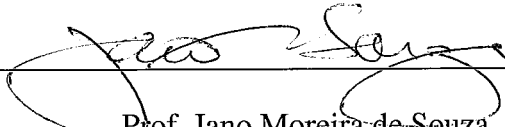


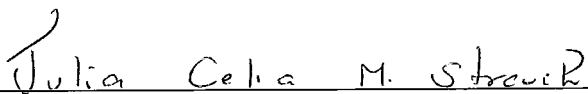
*BILL OF EXPERIMENTS: UM SISTEMA COLABORATIVO PARA
EXPLICITAÇÃO, REUSO E PLANEJAMENTO DE WORKFLOWS CIENTÍFICOS*

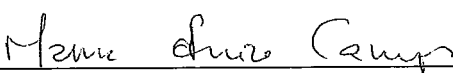
Leonardo Figueiredo Cardoso

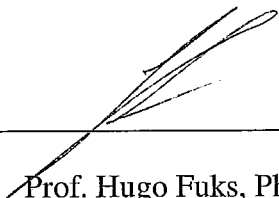
TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

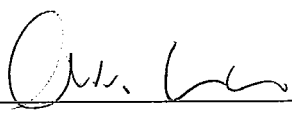
Aprovada por:


Prof. Jano Moreira de Souza, Ph.D.


Prof.^a. Júlia Célia Mercedes Strauch, D.Sc.


Prof.^a. Maria Luiza Machado Campos, Ph.D.


Prof. Hugo Fuks, Ph.D.


Dr. Alberto Sulaiman Sade Junior, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

JUNHO DE 2003

CARDOSO, LEONARDO FIGUEIREDO

Bill of Experiments: Um Sistema Colaborativo Para Explicitação, Reuso e Planejamento de Workflows Científicos [Rio de Janeiro] 2003

XII, 128p., 29,7 cm (COPPE/UFRJ, M.Sc., Engenharia de Sistemas e Computação, 2003)

Tese – Universidade Federal do Rio de Janeiro, COPPE

1. *Workflows* científicos
2. Bancos de Dados Científicos
3. Gestão do Conhecimento Científico

I. COPPE/UFRJ II. Título (série)

Aos meus amados pais Nelson e Luci.

Agradecimentos

A Deus, pela ajuda, pela saúde, pela vida.

Aos meus amados pais e irmão, sempre presentes, interessados e dando-me apoio incondicional para a conclusão deste trabalho.

Ao professor Jano Moreira de Souza, pelas idéias originais desta dissertação e por sua valiosa orientação e esforço para que esta fosse de qualidade.

À professora Júlia Strauch, pela incomensurável ajuda, especialmente na parte final, na preparação deste trabalho e por seus ensinamentos.

Ao doutor e amigo Alberto Sulaiman, por ter participado ativamente da preparação das raízes desta dissertação e por ser uma pessoa sempre otimista e generosa. “Sucesso a todos nós”.

À professora Maria Luiza Campos, por ter aceitado participar desta banca e por ter me mostrado os primeiros caminhos da pesquisa científica no projeto de fim de curso da graduação.

Ao professor Hugo Fuks, por ter aceitado participar desta banca.

À Samara, que soube compreender as minhas ausências devido à preparação deste trabalho.

Aos meus grandes amigos que estiveram juntos nesta caminhada, em especial os da COPPE: Leonardo Murta, Humberto José Vieira, Wanderson Araújo da Silva, Valdino de Azevedo Júnior e Victor Teixeira de Almeida.

À Patrícia Leal, pela valiosa amizade conquistada durante esses anos e por seu trabalho sério e competente como secretária da linha de banco de dados.

A todos os integrantes do Projeto SPeCS, que sempre contribuíram muito para a minha formação.

Ao projeto COPPETec, que me permitiu continuar em contato com o mercado de trabalho. Um abraço especial para Rosana Lopes de Brito, Leandro de Brito e Silva,

Marco Aurélio Ferreira Borges, Fernanda Jamus, Deborah Kalume e José Roberto Blaschek.

À CAPES e ao governo brasileiro, pelo apoio financeiro.

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

BILL OF EXPERIMENTS: UM SISTEMA COLABORATIVO PARA
EXPLICITAÇÃO, REUSO E PLANEJAMENTO DE WORKFLOWS CIENTÍFICOS

Leonardo Figueiredo Cardoso

Junho/2003

Orientadores: Jano Moreira de Souza
Júlia Célia Mercedes Strauch

Programa: Engenharia de Sistemas e Computação

Com a evolução das relações de trabalho ao longo das últimas décadas, os requisitos de gerenciamento também avançaram. Ao mesmo tempo em que mecanismos foram criados para apoiar ou automatizar a realização do trabalho, outros surgiram para o gerenciamento do seu fluxo. Quando se observa o trabalho de um cientista que realiza experimentos, compostos por uma série de atividades em seqüência, necessitando de dados iniciais e gerando resultados, surge a motivação para aplicação dos conceitos de *workflow* para o trabalho científico. Esta dissertação apresenta um sistema colaborativo para explicitação, reuso e planejamento de *workflows* científicos inerente aos experimentos. Este sistema é definido na ferramenta *Bill of Experiments* (BOE) e utiliza conceitos advindos de diversas áreas de pesquisa, tais como *workflow*, bancos de dados científicos, trabalho cooperativo auxiliado por computador (CSCW), gestão do conhecimento e inteligência artificial. O BOE é capaz de apontar melhores alternativas para a realização de experimentos científicos baseando-se em critérios de custo, prazo de obtenção e qualidade, em uma analogia com o processo de *Bill of Materials*. O sistema ainda age como um portal de *internet* que une pesquisadores, instituições e empresas. Apresenta-se um estudo de caso na área de agrometeorologia, sendo um zoneamento agroclimático o *workflow* científico em questão.

