dos sistemas disponíveis

Sistema	Categoria do sistema	Base de dados	Identificador	Características	Estabilidade
SEC	Sistema especialista Usuários: médicos de ambulatório	Banco de dados – ACCESS	Número caso	Modelo de dados simples, e apenas 1 índice	Mais de 1500 casos catalogados
HiperClinica	Aplicação de Hipermídia Usuários: médicos de ambulatório, residentes, internos	Banco de Dados – ACCESS	Registro HUPES	Grande número de campos textos	Alimentação não regular
TeleCardio	Aplicação de teleconsulta Usuários: médicos de ambulatório e enfermaria	Banco de Dados - ACCESS	Registro HUPES	Dados estruturados do prontuário	Em implantação
SIGAH Laudos (legado)	Sistema de informação Usuários: médicos especialistas em exames	Banco de dados Unify	Registro FBC	Todo o laudo do exame é armazenado em um campo binário	Base de dados legada.
SIGAH – Multimídia Dignóstico	Sistema de informação Usuários: médicos especialistas em exames	Banco de dados – SYBASE	Registro HUPES	Modelo de dados complexo, e uso de descritores para todos os dados	Mais de 20000 exames de pacientes
EPICOR	Aplicação da pesquisa Usuários: médicos de enfermaria (UTI)	Banco de dados – ACCESS	Registro HUPES	Conjunto específico de dados estruturado	Registros de casos escolhidos
ACID	Aplicação da pesquisa Usuários: médicos de ambulatório	Banco de dados – ACCESS	Registro HUPES	Conjunto específico de dados estruturado	Registros de casos escolhidos
Miocardopatias	Aplicação da pesquisa Usuários: médicos de ambulatório	Banco de dados – ACCESS	Registro HUPES	Conjunto específico de dados estruturado	Registros de casos escolhidos

5.3.1.2. Preparação para a análise

Após o estudo das bases de dados disponíveis foram verificadas as bases com potencial para serem utilizadas. Com isto, também, foi possível inferir na definição de alguns grupos de interesse, que seriam os grupos de usuários das aplicações em questão: médicos de ambulatório, médicos especialistas de exames, médicos de enfermarias, médicos de UTI. A definição dos grupos foi importante na definição dos participantes da primeira reunião de apresentação do trabalho.

As tecnologias que poderiam ser utilizadas foram avaliadas segundo a tecnologia disponível na instituição, utilizadas nas aplicações existentes, na utilidade de outras que seriam específicas para a aplicação que estava sendo desenvolvida (como é o caso da ferramenta OLAP), e das necessidades que poderiam existir em termos do que já estava sendo feito individualmente, como, por exemplo, análises estatísticas.

Foi feita uma apresentação geral da tecnologia e da motivação para a implantação do ambiente na UCCV/FBC. Esta apresentação definiu os grupos de negócio – especialistas interessados em síndromes coronarianas agudas, em cardiopatias isquêmicas e dislipidemias ou em miocardiopatias - motivando aqueles que ainda não haviam sido consultados e deixando claras as necessidades de limpeza dos dados e qualidade no armazenamento da informação, para garantir a qualidade do ambiente.

Os grupos de negócio foram definidos como grupos formados por membros de grupos de interesse diversos da instituição, que direcionam seus interesses em negócios específicos (tabela 5-4).

Tabela 5-4 - Definição de áreas de interesse e de negócio

Área de interesse	Área de Negócio
Médicos especialistas em exames	Insuficiência Cardíaca e Miocardiopatias
Médicos de ambulatório Médicos de UTI	Síndrome Coronariana Aguda
Médicos de enfermaria Médicos de ambulatório	Cardiopatias Isquêmicas e Dislipidemias
Médicos especialistas em exames	Hipertensão
Médicos especialistas em exames	Arritmias
Médicos especialistas em exames	Hemodinâmica
Médicos especialistas em exames	Medicina Nuclear

Indiretamente ainda havia o grupo com interesses administrativos que viria manifestar-se mais tarde, da mesma forma que este nível de interesse dentro dos grupos definidos.

5.3.1.3. Levantamento dos requisitos

Todas as informações coletadas nas etapas anteriores foram trazidas para as etapas seguintes e apresentadas aos usuários de interesse. A partir da reunião de apresentação e motivação dos participantes, foram determinados os grupos de negócio e agendadas reuniões individuais com cada um deles.

O objetivo das entrevistas com os grupos de negócio é definir os dados de interesse para as suas atividades. Uma recomendação de Kimball (KIMBALL, 1996c) é em momento algum perguntar “quais as consultas que serão do seu interesse?”, a qual é de extrema importância, em se tratando do domínio médico, que tem por característica o raciocínio não estruturado e a subjetividade nas inferências. Esta é uma pergunta vaga, e que normalmente seremos levados a fazer. O mais correto é partir para conhecer o negócio e as atividades envolvidas pelo grupo. Desta forma fica mais difícil para o analista, mas é mais fácil para o usuário definir a sua participação no processo.

Com as informações iniciais foi possível se construir um protótipo e, com isso, num segundo encontro foi possível definir com mais segurança as proposições de valor. Nesta segunda avaliação, foi possível esclarecer os aspectos de integração de dados.

O modelo de dados generalizado na etapa de definição das fontes de dados serviu de apoio no levantamento de requisitos gerais. Como a organização é constante, os usuários só precisavam expor as suas atividades e contribuições nas cadeias de valor, definindo as proposições que mais se adequavam a seus interesses. Mais detalhes podem ser vistos no anexo 1.

Entre os grupos de negócio, foi observada uma forte tendência ao nível de pesquisas epidemiológicas. Todos eles já conduzem atividades de pesquisa em seu dia a dia, seja através de suas bases de dados particulares ou utilizando ferramentas anexas (MSExcel, SPSS). Da mesma forma que alguns apresentavam interesse em rastrear informações de seus pacientes caso estes tenham estado em mais de um setor ou mesmo em outras ocasiões para análise de histórico, configurando assim o nível de informações de paciente.

Tabela 5-5 – Grupos de negócio e dados de interesses

Grupos de Negócio	Arritmias	Cardiopatias Isquêmicas	Dislipidemias	Hemodinâmica	Hipertensão	Medicina Nuclear	Insuficiência Cardíaca e Miocardiopatias	Síndrome Coronariana Aguda
Arritmias								
Cardiopatias Isquêmicas			Pacientes com/sem eventos coronariano prévio e dados (medicamentos, sintomatologia, dislipidemia, diabetes, sexo, idade)					Pacientes c/ dor torácica e enzimas cardíacas Pacientes c/ angina instável e (ECG, ECO, CATE) Pacientes c/ IAM e curva enzimática e prognósticos
Dislipidemias								
Hemodinâmica								
Hipertensão								
Medicina Nuclear								
Insuficiência Cardíaca e Miocardiopatias		Pacientes com ICC e sua classe funcional	Pacientes com ICC e sua classe funcional		Pacientes com ICC e sua classe funcional			Pacientes da UCO com ICC e etiologia da ICC
Síndrome Coronariana Aguda		*	Dados de dislipidemias de pacientes com infarto ou angina (EPICOR)	Coronariografia de pacientes com angina (EPICOR)			Pacientes com ICC isquêmica	

5.3.1.4. Negociação dos dados

Com a diversidade de grupos de negócio, cada um sendo responsável por uma fonte de dados, não seria difícil despertar interesses mútuos nos dados alheios. Na verdade, partindo-se da natureza da maioria das bases utilizadas nesta fase, seria natural que os grupos apresentassem interesses apenas em dados de fontes de outros grupos (Figura 5-2).

Devido à dificuldade de se reunir todos os grupos em um só momento para que fossem discutidas as trocas de dados, foram apresentadas as necessidades separadamente, para cada grupo, já em termos de proposição de valor. O grupo presente avaliava os dados tratados pelos outros grupos de negócio que pudessem agregar algum valor comum aos seus.