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

BILL OF EXPERIMENTS: A COLLABORATIVE SYSTEM FOR EXPLICITATION,
REUSE AND PLANNING OF SCIENTIFIC WORKFLOWS

Leonardo Figueiredo Cardoso

June/2003

Advisors: Jano Moreira de Souza
Júlia Célia Mercedes Strauch

Department: Computer and Systems Engineering

With the evolution of work relations through the last decades, the management requirements also evolved. At the same time that procedures were created to support or to automate the execution of the work, others appeared for the management of its flow. When a scientist's work is observed, in the execution of experiments, composed by series of activities in a flow, needing startup data and generating results, the motivation appears for application of workflow concepts to the scientific work. This thesis presents a collaborative system aiming at explicitation, reuse and planning of scientific workflows inherent to the experiments. We developed the Bill of Experiments (BOE) tool that uses concepts of several research areas, such as workflow, scientific databases, computer supported cooperative work (CSCW), knowledge management and artificial intelligence. The BOE tool is capable of pointing out the best alternatives for the accomplishment of scientific experiments based on cost criteria, deadline attainment and quality, in an analogy with the Bill of Materials process. The system acts as an internet portal that unites researchers, institutions and companies. This work presents a case study in the agrometeorology, with an agroclimatic zoning as the scientific workflow being analyzed.

Índice

Capítulo 1 - Introdução	1
1.1 Motivação	1
1.2 Objetivos	4
1.3 Contexto	5
1.4 Organização dos Capítulos	7
Capítulo 2 - Revisão da Literatura	8
2.1 Bancos de Dados Científicos	8
2.1.1 Definição	8
2.1.1.1 Gerência de Experimentos	11
2.1.1.2 Arquivamento e Recuperação de Dados	11
2.1.1.3 Recuperação de Literatura	12
2.1.1.4 Recuperação de Fatos Baseados na Literatura	12
2.1.1.5 Recuperação de Fatos Não Baseados na Literatura	13
2.1.2 Gerência de Experimentos Científicos	13
2.1.2.1 Ciclo de Vida de Experimentos	14
2.1.2.2 Trabalhos Relacionados	16
2.2 CSCW e Workflow	23
2.2.1 CSCW, Cooperação e Colaboração	23
2.2.2 Tecnologia <i>Workflow</i>	25
2.2.3 Classificação de <i>Workflows</i>	27
2.2.3.1 <i>Workflows ad-hoc</i>	27
2.2.3.2 <i>Workflows</i> Administrativos	27
2.2.3.3 <i>Workflows</i> de Produção	27
2.2.4 Funcionamento de <i>Workflows</i>	28
2.2.5 <i>Workflows</i> Científicos	29
2.2.6 Trabalhos Relacionados	32
2.3 Gestão do Conhecimento	34
2.3.1 Conhecimento	34
2.3.2 Tipos de Conhecimento	39
2.3.2.1 Conhecimento Popular	39
2.3.2.2 Conhecimento Religioso	39
2.3.2.3 Conhecimento Filosófico	39
2.3.2.4 Conhecimento Científico	40
2.3.2.5 Conhecimento Tácito e Explícito	41

2.3.3	O Processo de Gestão de Conhecimento	42
2.3.4	Definições de Gestão do Conhecimento	43
2.3.5	Ferramentas de Gestão do Conhecimento	49
2.3.6	Gestão do Conhecimento Científico	52
2.3.6.1	Trabalhos Relacionados	54
Capítulo 3 - Proposta da Ferramenta <i>Bill of Experiments</i>		59
3.1	Introdução	59
3.2	Analogia ao <i>Bill of Materials</i> (BOM)	60
3.3	Arquitetura da Ferramenta BOE	63
3.3.1	Portal do Pesquisador	65
3.3.2	Portal do Fornecedor	66
3.3.3	Projeto do <i>Workflow</i>	68
3.3.4	Módulo de Gerência de Qualidade	71
3.3.5	Módulo de Inferência	76
3.3.6	Módulo de Busca	81
3.3.7	Camada de Banco de Dados	82
3.3.8	Implementação da Arquitetura	83
Capítulo 4 - Estudo de Caso e Exemplos de Aplicação		87
4.1	Estudo de Caso na Agrometeorologia	87
4.2	Exemplo de Aplicação 1 – Problemas Inversos	92
4.3	Exemplo de Aplicação 2 – Área Biológica	95
4.3.1	Consultando <i>Workflows</i> Alternativos	95
4.3.2	Melhores Alternativas para a Citometria de Fluxo	96
Capítulo 5 - Conclusões		100
5.1	Contribuições	101
5.2	Trabalhos Futuros	102
Referências Bibliográficas		104
Apêndice A – Diagrama de Casos de Uso da Ferramenta BOE		112
Apêndice B – Modelo de Classes do Módulo de Projeto do <i>Workflow</i>		114
Apêndice C – Programa Prolog Criado pelo Módulo de Inferência		116
Apêndice D – Esquema Lógico da Camada de Banco de Dados		127

Índice de Figuras

Figura 1.1 - Arquitetura SPeCS	6
Figura 2.1 - Exemplo de Experimento Científico (CHEN & MARKOWITZ, 1995a) modificado	14
Figura 2.2 - Ferramentas e Arquitetura OPM (CHEN & MARKOWITZ, 1995a)	17
Figura 2.3 - Arquitetura do LSEA (CAVALCANTI, MATTOSO et al., 2002)	20
Figura 2.4 - Arquitetura do Sistema CHIMERA (FOSTER, VOCKLER et al., 2002)	22
Figura 2.5 - Relacionamentos entre a terminologia básica de <i>workflows</i> (WFMC, 1999)	28
Figura 2.6 - Escala de valores de dado, informação, conhecimento e ação (SANTOS & CERANTE, 2000) – modificado.	37
Figura 2.7 - Espiral do Conhecimento	47
Figura 2.8 - Arquitetura do ambiente AGROMET	58
Figura 3.1 - Diagrama de classes exibindo o auto-relacionamento entre peças e seus componentes com o desdobramento em um agregado recursivo	61
Figura 3.2 - Arquitetura da Ferramenta BOE	64
Figura 3.3 - Portal do Pesquisador Utilizado para Definir um <i>Workflow</i>	65
Figura 3.4 - Portal do Pesquisador Efetuando uma Busca por Palavra-Chave	66
Figura 3.5 - Portal do Fornecedor Utilizado para Cotação de Dados Científicos	67
Figura 3.6 - Exemplo de <i>Workflow</i> Científico Representado no Diagrama	70
Figura 3.7 – Gerenciamento de Artefatos Ligados à uma Transição no Projeto do <i>Workflow</i>	70
Figura 3.8 - Avaliação Colaborativa da Qualidade de Dados Científicos	72

Figura 3.9 - Uso do Módulo de Gerência de Qualidade	74
Figura 3.10 - Exemplo de Atividades de um <i>Workflow</i>	78
Figura 3.11 – Exemplo de Resultado do Módulo de Inferência em XML	81
Figura 3.12 - Esquema Utilizado para Realização das Inferências	85
Figura 4.1 - Representação Gráfica do <i>Workflow</i> de Zoneamento Agroclimático (Parte 1)	90
Figura 4.2 - Representação Gráfica do <i>Workflow</i> de Zoneamento Agroclimático (Parte 2)	91
Figura 4.3 - Resultado do Planejamento para o Zoneamento, mostrando explicações	91
Figura 4.4 - <i>Workflow</i> para Determinação de Dados de Reservas Petrolíferas	93
Figura 4.5 - Resultado do Planejamento do <i>Workflow</i> para Determinação de Dados de Reservas Petrolíferas	94
Figura 4.6 - Resultado da Busca por <i>Workflows</i> para Citometria de Fluxo	96
Figura 4.7 - Representação de um <i>Workflow</i> para Citometria de Fluxo	97
Figura 4.8 - Melhores Alternativas para o Histograma de Citometria de Fluxo	98
Figura 4.9 - Melhores Alternativas para Células Preparadas	99

Índice de Tabelas

Tabela 2.1 - Módulos do Ambiente ZOO	18
Tabela 2.2 - Sistemas x Atuação nas Etapas do Ciclo de Vida de um Experimento	23
Tabela 2.3 - Quadro comparativo entre <i>workflows</i> científicos e de escritório	31
Tabela 2.4 - Características de dado, informação e conhecimento SANTOS & CERANTE apud (DAVENPORT & PRUSAK, 1998)	38
Tabela 2.5 - Interseções das dimensões de tipificações de conhecimento	42
Tabela 3.1 - Analogia BOM e BOE	63
Tabela 3.2 - Principais Problemas Associados às Características de Qualidade	75

Capítulo 1 - Introdução

1.1 Motivação

Através dos tempos, as relações de trabalho sofreram diversas evoluções. Entretanto, a atribuição das tarefas mais complexas para os colaboradores de competência específica é uma tendência que não deve se extinguir. Normalmente, supervisores ou gerentes atribuem o trabalho a ser realizado a determinados recursos com base em critérios de treinamento, habilidades individuais e experiências anteriores. Inicialmente, tais recursos eram somente as pessoas, apoiadas por ferramentas como máquinas de escrever, de calcular, formulários padronizados etc. Eventualmente, uma atividade ou outra era automatizada (por exemplo, a totalização de quantias de uma venda), mas somente após um funcionário entrar com os dados necessários. Embora a realização do trabalho fosse automatizada mesmo que parcialmente, pouco mudou na forma de gerenciamento: gerentes delegam tarefas e monitoram desempenho.

Nos últimos anos, mecanismos foram desenvolvidos não somente para apoiar ou automatizar a realização do trabalho, mas também para o gerenciamento do seu fluxo (denominado *workflow*), contemplando a automação completa para a realização das tarefas, as informações que trafegam em decorrência do andamento do trabalho, a definição exata de objetivos, dentre outros.

O gerenciamento do *workflow*, o qual será referenciado simplesmente como *workflow*, pode ser aplicado utilizando-se técnicas “manuais”, isto é, sem a intervenção de computadores ou ferramentas, ou com a utilização dessas últimas, atualmente a forma de aplicação mais difundida. É até difícil a dissociação do conceito de *workflow* da sua implementação e utilização por ferramentas computadorizadas, dado que essa tecnologia busca utilizar todas as facilidades oferecidas para melhor servir a seus propósitos. Em resumo, num ambiente de gerenciamento de fluxo de trabalho, todo o processo é gerenciado por um sistema computadorizado que delega atividades e acompanha sua realização.

Não há dúvidas de que a tecnologia de gerenciamento do fluxo de trabalho afeta diretamente a maneira pela qual este é realizado. Automatizando processos, pode-se **reduzir o nível de interação humana** necessária, bem como seus **custos**

operacionais com papel, por exemplo, visto que o uso de documentos eletrônicos torna-se uma realidade. **Reduz-se também o tempo de conclusão** desses processos, pois a distribuição de informações, documentos e itens de trabalho é feita automaticamente. O **controle** sobre tais processos também **aumenta**, visto que informações estatísticas sobre a sua execução podem ser mantidas e acessadas por um gerente ou supervisor para a **tomada de decisão** quanto à **melhoria** de suas características e de seu **desempenho** geral.

A tecnologia *workflow* foi concebida inicialmente para o trabalho em empresas, na denominada “automação de escritório”. Entretanto, quando se observa o trabalho de um cientista que realiza experimentos, compostos por uma série de atividades em seqüência, necessitando de dados iniciais e gerando resultados, acontece a motivação para aplicação dos conceitos de *workflow* para o trabalho científico. A especificação, monitoramento e controle de experimentos complexos, bem como a análise de resultados experimentais gerados são etapas importantes no ambiente de pesquisa. O controle de execução de atividades complexas é reconhecidamente a principal funcionalidade de um sistema gerenciador de *workflows*. Estas afirmações comprovam a tendência de sucesso na aplicação desse tipo de tecnologia no ambiente científico.