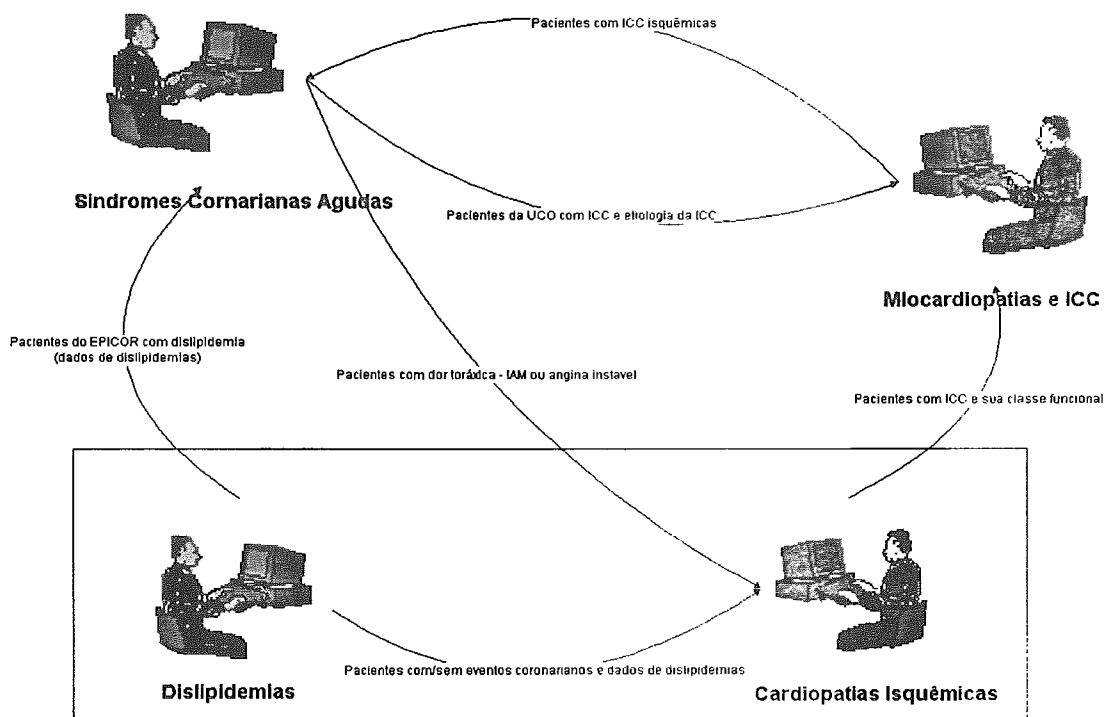


Figura 5-2 - Cenário da negociação

Foi verificado que os grupos cediam informações entre si apenas para os pacientes que fossem comuns às bases de dados em questão. O grupo público, denominado aqui como FBC, não teria acesso às informações restritas aos grupos de negócio, embora estes também façam parte do grupo FBC. Entretanto os dados das bases de dados particulares poderiam ser cruzados livremente com os dados da base pública na visão de cada grupo de negócio.

5.3.1.5. Especificação dos requisitos

Após realizada a negociação entre os grupos, tem-se dados suficientes para iniciar a especificação dos perfis de acesso e dados restritos aos níveis de análise. Os casos de usos, também, já podem ser definidos e devem marcar claramente as fronteiras entre os grupos e a origem das informações de interesse de cada um.

Os fatos foram modelados para cada nível de análise de cada grupo. Os grupos particulares também utilizaram dados da base pública, os quais não necessitava de negociação. Um exemplo, de modelagem de um dos fatos para o nível de Instituição do grupo FBC e está no anexo 2.

A elaboração de um protótipo é importante para validação dos casos de uso e, por conseqüência, dos requisitos levantados. Exemplos de casos de uso foram colocados no anexo 2. Esta validação deve tratar não só os requisitos de informação como também as funções que estarão sendo disponibilizadas para manipulação.

Esta segunda avaliação conclui a fase de análise de requisitos e uma primeira versão da aplicação. A primeira fase do desenvolvimento tem ênfase nos dados e informações que serão contempladas. Os processos de construção e manutenção serão definidos após se disponibilizar para o usuário recursos suficientes para que seus requisitos possam ser melhor explorados.

5.4. Descrição do ambiente

Considerando-se as etapas descritas anteriormente, foi desenvolvido um ambiente para análise de dados em cardiologia com os conjuntos de dados e níveis de análise determinados por cada grupo de negócio que compõe a instituição e que tinha participado desta primeira abordagem. Como foi mencionado, após a primeira fase do desenvolvimento do ambiente, já se dispunha de uma primeira versão da aplicação para manipulação dos dados que estaria disponível para refinamentos e utilização restrita.

5.4.1. Arquitetura do ambiente

A arquitetura do ambiente foi sendo definida à medida em que a fase de análise de requisitos evoluía, seguindo as exigências do processo. A camada de definição e processos foi incrementada após a conclusão desta fase. As formas de apresentação dos dados foram incorporadas após a liberação da primeira versão (Figura 5-3).

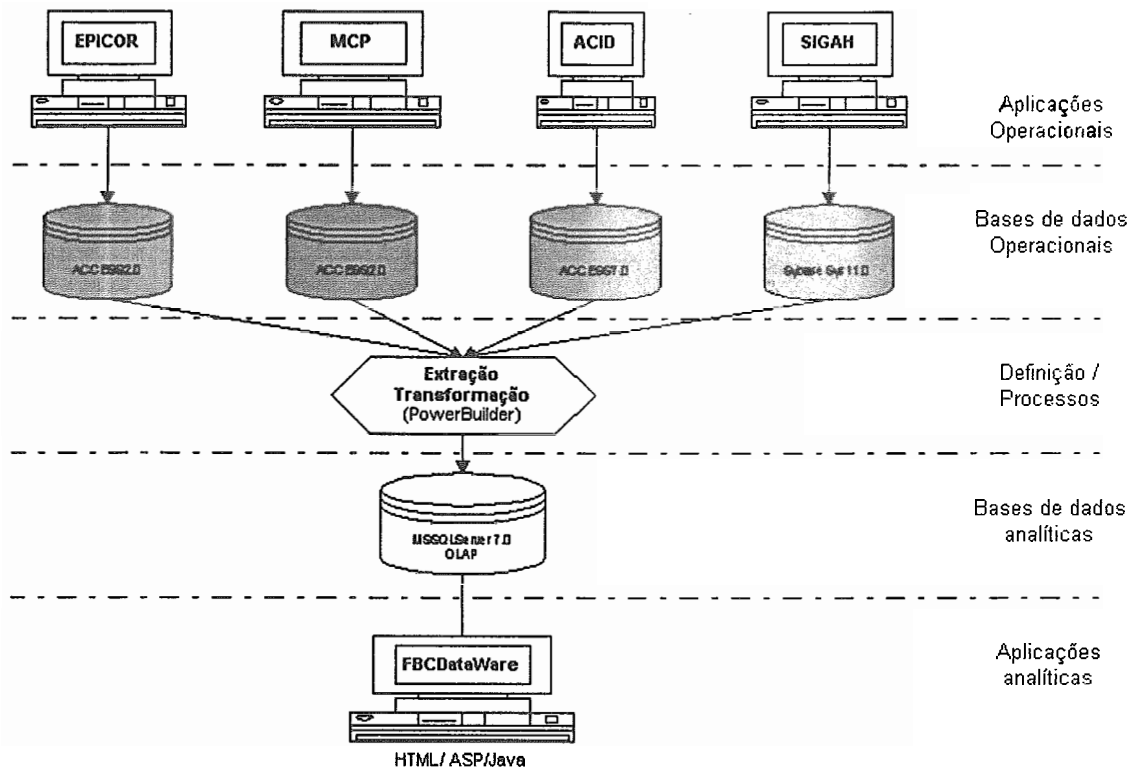


Figura 5-3 - Arquitetura do FBCDataWare

A escolha das ferramentas utilizadas foi feita por conveniência, de modo a aproveitar o que havia disponível (registro) na instituição. A existência, na UCCV/FBC, de uma rede intranet facilitou a definição da aplicação de análise e a implantação da mesma.

Na primeira versão foram utilizadas as quatro bases de dados escolhidas na atividade de avaliação das fontes de dados. Estas bases atingem quatro grupos distintos, os quais terão seus respectivos perfis de acesso no FBCDataWare.

Os processos de extração e transformação foram manuais através de rotinas de importação de dados e *scripts* de transformação direcionados à carga dos fatos e dimensões. Da mesma forma a definição e processamento de cubos na ferramenta OLAP.

Detalhes da aplicação serão mostrados na próxima seção.

5.4.2. O FBCDataWare

Uma característica importante desta aplicação, é a definição de perfis de acesso para os grupos de negócio. Estes perfis delimitam os grupos de dados que podem ser manipulados por cada grupo de negócio em seus respectivos níveis de análise, conforme determinados durante a análise de requisitos.

Os perfis são organizados de modo que cada usuário faz parte de pelo menos um grupo e cada grupo trabalha com pelo menos um nível de análise. Cada nível possui pelo menos um tema definido (cubo) (Figura 5-4).

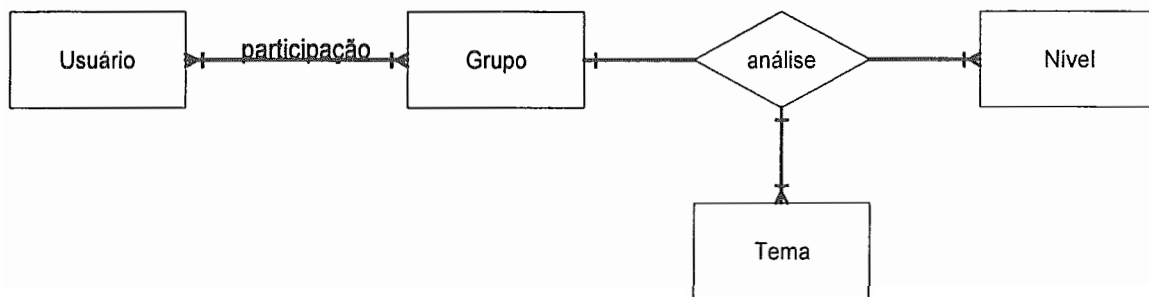


Figura 5-4- Modelo para a definição de perfis

Na entrada do sistema o usuário escolhe o grupo do qual faz parte. Um usuário pode fazer parte de mais de um grupo, visto que o grupo público FBC abriga todos os especialistas da instituição e estes podem, também, fazer parte de grupos particulares (Figura 5-5).