Quando se cogita o uso de *workflows* na pesquisa científica, implicitamente está sendo posto em discussão o papel de um banco de dados científicos de gerência de experimentos. Normalmente, um *workflow* tem uma definição formal em que suas atividades e fluxos são estabelecidos. Este estabelecimento necessita de armazenamento e isto se dá normalmente em um ambiente de banco de dados. ROBBINS (1990) define este tipo de ambiente como um banco de dados científicos de gerência de experimentos. Esta associação de conceitos torna importante a análise deste tipo específico de aplicação de banco de dados.

Usuários de banco de dados científicos trabalham com uma larga variedade de ambientes, desde um único computador pessoal ou estação de trabalho, passando por supercomputadores, e chegando a sistemas heterogêneos configurados em rede, tais como *grids* (FOSTER, KESSELMAN et al., 2001). Da mesma forma, os requisitos de flexibilidade, eficiência e capacidades de comunicação variam drasticamente. Cada vez mais, os cientistas se dão conta de que o seu trabalho requer dados de diversos tipos e de

diferentes fontes. Assim, os sistemas de bancos de dados científicos devem tratar um ambiente complexo e em constante evolução, além de localizar e recuperar informações e conhecimentos úteis de forma que um pesquisador consiga tirar proveito deles para um correto planejamento e execução de suas atividades.

A era do conhecimento, também chamada de era da informação, nasce em um ambiente de mudanças com uma economia baseada não mais em recursos naturais e físicos, como na era industrial, mas baseada em recursos como o conhecimento e a comunicação. A comprovação da importância do conhecimento para o crescimento organizacional tem gerado nas empresas a preocupação em gerenciar este precioso recurso da melhor forma possível. Desta forma surgiu a disciplina Gestão do Conhecimento (GC), que se propõe a oferecer instrumentos que auxiliem as empresas a transformar o conhecimento em uma fonte de vantagem competitiva.

Da mesma forma que a tecnologia *workflow*, a GC foi inicialmente concebida para dar apoio a processos de explicitação de conhecimento em empresas, ou seja, o conhecimento deveria se tornar um ativo empresarial e não existir somente no cérebro dos funcionários (DAVENPORT & PRUSAK, 1998; NONAKA & TAKEUCHI, 1995; SVEIBY, 2001). Conforme será visto mais adiante nesta dissertação, é característica do conhecimento científico a sua existência na forma tácita, de difícil representação e transmissão. Este fato traz motivação para que se criem mecanismos que dêem tratamento adequado ao conhecimento científico. Tais mecanismos precisam ser capazes de tornar explícito o que é tácito, promovendo reuso e colaboração entre pesquisadores de diferentes áreas.

CORMICAN & O'SULLIVAN (2000) argumentam que o uso da tecnologia da informação está tendendo para o uso em grupo (decisão em grupo, listas de discussão, etc). Nos últimos anos, a *world wide web (web)* teve um espantoso aumento de tamanho, se tornando um canal de comunicação bastante interativo. Com o adequado uso da *web* fica possível o trabalho em grupo para geração de idéias, discussão de problemas e tomada de decisão, nos quais indivíduos podem colaborar uns com os outros independentes da sua localização geográfica.

A motivação deste trabalho está na possibilidade do uso eficiente dos benefícios da tecnologia *workflow* para propósitos científicos através da aplicação de

conceitos advindos de bancos de dados científicos (gerência de experimentos científicos e planejamento), gestão do conhecimento (explicitação e reuso de conhecimento científico) e colaboração, usando a *web* como principal canal de comunicação.

1.2 Objetivos

O objetivo principal desta dissertação é a definição e implementação de um sistema colaborativo para explicitação, reuso e planejamento de *workflows* científicos. Este sistema é definido na ferramenta *Bill of Experiments* (BOE) e utiliza conceitos advindos de diversas áreas de pesquisa, tais como *workflow*, bancos de dados científicos, trabalho cooperativo auxiliado por computador (*Computer Supported Cooperative Work – CSCW*), gestão do conhecimento e inteligência artificial.

Conforme será visto nos próximos capítulos, a explicitação de conhecimento é um dos tópicos de pesquisa mais importantes na área de GC. Acredita-se que a existência de definições de experimentos científicos em um banco de dados científicos seja uma forma de conhecimento explícito (CARVALHO & FERREIRA, 2001; NELH, 2003). Dispondo deste tipo de conhecimento, a combinação e reuso são incentivados quando se propõe um ambiente colaborativo. Ao mesmo tempo, o efetivo planejamento e realização de um experimento por meio de técnicas de *workflow* são de grande valia para que se obtenha agilidade no trabalho científico, com a geração otimizada de resultados experimentais.

A ferramenta BOE é um sistema *web* direcionado para cientistas e empresas fornecedoras de artefatos científicos. Ela é capaz de apontar melhores alternativas no planejamento da realização de experimentos científicos baseando-se em critérios de custo, prazo de obtenção e qualidade, em uma analogia com o processo de *Bill of Materials* (BOM). Para uma melhor compreensão desta proposta, estabelece-se uma analogia entre a preparação do experimento científico e a compra de itens para montagem de componentes. O BOE age como um portal de *internet* que une pesquisadores, instituições e empresas em um paradigma semelhante ao *Business to Business*¹ (B2B).

¹ Modalidade de comércio eletrônico praticada entre empresas que negociam produtos e serviços diretamente, utilizando a *internet* como principal meio de comunicação. Outra tradicional modalidade de

Outro objetivo, não menos importante, da ferramenta BOE é prover definições de *workflow*, dados, modelos, programas, e outros artefatos que digam respeito a experimentos científicos. Na inexistência destes, com o auxílio de um ambiente colaborativo, o BOE orienta o pesquisador na sua obtenção, tornando-se uma ferramenta para melhorar o acesso, compartilhamento e conseqüente inovação do conhecimento científico.