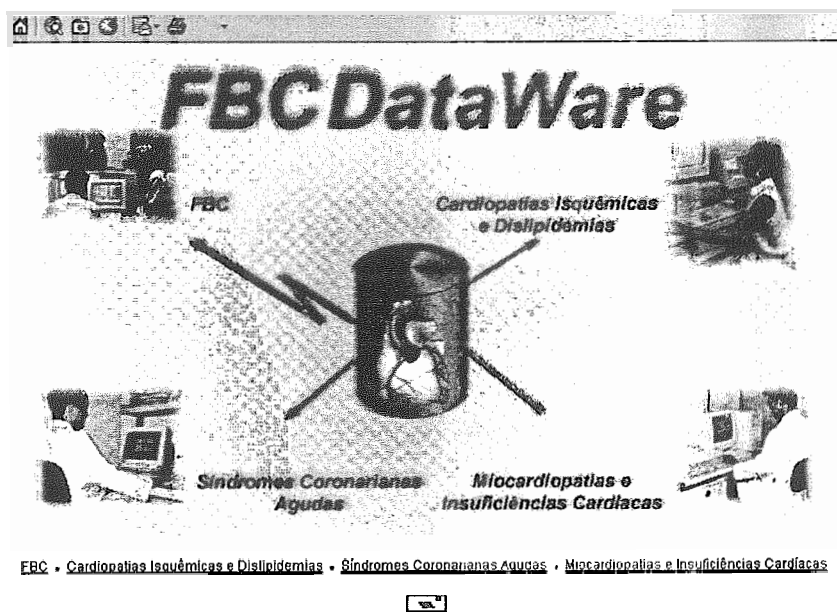


Figura 5-5 - FBCDataWare - Tela inicial - Acesso aos grupos

Uma vez definido o grupo, o usuário deve passar por uma autenticação neste grupo através de login e senha previamente cadastrados (Figura 5-6).

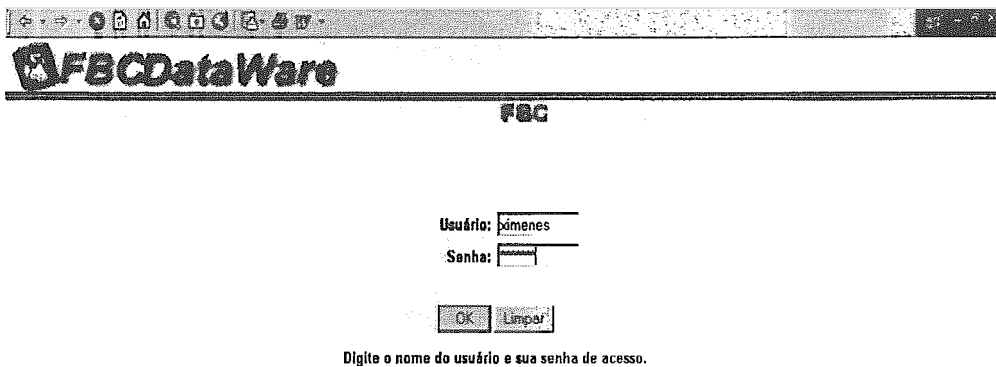


Figura 5-6 - FBCDataWare - Tela de Login - Autenticação no grupo

Feito isto, o usuário estará numa área de acesso restrito ao grupo que foi escolhido. A tela inicial do grupo está sendo usada para a publicação de alguns índices de avaliação da instituição (Figura 5-7). Estes índices não fazem parte necessariamente do modelo, mas servem para informar imediatamente ao grupo a situação de alguns fatores, até que o usuário defina em que perfil irá trabalhar.

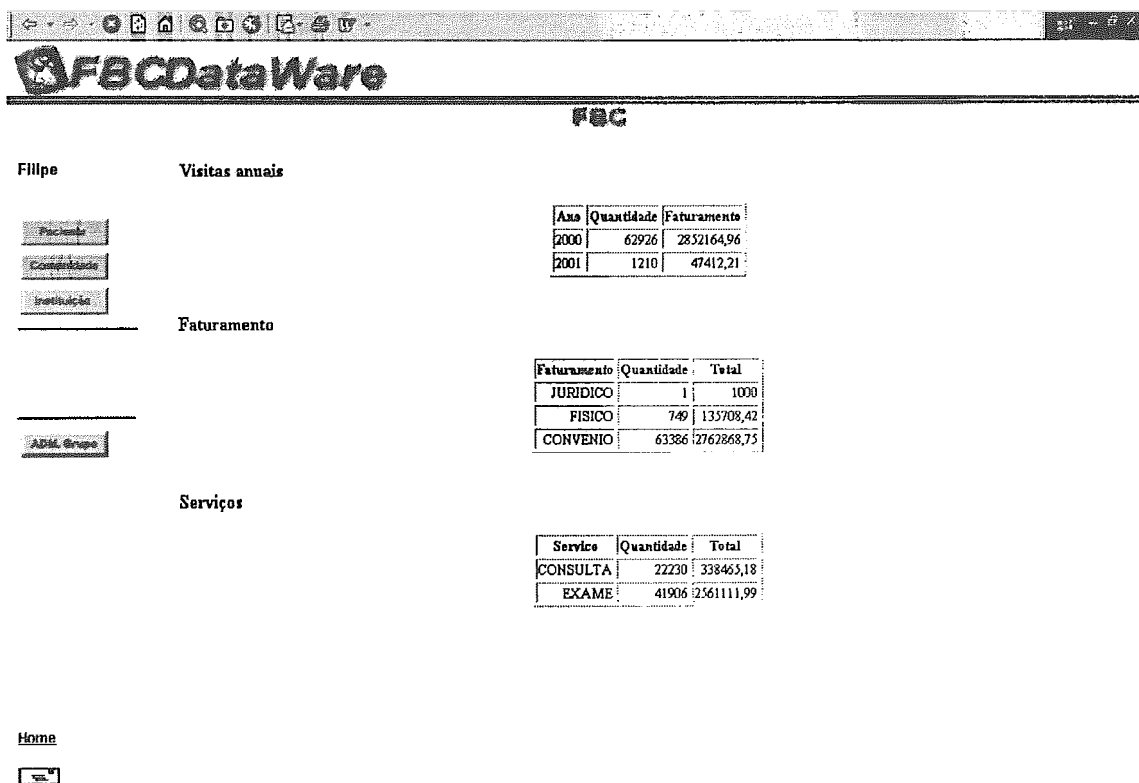


Figura 5-7 - FBCDataWare - Tela inicial do grupo

Na área do perfil escolhido, podem ser utilizadas consultas pré-definidas pelo próprio usuário ou que foram definidas por outros integrantes do grupo e publicadas para que todos pudessem utilizá-la (Figura 5-8).

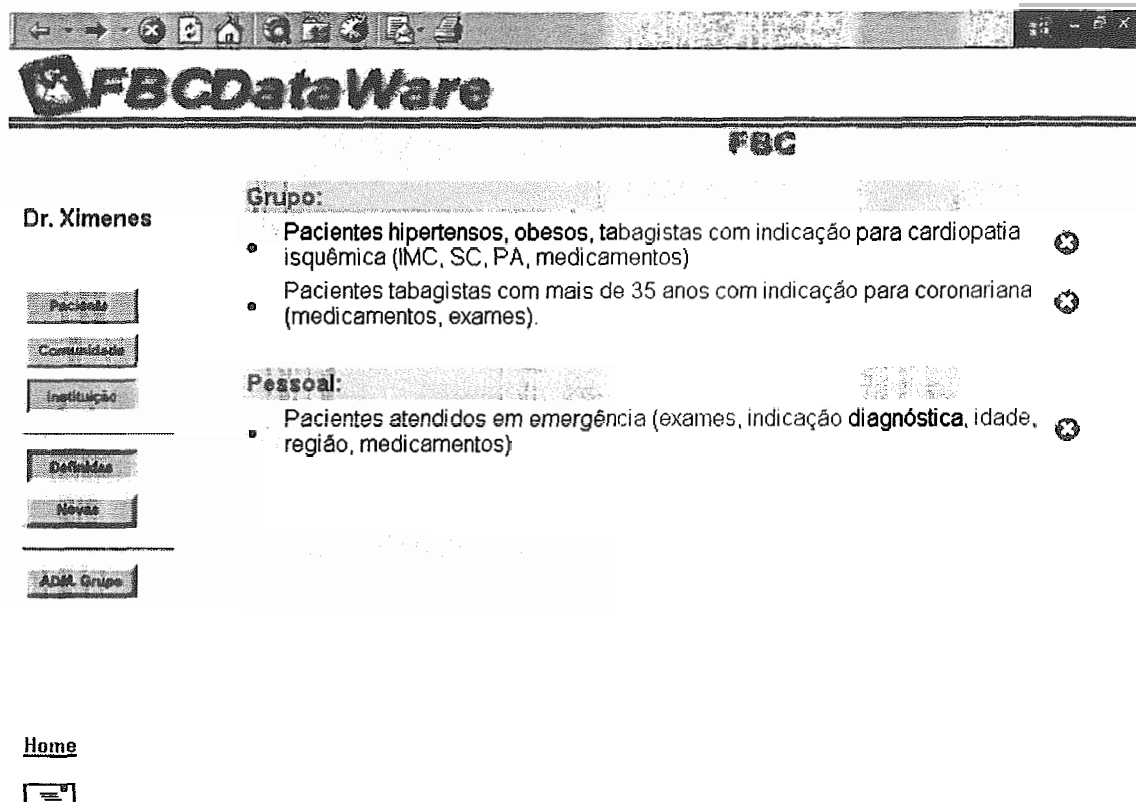


Figura 5-8 – FBCDataWare - Apresentação consultas definidas para o grupo

Caso o usuário ou o grupo não possua consultas pré-definidas, o sistema disponibiliza uma interface para que o usuário possa trabalhar definindo suas próprias consultas (Figura 5-9).