1.3 Contexto

A ferramenta BOE está inserida no contexto do ambiente SPeCS (*Spatial Decision Collaborative Support*) (PALMA, 2002; PALMA, SOUZA et al., 2001). O SPeCS, desenvolvido na COPPE / UFRJ, provê um ambiente de trabalho cooperativo comum, flexível e de fácil utilização, em que os membros de um grupo podem estar espalhados geograficamente em diversos ambientes heterogêneos e mesmo assim são capazes de interagir durante um processo de tomada de decisão.

A arquitetura do SPeCS é composta por quatro camadas: Ferramentas de Decisão, Ferramentas de Conhecimento, Camada de Serviços Básicos e Repositório de Conhecimento (Figura 1.1).

No SPeCS, as ferramentas de decisão representam e coordenam as atividades envolvidas no processo de tomada de decisão, ou seja, desde a fase de definição do problema até a fase de especificação e documentação de todas as eventuais soluções levantadas durante as discussões. Estas ferramentas ainda possibilitam aos membros do grupo priorizar os assuntos no ambiente, realizar diferentes tipos de votação, avaliar diferentes critérios, elaborar documentos em grupo, realizar seções de *brainstorming*, analisar projetos, ordenar soluções e permitir que argumentações e informações utilizadas durante as discussões sejam georeferenciadas. Em (CASTRO, OLIVEIRA et al., 2002), a ferramenta **Decisio**, que implementa diversas funcionalidades das ferramentas de decisão propostas para o SPeCS é apresentada com um estudo de caso na Agrometeorologia.

comércio eletrônico é denominada *Business to Consumer* (B2C). Ela representa o negócio entre um comerciante e um consumidor final.

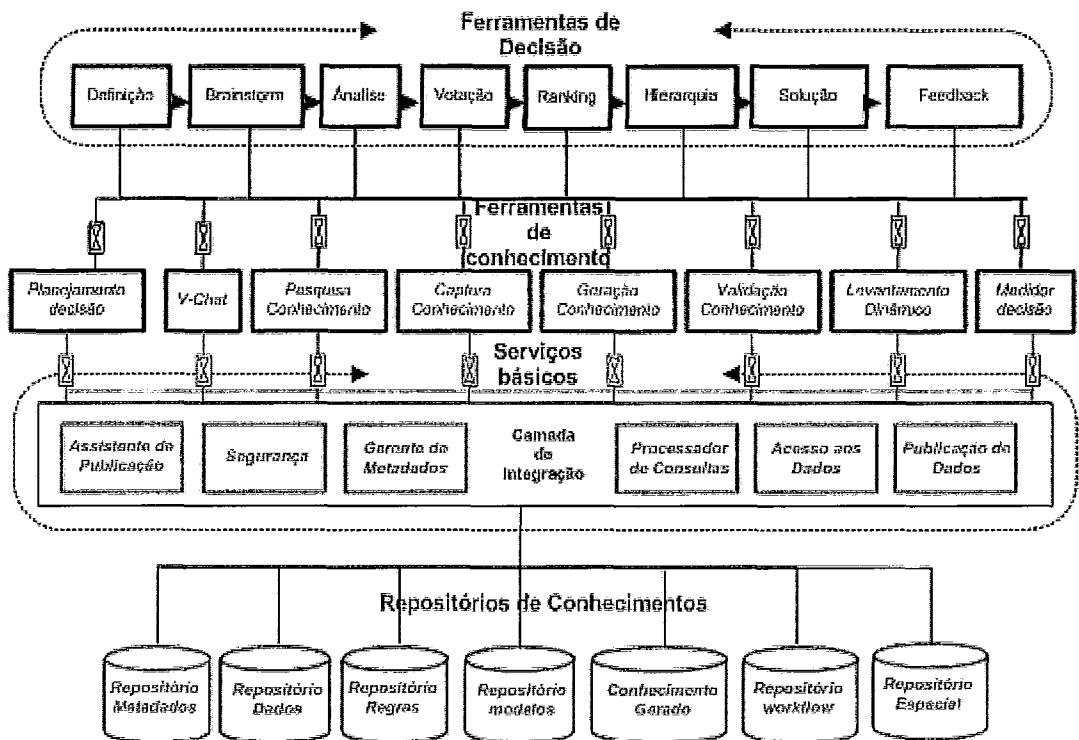


Figura 1.1 - Arquitetura SPeCS

As ferramentas de conhecimento contemplam mecanismos para o planejamento e medições de decisões, bate-papo visual, controle de questionários e gestão do conhecimento. Algumas destas ferramentas podem ser configuradas de acordo com o fluxo de trabalho que o grupo segue durante a execução de suas tarefas. A ferramenta **Epistheme** (OLIVEIRA, SOUZA et al., 2002) trata da aquisição, identificação, integração, validação e criação de conhecimento em ambiente científico. O **BOE**, tema desta dissertação, também figura na camada de ferramentas de conhecimento e provê mecanismos para **planejamento de decisão**. Com menos intensidade, o BOE atua no lado tecnológico da gestão do conhecimento, auxiliando a **pesquisa** e a **captura** deste.

A camada de serviços básicos do SPeCS permite a integração e o compartilhamento de bases de dados heterogêneas e distribuídas através de mediadores que possibilitam a interoperabilidade entre os diversos repositórios de dados espalhados através da *web*, mantendo a autonomia dos mesmos. A ferramenta **X-Arc** (PINTO, SOUZA et al., 2001) implementa a camada de integração de dados do SPeCS.

1.4 Organização dos Capítulos

A presente dissertação está organizada em cinco capítulos. O capítulo dois traz uma revisão da literatura a respeito dos temas relacionados com esta pesquisa: bancos de dados científicos, CSCW, *workflow* e gestão do conhecimento científico. A proposta da ferramenta BOE e sua arquitetura são apresentadas no capítulo três. No capítulo quatro, são apresentados um estudo de caso aplicado na área de agrometeorologia, em uma cooperação com a Embrapa Solos, e exemplos de utilização do ambiente em outras situações reais. Finalmente, o capítulo cinco conclui esta dissertação apontando as contribuições e possíveis melhoramentos neste trabalho.

Capítulo 2 - Revisão da Literatura

Para haver uma boa compreensão a respeito dos conceitos estudados para a implementação da ferramenta BOE, é importante realizar uma revisão do que foi objeto de estudo nos últimos anos nas áreas relacionadas com esta pesquisa.

Em primeiro lugar, os bancos de dados científicos, especialmente a área de gestão de experimentos científicos (*Scientific Experiment Management*), serão examinados. Em seguida, a tecnologia de fluxos de trabalho (*workflow*) será discutida e associada com trabalhos que apóiam o ciclo de vida de um experimento científico. Finalmente, a gestão do conhecimento será definida e discutida no âmbito organizacional e científico para que seja possível entender como técnicas desta tecnologia emergente podem ser aplicadas para atingir os objetivos deste trabalho.

2.1 Bancos de Dados Científicos

Esta seção irá discutir os trabalhos realizados na área de bancos de dados científicos (*Scientific Databases – ScDBs*), que está intimamente relacionada com a pesquisa desenvolvida neste trabalho.

2.1.1 Definição

Através de sensores, experimentos e simulações, dados científicos vêm crescendo em volume e complexidade em uma escala acentuada. O custo da produção de dados é muito alto: satélites, aceleradores de partículas e centros de supercomputação representam fontes de geração de informação que coletivamente custam bilhões. Na década anterior houve um crescimento enorme em capacidade de computação, diminuíram-se os custos dos espaços de armazenamento e o acesso a redes de comunicação rápidas e confiáveis se tornou uma realidade. Sem meios efetivos de recuperação, análise e manipulação desses dados, o investimento feito não trará os benefícios esperados para a sociedade.

ZEMANKOVA & IOANNIDIS (1994) definem bancos de dados científicos como um repositório de conhecimentos de duas categorias: aqueles que já existem e os que aguardam descoberta. O surgimento de disciplinas baseadas em ciência

computacional atesta a tendência de derivação de conhecimento científico a partir de dados, em conjunto com as abordagens mais tradicionais que utilizam experimentos.

Bancos de dados científicos são utilizados para armazenamento e acesso a dados que descrevem objetos de interesse científico, como proteínas, seqüências de DNA, mapas de regiões e experimentos científicos. A maioria das dificuldades no acesso a informação nos bancos de dados científicos advém do enorme volume de dados envolvidos. Entretanto a organização destes dados é também um problema sério. Usuários de diversas comunidades científicas observam relacionamentos diferentes entre os mesmos conjuntos de dados. Além disso, esta área não-convencional necessita de informações adicionais que se encontram em esboços gráficos, diagramas, mapas, fotografias, sons etc, e que não podem ser automaticamente manipuladas de forma estruturada na rígida estrutura dos bancos de dados relacionais.

Tradicionalmente, a pesquisa sobre projeto, desenvolvimento, gerência e uso de bancos de dados se focou em conceitos e requisitos críticos para o ambiente de negócios. Conseqüentemente, as tecnologias de bancos de dados atuais sofrem limitações para dar suporte a aplicações de natureza científica que são caracterizadas por diversos e complexos tipos de dados, enorme volume de dados, necessidade de processamento especial de dados, distribuição em nível mundial de usuários e informações, e extrema heterogeneidade entre as plataformas adotadas.

Existem significativas diferenças entre dados científicos e dados convencionais, geralmente gerados no ambiente de negócios citado no parágrafo anterior. Esforços para identificação destas diferenças (FRENCH, JONES et al., 1990; PFALTZ, 1990) foram feitos com o objetivo de dar o tratamento adequado aos dados científicos. As principais diferenças identificadas por estes trabalhos são:

- **Volume de Dados** – embora seja conhecida a existência de enormes bancos de dados convencionais em empresas, normalmente os dados que residem nestes sistemas são gerados a partir de procedimentos manuais, por exemplo, um recebimento de nota fiscal, ou liberação de uma quantidade de material de um estoque. Por outro lado, dados científicos são freqüentemente gerados de maneira automática, através de sensores remotos ou mesmo por dispositivos que permitem a realização automática de experimentos, como aceleradores de

partículas. Estes fatos, embora notados no início da década passada, ainda são válidos, pois ao passo em que o volume de dados convencionais teve um aumento considerável com a evolução de tecnologias (telefonia móvel, pregões e bolsas de valores eletrônicos etc), os dispositivos científicos da atualidade também evoluíram e passaram a exigir mais espaço e eficiência no seu armazenamento. Conclui-se, então, que o volume de dados científicos tende a ser maior do que o de dados convencionais.

- **Valor Temporal** – muitas vezes dados convencionais gerados em ambientes de negócios perdem seu valor após dias, meses ou anos. Tal fato não acontece com dados científicos, pois estes dificilmente perdem o seu valor ao longo do tempo. Por exemplo, a situação de vendas de uma empresa no mercado em 1960 dificilmente terá utilidade para a tomada de alguma decisão nos dias atuais. Entretanto, dados científicos bastante antigos podem ser utilizados em experimentos em dias atuais, como no caso de estudo de séries temporais de quantidade de chuvas em uma região, na ocasião da realização de um zoneamento agroclimático.
- **Maneira de Agregação** – um banco de dados convencional tipicamente envolve um número limitado de valores de dados (entenda-se por valor de dados como um número ou uma *string*). No modelo relacional, estes valores de dados são agregados em tuplas e estas em relações, formando dois níveis de agregação. Conjuntos de dados científicos necessitam de ao menos mais um nível de agregação pois normalmente ocorrem em forma de vetores ou matrizes. Além deste fato, o volume de dados citado acima induz a criação prévia de agregações de dados científicos para permitir facilidade de processamento posterior. Esta característica foi inicialmente apresentada na área de bancos de dados estatísticos (MICHALEWICZ, 1992) e passou a ser utilizada em larga escala nos ambientes convencionais com o advento da tecnologia de *Data Warehouses*.
- **Necessidade de Metadados** – O nível de detalhe dos metadados científicos precisa ser muito maior do que os convencionais. Neste ambiente, os dados são pesquisados e encontrados a partir de seus metadados e não dos dados em si. Por exemplo, os metadados de um experimento científico consistem de uma coleção

de resultados, uma série de etapas e uma descrição estruturada do contexto em que ele foi conduzido.