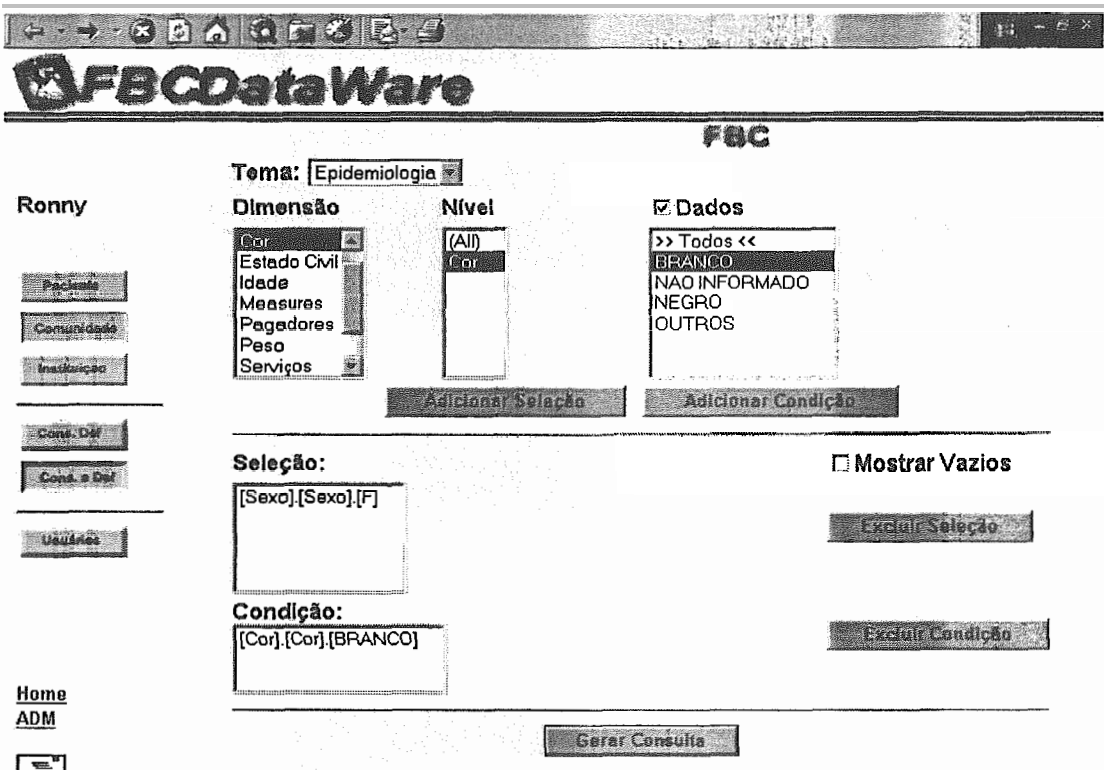



Figura 5-9 - FBCDataWare - Tela de Definição de consultas

Após a elaboração da consulta ou execução de alguma já pré-definida, os dados são apresentados em forma de tabela, e podem ser impressos e efetivamente utilizados (Figura 5-10). Algumas consultas podem ser apresentadas em forma de gráfico (Figura 5-11).

FBCDataWare - Consulta - Microsoft Internet Explorer

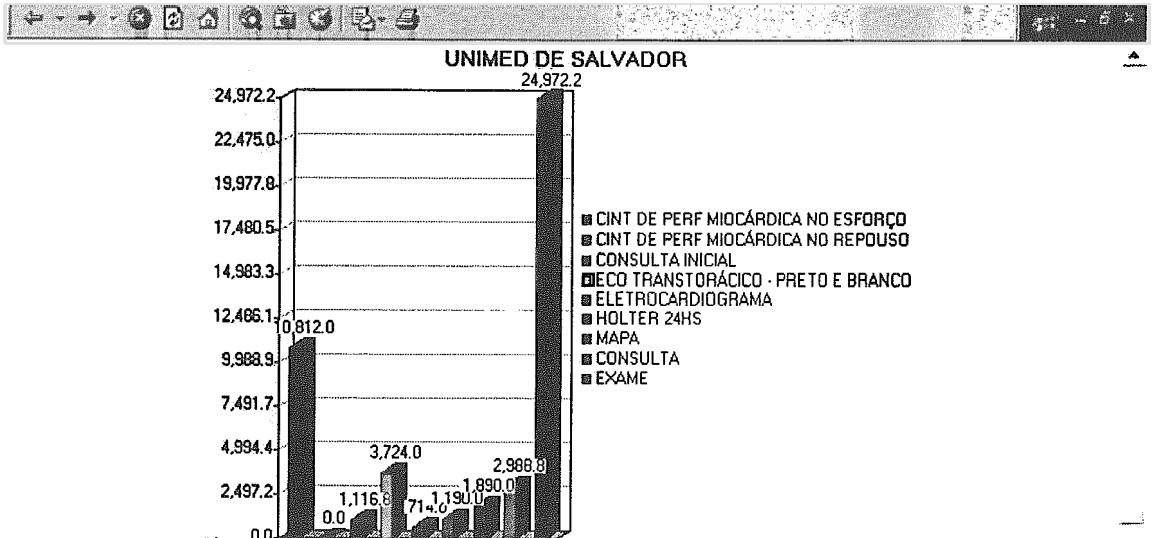
Consulta: Sem Título

Condições:
NÃO especificada.



	Todos os Serviços	CAT CAM CARD D/E C/EST CINECORONARIOGRAF	CAT CAM CARD D/E E CINEANGIOGRAFIA	CAT CAM ESQ COM EST CINECORON	CAT D/E C/ CINEANGIO E CORONARIOGRAFIA	HOLTER 24HS	RAIO X DE TÓRAX	CONSULTA	EXAME
SAÚDE BAMERINDUS	386,40							112,00	274,40
SAÚDE BRADESCO	39.366,47				560,00	630,00		3.630,12	35.736,35
SULAMÉRICA SERVIÇOS MÉDICOS LTDA	136.983,69				3.780,00	3.442,50	11,34	17.831,34	119.152,35
TELEMAR	368,06							120,00	248,06
UNIBANCO	1.992,53	270,00		179,55				200,00	1.792,53
CONVENIO	958.737,63	7.720,00	900,00	18.458,86	7.112,00	34.404,32	365,60	122.737,91	835.999,72
FÍSICO	5.211,00			2.800,00				236,00	4.975,00
PACIENTE	64.234,42			3.800,00		14.730,00	15.000,00	4.416,00	59.818,42

Figura 5-10 - FBCDataWare - Apresentação do resultados da consulta em tabela



Geral

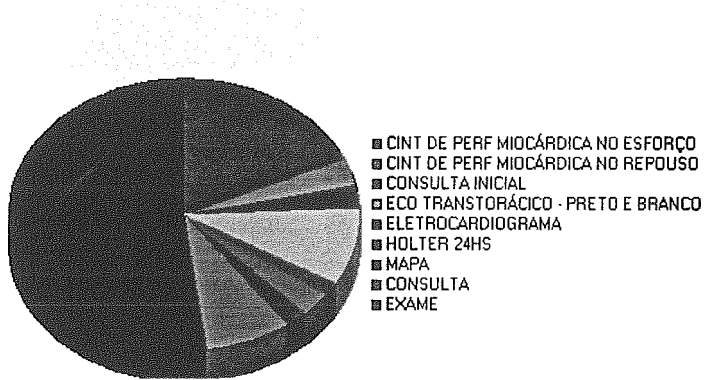


Figura 5-11 - FBCDataWare - Apresentação do resultados da consulta em gráfico

Algumas funcionalidades de gerência dos grupos também foram implementadas, e, a princípio, disponibilizadas para que um usuário do grupo possa alterar sua senha ou incluir mais integrantes (Figura 5-12).