Quando se discute o termo banco de dados científicos, alguma confusão pode surgir em razão das diferentes noções que as pessoas têm do que realmente constitui um banco de dados científicos. Devido a isso, é importante tentar impor uma categorização para a área e examinar cada tipo identificado.

As próximas subseções descrevem os tipos de bancos de dados científicos identificados por ROBBINS (1990) na sua classificação funcional.

2.1.1.1 Gerência de Experimentos

Sistemas de gerência de experimentos são utilizados para armazenar e gerenciar dados experimentais para análise imediata e para planejamento de experimentos correlatos. Normalmente, são utilizados pelos próprios pesquisadores que são responsáveis pela geração dos dados. *Laboratory Information Management Systems* (LIMS), que auxiliam o pesquisador no seu dia a dia, são bons exemplos de sistemas de gerência de experimentos.

Um sistema de gerência de experimentos deve ser flexível e eficiente porque protocolos científicos mudam constante e significativamente ao longo do tempo, e esta classe de sistemas é usada dinamicamente para planejamento e interpretação de experimentos em andamento. Deve ser possível a rápida modificação de objetos destes sistemas e ainda existir uma interface simples que permita aos usuários não avançados em computação desempenhar os seus trabalhos de maneira produtiva.

2.1.1.2 Arquivamento e Recuperação de Dados

Estes sistemas são utilizados para armazenamento, gerência e recuperação de dados (muitas vezes sem qualquer tratamento – *raw data*) para análises posteriores. Em geral, gerenciam dados para uso a longo prazo e por muitos pesquisadores, inclusive aqueles que não participaram do trabalho de coleta e geração. Um bom exemplo destes sistemas é o armazenamento de dados de telemetria de sondas espaciais.

A coleção destes dados somente tem valor quando é acompanhada de documentação e metadados apropriados. Somente com estas características que

cientistas interessados poderão descobrir que determinado dado armazenado no sistema é de fundamental importância para a sua pesquisa.

2.1.1.3 Recuperação de Literatura

Sistemas de recuperação de literatura se destinam a prover acesso à literatura científica. Eles têm muita interseção com sistemas de recuperação de informação e bancos de dados. Exemplos atuais de sistemas de recuperação de literatura são o PubMed (<http://www.ncbi.nlm.nih.gov/PubMed/>) para a literatura médica e o CiteSeer (<http://citeseer.nj.nec.com/cs>) que indexa publicações de ciência da computação.

O objetivo destes sistemas é auxiliar cientistas na busca por literatura potencialmente útil.

2.1.1.4 Recuperação de Fatos Baseados na Literatura

Estes sistemas oferecem uma possibilidade importante de pré-interpretação de fatos existentes na literatura. Primeiramente, leitores *experts* lêem e combinam a literatura para extrair fatos. Em seguida, estes fatos são incluídos em um banco de dados. Um exemplo de fato do meio biológico é: “o corpo humano é capaz de sintetizar vitamina D quando exposto aos raios solares”.

Sistemas de Conhecimento Público Não-Descoberto (*Undiscovered Public Knowledge – UPK*), originalmente propostos por SWANSON (1991), são bons exemplos deste tipo de bancos de dados. Nestes sistemas, existe uma busca por conhecimento que está no domínio público, mas que ainda é desconhecido a qualquer pessoa porque é somente acessível de forma não agregada, devendo ser recuperado, montado e interpretado.

Conhecimento pode ser publicado, contudo desconhecido a qualquer um porque não tem nenhuma representação explícita. Por exemplo, as relações lógicas implícitas entre dois pedaços separados de informação publicada. Um exemplo prático: Documento 1 contém evidências para aceitar a hipótese que A causa B. O documento 2 evidencia que a hipótese B causa C. Logo, embora explicitamente não existente, a hipótese A causa C pode ser sugerida.

Com a evolução da *web*, esta passou a ser a principal fonte de busca por conhecimento em sistemas UPK. Técnicas de *web* mining e de descoberta de conhecimento são empregadas para a implementação desta categoria de sistemas.

2.1.1.5 Recuperação de Fatos Não Baseados na Literatura

Este tipo de sistema se diferencia do anterior porque apenas fatos que nunca apareceram na literatura são tratados. Exemplos de tais fatos não publicados são: artigos internos, relatórios internos de projetos, hipóteses não testadas, sugestões, conclusões preliminares, idéias soltas, *insights* etc.

Dado que não existe o controle rígido exigido para uma publicação, esquemas de revisão são empregados com a finalidade de se garantir a boa qualidade dos fatos existentes na base.

2.1.2 Gerência de Experimentos Científicos

Como exposto na seção anterior, a gerência de experimentos é uma classe importante dentro do contexto de bancos de dados científicos. Conforme será exposto no capítulo três desta dissertação, a gerência de experimentos científicos tem um importante papel neste trabalho, pois existe um banco de dados de protocolos científicos na ferramenta BOE. Esta seção apresenta as características destes sistemas e trabalhos relacionados com esta pesquisa.

Geralmente, experimentos científicos são complexos, apresentam alta interatividade, consomem tempo e necessitam de instrumentos e locais apropriados. Para os cientistas envolvidos no processo é essencial o acompanhamento da geração de resultados para que seja possível: a verificação destes em um passo posterior, a reprodução do experimento por outros pesquisadores, o esclarecimento de dúvidas, e a criação de novos experimentos relacionados. Desta forma, são necessárias ferramentas de gerência de experimentos científicos eficientes que acelerem e melhorem a qualidade da produção de artefatos científicos, que evitem desnecessária duplicação de esforços, e que melhorem o próprio entendimento da ciência.