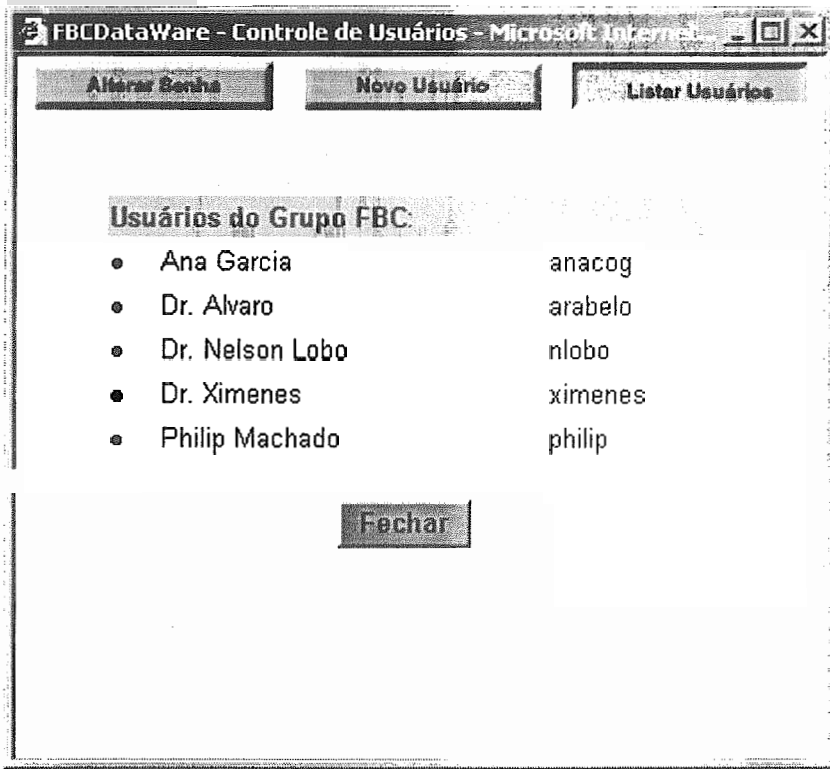


Figura 5-12 - FBCDataWare - Controle de usuários do grupo

Outras funcionalidades mais sofisticadas e que envolvem definição dos perfis, estão disponíveis apenas para usuários administradores que, provavelmente, não serão médicos (Figura 5-13).

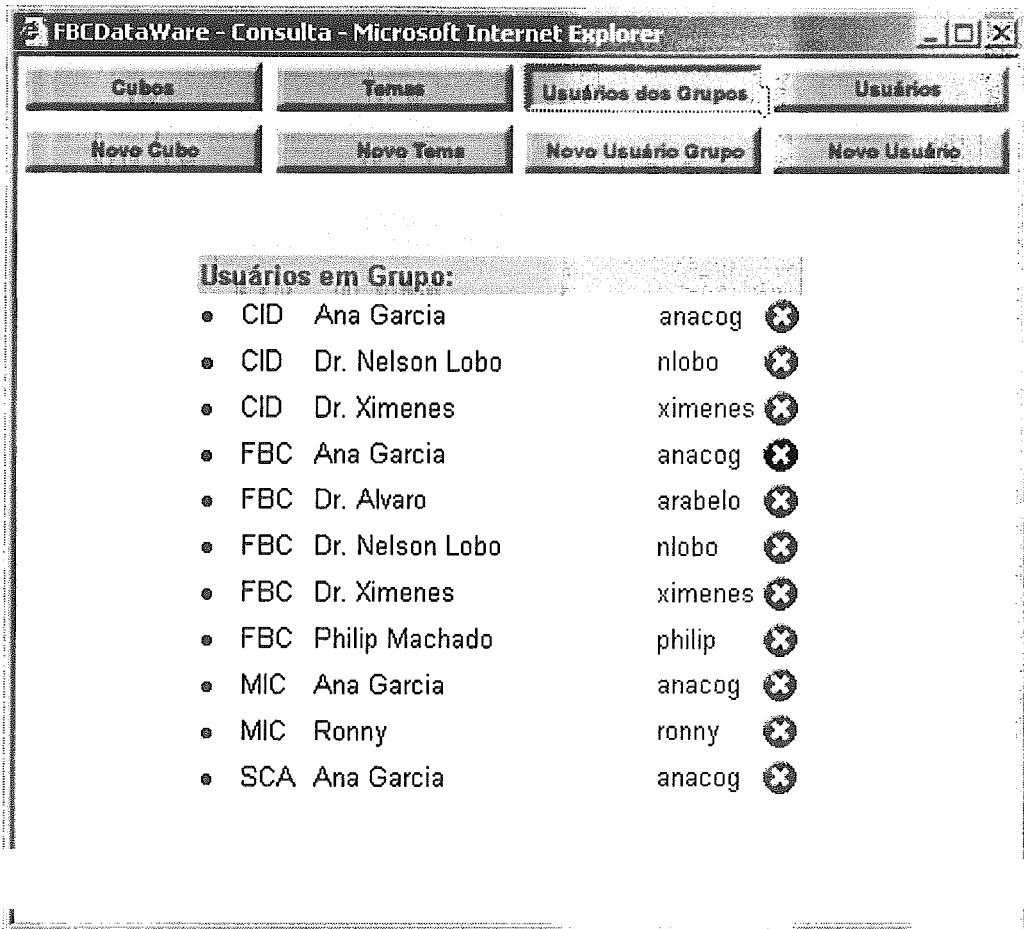


Figura 5-13 - FBCDataWare - Administração de Grupos e Perfis

E importante ressaltar que ao final da fase de organização, o produto gerado era muito simples e preparado para melhoramentos em todos os sentidos, sendo que o principal objetivo era apoiar o andamento da fase de levantamento de requisitos. Era de grande importância para aproximar o usuário do que poderia ser um produto final, de modo que já tratava dados reais que poderiam ser efetivamente usados, mesmo que de maneira simples.

5.5. Considerações Finais

A experimentação do processo foi importante no sentido de adequá-las ao domínio médico. Os usuários dessa área possuem características particulares dadas pela sua forma de raciocínio, característica do tipo de atividade que desempenha. Sendo assim, pelo ambiente de *data warehouse* não ser um tipo de aplicação ao qual eles estão acostumados a lidar, o desenvolvimento da fase de análise de requisitos deveria acontecer da maneira mais ágil e produtiva possível, visto que para estes

profissionais não é comum a regularidade de tempo disponível para as atividades extras. Com a mesma agilidade deveriam se tornar claros os objetivos do ambiente.

O produto final já possuía características do processo de transformação das informações da instituição, mesmo ainda sem utilizar todos os recursos de uma ferramenta OLAP, como navegação hierárquica e quebras dinâmicas de grupos de dados. O primeiro ciclo da análise de requisitos tinha como objetivos principais a definição de grupos de negócios e caracterização do fluxo de informações para cada grupo, desta forma, o produto adquirido desta iteração auxiliava o esclarecimento das características e motivava melhorias com maior fundamento.

6. Conclusão

Este trabalho tratou de aspectos relacionados à organização de um ambiente para apoiar atividades de pesquisa e tomada de decisão, abordando os níveis de análise de dados que podem ser encontrados numa instituição de saúde. Este ambiente tem características de *data warehouse* e repositório de dados clínicos, pelas particularidades da área de atuação.

A pesquisa foi direcionada para a exploração de formas de desenvolvimento e organização de informações em um *data warehouse*. Foi feito um estudo sobre o estado da arte desta tecnologia e dos diversos sistemas de informação encontrados na área médica. Outro aspecto avaliado foi o processamento de informações de uma organização e o desenvolvimento de atividades destinadas a contribuir com algum valor para o negócio em questão. Este aspecto auxiliou a identificação e o mapeamento do processamento de informações de uma instituição de saúde. Na literatura existem poucos trabalhos voltados para a organização de *data warehouse* clínico baseados nos níveis de análise da informação que podem ser constatados na instituição.

Uma contribuição relevante deste trabalho é a organização de uma seqüência de atividades para a análise de requisitos de um *data warehouse* baseado nos níveis analíticos de uma instituição. O detalhamento destas atividades foi possível baseado no estudo das atribuições necessárias para a construção de um *data warehouse* e nos requisitos de ambientes desta categoria. Este processo foi experimentado e possibilitou a construção de um *data warehouse* para cardiologia na UCCV/FBC.

A experimentação do processo de organização proposto foi de grande importância para validação de todas as etapas definidas. A instituição possui fortes elementos para a realização de um projeto deste tipo e soube perceber sua importância e contribuições incentivando o seu desenvolvimento. Foi disponibilizada uma primeira versão do ambiente que ainda necessita de melhoramentos, principalmente por se ter cumprido apenas a

primeira fase do desenvolvimento. Ainda assim, pode ser utilizada para auxiliar o refinamento dos requisitos.

A concretização deste trabalho gerou elementos essenciais para outros projetos. Sendo assim, muito ainda pode ser feito, principalmente no que se refere a:

- Aprofundar as técnicas de refinamento das funcionalidades do ambiente;
- Caracterizar os níveis de análise em termos de adequação dos dados, facilitando uma possível automatização do enquadramento dos requisitos;
- Definir atividades de manutenção do ambiente de acordo com os níveis de análise definidos, para evitar uma descaraterização após sucessivas evoluções;
- Estudar o impacto da utilização do ambiente para a instituição;
- Estender o ambiente a outros grupos da instituição e avaliar os aspectos desta ampliação.

A experiência no desenvolvimento deste trabalho mostrou que à medida que os usuários vão conhecendo a tecnologia que está sendo proposta mais idéias e interesses vão surgindo, aumentando assim as perspectivas de trabalhos e pesquisas que podem ser desenvolvidas.

Referências

- AGOSTA, L., 1999, **The Essential Guide to Data Warehousing**, 1ª ed, Chicago, IL, Prentice Hall.
- AGRAWAL, R., Gupta, A., Sarawagi, S., 1997, **Modeling Multidimensional Databases**, Research Report, IBM Almaden Research Center, San Jose, USA,
- ANDERSON, J.G., 1998, "Clearing the Way for Physicians' Use of Clinical Information Systems", **Communications of ACM**, v. 40 n. 8 (Aug), pp. 83-90
- ASARO, P. V., Herting Jr, R. L., Roth, A. C., et al., 1999, "Effective Audit Trails – A Taxonomy For Determination of Information Requirements", In: **Proceedings of American Medical Informatics Association Annual Symposium**, pp. 663 - 665, Washington, DC - USA, November
- BARROWS, R. C., Clayton, P. D., 1996, "Privacy, Confidentiality, and Electronic Medical Records", **Journal of American Medical Informatics Association**, v. 3 n. 02 (mar/apr), pp. 139 – 148
- BERNDT, D. J., Hevner, A. R., 1998, "CATCH/IT:A Data Warehouse to Support Comprehensive Assessment for Tracking Community Health", In: **Proceedings of the American Medical Informatics Association Annual Symposium**, pp. 555-560, Orlando - USA, November
- BOEHM, B., Egyed, A., Kwan, J. et al., 1998, "Using the WinWin Spiral Model: A Case Study", **IEEE Computer**, v. 31, n.07 (Jul), pp. 33-44.
- BOHN, K., 1997, "Converting Data for Warehouses", **DBMS Magazine**, (url: <http://www.dbmsmag.com/9706d15.html>) v. 10, n. 06 (Jun), on line

- BONTEMPO, C., Zagelow, G., 1998, "The IBM Data Warehouse Architecture", **Communications of ACM**, v. 41, n. 9 (Sep), pp. 38-51.
- CALVANESE, D., Giacomo, G., Lenzerini, M. *et al*, 1998, "Source Integration in Data Warehousing", In: **Proceedings of the 9th Int. Workshop on Database and Expert Systems Applications (DEXA-98)**, IEEE Computer Society Press, pp. 192-197, Vienna – Austria, August
- CHU, W. W., Leong, I. T., Taira, R. K., 1994, "A Semantic Modeling Approach for Image Retrieval by Content.", **VLDB Journal**, v. 3, n.04 (Apr), pp. 445-477.
- CODD, E. F., Codd, S. B., and Salley, C. T., 1994, Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate, **warehouse.chime-net.org/software/datastore/dataaware/coddc0.html**.
- DAPHNE, 1998, "Building Corporate Business Intelligence: From Datamarts to the Enterprise Data Warehouse", url: www.daphnesoft.com/daphnewhitepaper.html.
- DE MARSICOI, M., Cinque, L., Levialdi, S., 1996, "Indexing pictorial documents by their content: a survey of current techniques", **Image and Vision Computing**, v. 15, n. 02, (May), pp. 119-141.
- DOYLE, J. D., 1997, "Informed Decision Making Through Data Warehousing", url: dhrinfo.hr.state.or.us/intranet/tands/Dwpap/DWWHITEP.htm
- FIRESTONE, J. M., 1997, **Evaluating OLAP Alternatives**, Report 4, Executive Information Systems, Inc., March
- FLUM, J., 1995, "Query by Image and Video Content: The QBIC System", **IEEE Computer**, v. 28, n. 9 (Jan), pp. 23-30.
- GAMA, C., 1996, **Sistema Cooperativo Hipermedia de Apoio à Sessão Clínica**, Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

- GARDNER, S. R. 1998, "Building the Data Warehousing", **Communications of the ACM**, v. 41, n. 9 (Sep), pp. 52-59
- GEISSBUNHLER, A., 1998, "Clinical Information Systems – What is the Bottom Line? ", **Journal of the American Medical Informatics Association**, v. 5 n. 6 (Dec), pp. 585-586
- GILL, H. S., Rao, P. C., 1996, **The Official Client/ Server Computing Guide to Data Warehousing**, USA, Que Corporation.
- INMON, W. H., 1997, **Como Construir o Data Warehouse**, 2nd ed., Editora Campus.
- JAJODIA, S., Ammann, P., McCollum, C. D., 1999, "Surviving Information Warfare Attacks", **IEEE Computer**, v. 32, n. 4 (Apr), pp. 57-63.
- KAHN, M. G., 1998, "Three perspectives on Integrated Clinical Databases", **Yearbook of Medical Informatics 1998**, pp. 289 – 294
- KELSO, S. W., Myers, J. E., 1998, "Integrating Medication Management into the Electronic Patient Record", Metadata Company, url: www.meta-data.com/sk_tep6.html
- KESHAVJEE, K., Willison, D.J., Holbrook, A.M., Kyba, R., 1999, "Communicating Privacy and Confidentiality Issues to Physicians and Patients in a Research Study Using Electronic Medical Records", In: **Proceedings of American Medical Informatics Association Annual Symposium**, pp.1098, Washington, DC – USA, November
- KIMBALL, R., 1996, **The Data Warehouse Toolkit**, USA, John Wiley & Sons, Inc.
- KIMBALL, R., 1996a, "Letting the Users Sleep, Part 1. Nine Decisions in the Design of a Data Warehouse", **DBMS Magazine**, (url: <http://www.dbmsmag.com/9612d05.html>) v. 9, n. 12 (Dec), (on line)
- KIMBALL, R., 1996b, "Slowly Changing Dimensions", **DBMS Magazine**, v. 9, n. 4 (Apr), (on line)

- KIMBALL, R., 1997a, "A Dimensional Modeling Manifesto", **DBMS Magazine**, (url: <http://www.dbmsmag.com/9708d15.html>) v. 10, n. 08 (Aug) , (on line)
- KIMBALL, R., 1997b, "Letting the Users Sleep, Part 2. Nine Decisions in the Design of a Data Warehouse", **DBMS Magazine**, (url: <http://www.dbmsmag.com/9701d05.html>) v. 10, n. 01 (Jan), (on line)
- KIMBALL, R., 1998a, "Brave New Requirements for Data Warehousing", **Intelligent Enterprise Magazine - Warehouse Architect**, (url: http://www.intelligententerprise.com/db_area/archives/1998/9810/warehouse.shtml) v. 1, n. 01 (Oct), (on line)
- KIMBALL, R., Reeves, L., Ross, M., Thornthwaite, W., 1998b, **The Data Warehouse Life Cycle Toolkit**, USA, John Wiley & Sons, Inc.
- LIMA, K., 1999, **Uma Aplicação de Telemedicina para Cardiologia**, Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- MARCO, D., 1998, "Managing Meta Data", **DM Review (on line)**, url: <http://www.dmreview.com/master.cfm?NavID=55&EdID=699> March, (on line)
- MATTISON, R., 1997, **Data Warehousing and Data Mining for Telecommunications**, MI, Artech House.
- NADKARNI, P. M., Brandt, C., 1998, "Data Extraction and Ad Hoc Query of an Entity – Attribute – Value Database", **Journal of the American Medical Informatics Association**, v. 5 n. 6 (Dec), pp 511-527.
- OLIVEIRA, K., 1995, **Avaliação da Qualidade de Sistemas Especialistas**, Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- ORR, K., 1996: "Data Warehouse Technology", url: <http://www.kenorrinst.com/pg%2033%20d.w.%20whitepaper.htm>, revised edition: 2000

- PEDERSEN, T. B., Jensen, C. S., 1998, "Clinical Data Warehousing - a Survey" In **Proceedings of the VIII Mediterranean Conference on Medical and Biological Engineering and Computing (Medicon98)**, pp. 150-160, Limassol - Cyprus, June.
- PEDERSEN, T. B., Jensen, C. S., 1998, "Research Issues in Clinical Data Warehousing" In: **Proceedings of the Tenth International Conference on Scientific and Statistical Database Management (SSDBM98)**, pp. Capri, Italy, July.
- PENDSE, N., 2000, "The origins of today's OLAP products", url: www.olapreport.com/origins.htm.
- PERKINS, A., 1996, "Developing a Data Warehouse The Enterprise Engineering Approach", Visible Systems Corporation, url: www.ozemail.com.au/~visible/papers/dw.htm.
- RAGHUPATHI, W., 1998, "Health Care Information Systems", **Communications of ACM**, v. 40 n. 8 (Aug), pp. 81-82
- SALTON, G., 1995, **Automatic Text Processing**, Reading, MA, Addison Wesley Publishing.
- SCHNEIDER, G., Winters, J. P., 1998, **Applying Use Cases**, CA, Addison-Wesley Publishing.
- SCULLY, K. W., Pates, R. D., Desper, G. S. et al., 1997, "Development of an Enterprise-Wide Clinical Data Repository: Merging Multiple Legacy Databases", In: **Proceedings of the American Medical Informatics Association Annual Fall Symposium**, pp. 32-36, Nashville-USA, October
- SEN, A., Jacob, V. S., 1998, "Industrial-Strength Data Warehousing", **Communications of ACM**, v. 41, n. 9 (Aug), pp. 29-31.
- STEAD, W. , 1998, "Medical Informatics – On the Path Toward Universal Truths", **Journal of the American Medical Informatics Association**, v. 5 n. 6 (Dec), pp. 583-584

- TEICH, J. M., 1998, "Clinical Information Systems for Integrated Helthcare Networks", In: **Proceedings of American Medical Informatics Association Annual Symposium**, pp. 19-28, Orlando-USA, November
- VALLE, C., 1997, **Sistemas de Acesso Público para Educação de Pacientes**, Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- van MULLIGEN, E. M., Satm, H., van Ginneken, A. M., 1998, "Clinical Data Entry", In: **Proceedings of American Medical Informatics Association Annual Symposium**, pp. 81-85, Orlando-USA, November
- WAND, Y., Wang, R. Y., 1996, "Anchoring Data Quality Dimensions in Ontological Foundations", **Communications of the ACM**, v. 39, n. 11 (Nov), pp. 32-37
- WATSON, H. J., Haley, B. J., 1998, "Managerial Considerations", **Communications of the ACM**, v. 41, n. 9 (Sep), pp. 32-37
- WEINTRAUB, W. S., Mckay, C. R., Riner, R. N., 1997, "The American College Cardiology National Database: Progress and Challenges", **Journal of American College Cardiology**, v. 29, n. 2 (Feb), pp. 459-466
- WELDON, J. L., 1997, "Data Warehouse Building Blocks", **BYTE**, v. 22, n. 1 (Jan), pp. 82-102
- YOUNESS, S., 2000, **Professional Data Warehousing with SQL Server 7.0 and OLAP Services**, New York, NY, WROX Press

Anexo 1

Anexo 1. Negociação das informações

Tabela 1 - Definição de grupos e entidades de interesse

	Visitas	Paciente	Eletrocardiograma	Teste Ergométrico	Holter	ECO Doppler	ECO STRESS	ECO Transesofágico	ECO p/ Cardiopatias Congênicas	Cintilografia (Ventriculografia)	Cintilografia de Perfusão miocárdica repouso	Cintilografia de Perfusão miocárdica esforço	Hemodinâmica	Dislipidemias
Arritmias														
Cardiopatias Isquêmicas		1	2			3					4	4	5	6
Dislipidemias														
Hemodinâmica														
Hipertensão														
Insuficiência Cardíaca e Miocardiopatias		1	2		3	4	5							6
Síndrome Coronariana Aguda		1	2	3		4	5			6	7	7	8	9

Síndrome Coronariana Aguda

1. Pacientes – Pacientes com cardiopatia tipo infarto / angina (EPICOR),
Pacientes com cardiopatia do tipo ICC isquêmica (Miocardiopatias),
Pacientes com dislipidemias (ACID)
2. ECG – Pacientes que apresentaram alterações de condução do tipo BLRD3, alterações isquêmicas e zona eletricamente inativa
3. Teste Ergométrico – Conclusão da ergometria
4. ECO Doppler– Pacientes com cardiopatia isquêmica, grau da cardiomiopatia dilatada, com disfunção sistólica do ventrículo direito, condição do segmento (normal = acinético; anormal = (hipocinético/ discinético), fração de ejeção
5. ECO Stress – Conclusão do exame:
 - Normal = Negativo
 - Inconclusivo
 - Anormal = positivo
6. Cintilografia (ventriculografia) – Fração de ejeção e conclusão
7. Cintilografia (Perfusão miocárdica) – Conclusão Normal/ Alterada e Observação sobre a Função Sistólica do VE. Ver observação no final
8. Hemodinâmica – Situação das coronárias quanto à lesão, o segmento e o tipo da lesão
9. Dislipidemias – Para pacientes do EPICOR
 - Data do Último perfil lipídico
 - Medicação em uso: Estatina
 - Fibratos
 - Sequestrantes do ácido biliar
 - Probucol

 - Col Total: (mg/dl)
 - HDL-C: (mg/dl)
 - LDL-C : (mg/dl)
 - Triglicérides : (mg/dl)
 - VLDL-C: (mg/dl)

Tipo de dislipidemia:

Hipercolesterolemia / Hipertrigliceridemia / Hiperlipidemia Mista / HDL-C Baixo (↓ do HDL-c) / não classificada

Tipo de medicação em uso: Nenhuma /

Estatina:

sinvastatina / lovastatina / fluvastatina / atorvastatina /

pravastatina / Cerivastatina / outras

Dose: (mg/dia)

Fibratos

bezafibrato / ciprofibrato / gemfibrozil / fenofibrato / outros

Dose: (mg/dia)

Sequestrantes do ácido biliar: (colestiramina)

Dose: (g/dia)

Probucol: Dose: (mg/dia)

Outras medicações que interferem no perfil lipídico

Diuréticos / Beta bloqueadores / Anticoncepcionais / Corticosteróides / Anabolizantes

Insuficiência Cardíaca e Miocardiopatias

1. **Pacientes –** Pacientes com cardiopatia do tipo ICC (Miocardiopatias),
Pacientes da UCO com ICC e etiologia da ICC (EPICOR),
Pacientes com ICC e dislipidemia (ACID)
Pacientes com hipertensão (Hipertensão)
2. **ECG – Dados dos exames dos pacientes**
 - Ritmo
 - Bloqueios
 - Sobrecargas
 - Arritmias
3. **Holter – Dados dos exames dos pacientes**
 - Frequência Cardíaca Média
 - Ritmo
 - BAV
 - Arritmia

Número Absoluto de EV

4. **ECO Doppler - Dados dos exames dos pacientes**

5. **ECO Stress - Dados dos exames dos pacientes**

AE

Distância E-Septo

Dados do VE

Morfologia do VE

Comprometimento VD

Comprometimento VE

PSVD

Alteração Segmentar VE

Insuficiência

Trombos

Função Diastólica

Cardiopatias Isquêmicas e Dislipidemias

1. **Pacientes – Pacientes com dor torácica (SIGAH, EPICOR)**

Pacientes com/ sem eventos coronariano (ACID)

Pacientes com IAM/ Angina instável (EPICOR)

2. **ECG – alterações isquêmicas**

3. **ECO – alterações de contratilidade**

4. **Cintilografia – defeitos de perfusão**

5. **Hemodinâmica – obstrução coronariana**

6. **Dislipidemias – Para pacientes com/sem evento coronariano**

medicamentos usados,

sintomatologia

hipercolesterolemia

hipertrigliceridemia

diabetes

.idade

sexo

Anexo 2

Anexo 2. Especificação dos Requisitos

Exemplo de modelagem de fato

Nível de análise:	Instituição
Grupo:	FBC
Fato:	Visitas na instituição: tf_visitas
Dimensões:	Serviços prestados pela instituição: td_servicos Prestadores de serviços para a instituição: td_prestadores Datas de visitas: td_tempo Pagadores dos serviços (convênios, pessoa jurídica, particular): td_pagadores Setores da instituição: td_setores
Fonte de dados	SIGAH

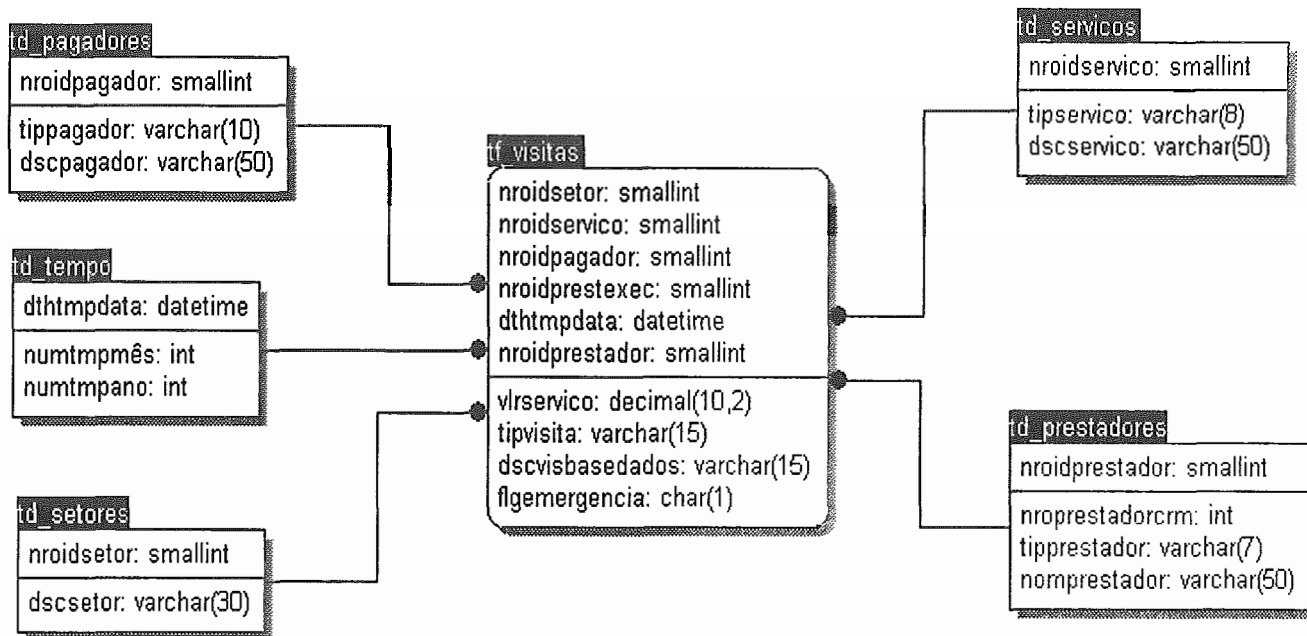
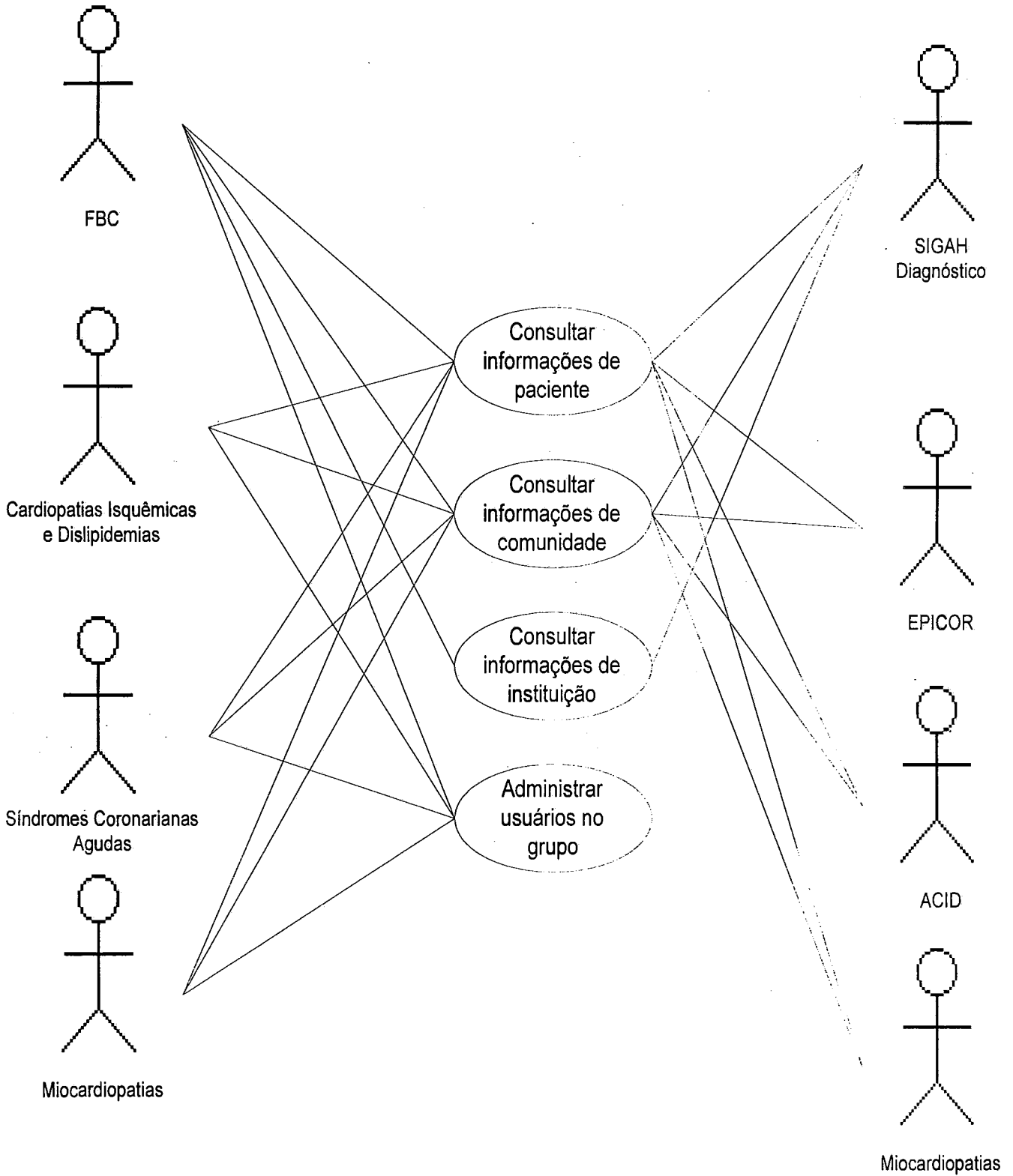


Figura 1 - Modelo do fato Visitas do grupo FBC

Anexo 2. Especificação dos Requisitos



Anexo 2. Especificação dos Requisitos

Atores

FBC: grupo de especialistas de cardiologia e administradores da FBC/UCCV.

Síndromes Coronarianas Agudas: grupo de pesquisa da FBC/UCCV para o estudo de pacientes infarto do miocárdio e angina da UCO

Cardiopatias Isquêmicas e Dislipidemias: grupo de pesquisa da FBC/UCCV para o estudo de pacientes do ambulatório de com cardiopatias isquêmicas e dislipidemias

Miocardopatias: grupo de pesquisa da FBC/UCCV para o estudo de pacientes com miocardopatias

Casos de Uso

- Consultar informações de paciente.

Atores:

FBC; Síndromes Coronarianas Agudas; Cardiopatias Isquêmicas e Dislipidemias;
Miocardopatias

Fluxo de eventos

Fluxo principal

1. O caso de uso começa quando o usuário escolhe um grupo de acesso.
2. O sistema pede nome e senha para validação do usuário no grupo escolhido
3. O sistema disponibiliza a área de trabalho do grupo
4. O usuário escolhe o perfil Paciente, para consultas específicas de paciente
5. O usuário escolhe área de definição de consultas
6. O sistema disponibiliza temas (cubos) específico para paciente
7. O usuário escolhe as dimensões, os níveis das dimensões e/ou dados específicos que deseja selecionar
8. O usuário escolhe os dados condicionais para a consulta
9. O sistema apresenta os dados disponíveis segundo os critérios informados

Fluxos alternativos

- O usuário pode desistir da consulta em qualquer ponto do caso de uso, voltando para o passo de escolha do grupo de acesso (1)
- O usuário pode decidir entrar como outro grupo antes informar nome e senha (2) e voltar para o passo de escolha do grupo de acesso (1)
- O usuário pode não fazer parte do grupo escolhido, nome inválido para o grupo, e não passar para a área de trabalho do grupo (3)
- O usuário pode errar a senha de acesso e não passar para a área de trabalho do grupo (3)
- O usuário pode escolher a área de consultas definidas, alterando o caso de uso nos passos (5), (6), (7), (8), (9), para:
 5. O usuário escolhe a área de consultas definidas e publicadas para o grupo
 6. O usuário escolhe a consulta desejada
 7. O sistema apresenta os dados disponíveis segundo os critérios da consulta escolhida
- O usuário pode decidir em qualquer ponto deste caso de uso acessar outro perfil dentro da área de trabalho do grupo.

Requisitos

Para que este caso de uso seja realizado deve haver pelo menos um cubo para o perfil paciente definido para o grupo.

Anexo 2. Especificação dos Requisitos

- Consultar informações de comunidade.

Atores:

FBC; Síndromes Coronarianas Agudas; Cardiopatias Isquêmicas e Dislipidemias;

Miocardopatias

Fluxo de eventos

Fluxo principal

1. O caso de uso começa quando o usuário escolhe um grupo de acesso.
2. O sistema pede nome e senha para validação do usuário no grupo escolhido
3. O sistema disponibiliza a área de trabalho do grupo
4. O usuário escolhe o perfil Comunidade, para consultas específicas de comunidade
5. O usuário escolhe área de definição de consultas
6. O sistema disponibiliza temas (cubos) específico para comunidade.
7. O usuário escolhe as dimensões, os níveis das dimensões e/ou dados específicos que deseja selecionar
8. O usuário escolhe os dados condicionais para a consulta
9. O sistema apresenta os dados disponíveis segundo os critérios informados

Fluxos alternativos

- O usuário pode desistir da consulta em qualquer ponto do caso de uso, voltando para o passo de escolha do grupo de acesso (1)
- O usuário pode decidir entrar como outro grupo antes informar nome e senha (2) e voltar para o passo de escolha do grupo de acesso (1)
- O usuário pode não fazer parte do grupo escolhido, nome inválido para o grupo, e não passar para a área de trabalho do grupo (3)
- O usuário pode errar a senha de acesso e não passar para a área de trabalho do grupo (3)
- O usuário pode escolher a área de consultas definidas, alterando o caso de uso nos passos (5), (6), (7), (8), (9), para:
 5. O usuário escolhe a área de consultas definidas e publicadas para o grupo
 6. O usuário escolhe a consulta desejada
 7. O sistema apresenta os dados disponíveis segundo os critérios da consulta escolhida
- O usuário pode decidir em qualquer ponto deste caso de uso acessar outro perfil dentro da área de trabalho do grupo.

Requisitos

Para que este caso de uso seja realizado deve haver pelo menos um cubo para o perfil comunidade definido para o grupo.