De uma maneira geral, experimentos científicos são caracterizados por propriedades como: tempo, local de realização e pessoas envolvidas. Além disso, são características marcantes de um experimento científico:

- a transformação de parâmetros de entrada em resultados experimentais;
- a possibilidade de definição de precedências em relação a outros experimentos, de forma que seus parâmetros de entrada possam ter sido gerados através de resultados de outro experimento, ou que seus resultados experimentais sejam aplicados a novos experimentos subsequentes². Esta característica permite a associação do conceito de experimento ao de *workflow*, que será visto mais adiante neste trabalho; e
- a hipótese de estar incluído em um experimento mais geral, visto como de mais alto nível.

CHEN & MARKOWITZ (1995a) apresentam um simples experimento científico para ilustração e entendimento das características deste. No exemplo, mostrado na Figura 2.1, fragmentos de DNA são utilizados como parâmetros de entrada que são analisados automática ou manualmente, resultando em conexões que finalmente são transformadas em mapas de continuidade, que é o resultado experimental esperado.

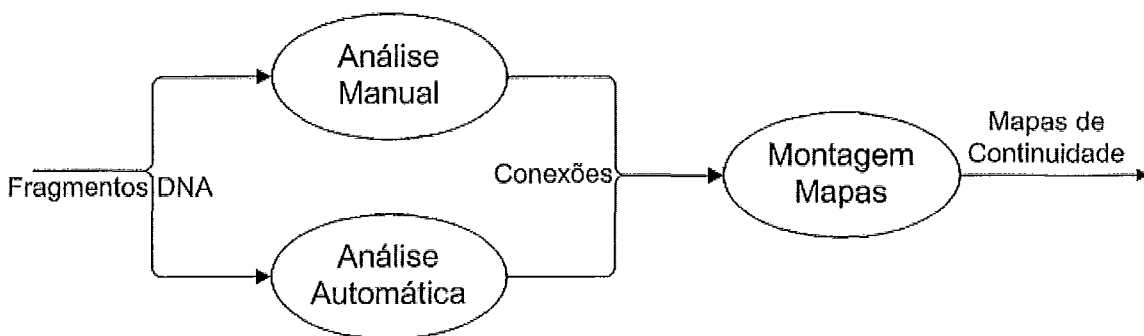


Figura 2.1 - Exemplo de Experimento Científico (CHEN & MARKOWITZ, 1995a) modificado

2.1.2.1 Ciclo de Vida de Experimentos

IOANNIDIS, LIVNY et al. (1996) argumentam que um experimento típico tem um ciclo de vida pré-definido, contendo as seguintes etapas:

² O rastreamento de dados científicos derivados (*data lineage* ou *data provenance*) é um problema bastante estudado na literatura (FOSTER, VOCKLER et al., 2002; SALTZ, 2002).

1. **Projeto do Experimento (*Experiment Design*)** – consiste na estrutura do experimento, especifica quais variáveis ou dados serão controlados e que resultados experimentais serão obtidos como saída.
2. **Coleta de Dados (*Data Collection*)** – consiste na execução do experimento, em que os resultados experimentais são efetivamente obtidos (coletados). Este trabalho é geralmente feito com auxílio de equipamentos laboratoriais ou com ferramentas de simulação.
3. **Análise de Dados (*Data Exploration*)** – os resultados experimentais são analisados e se formam conclusões a respeito do experimento conduzido. Ferramentas de recuperação e visualização de dados são importantes nesta etapa.

Tradicionalmente, experimentos científicos sempre foram desenvolvidos de maneira isolada, com cada cientista trabalhando com um conjunto limitado de modelos previamente estudado e analisado por si mesmo e por sua equipe. Com a evolução da tecnologia, dados científicos de diferentes fontes passaram a estar disponíveis, principalmente em meio eletrônico. A utilização de tais dados é muito importante para que cada pesquisador possa aprimorar o seu trabalho, analisando outras opções até então não pensadas. A evolução natural deste fato é que se passe a cogitar o compartilhamento de artefatos científicos, como a definição de protocolos de trabalho, modelos matemáticos, programas (implementação dos modelos) etc.

Por exemplo, a situação do mercado de serviços de meteorologia foi descrita por REPELLI (2001). Um mercado potencial de serviços de meteorologia criou-se uma vez que a meteorologia conseguiu estabelecer-se como uma “tecnologia”, e que seus usuários podiam usá-la no processo de tomada de decisões para maximizar seus lucros. Consumidores e fornecedores, produtos (informações) e seu fluxo são premissas para que se estabeleça o mercado descrito.

Devido a estes fatos, surge a necessidade de adicionar mais uma etapa no ciclo de vida de um experimento típico: a etapa de planejamento do experimento, que figura como segundo passo. Este trabalho propõe a extensão do ciclo de vida proposto por IOANNIDIS, LIVNY et al. (1996) da seguinte forma:

1. **Projeto do Experimento (*Experiment Design*)** – consiste na estrutura do experimento, especifica quais variáveis ou dados serão controlados e que resultados experimentais serão obtidos como saída. Ainda contempla a organização do fluxo de trabalho dentro do experimento, ou seja, a definição do seu *workflow* científico.
2. **Planejamento do Experimento (*Experiment Planning*)** – consiste no planejamento dos artefatos científicos que se deseja utilizar no experimento. Um dado base pode ser obtido a partir de diversos locais, e modelos matemáticos, criados por diferentes cientistas, podem ser executados em diferentes máquinas. A correta escolha destas fontes de artefatos pode levar a melhores resultados na etapa seguinte, de coleta de dados experimentais.
3. **Coleta de Dados (*Data Collection*)** – consiste na execução do experimento, em que os resultados experimentais são efetivamente obtidos (coletados). Caso exista alguma definição de fluxo de trabalho, esta deverá ter ser execução acompanhada nesta etapa. Este trabalho é geralmente feito com auxílio de equipamentos laboratoriais ou com ferramentas de simulação.
4. **Análise de Dados (*Data Exploration*)** – os resultados experimentais são analisados e se formam conclusões a respeito do experimento conduzido. Ferramentas de recuperação e visualização de dados são importantes nesta etapa.

A partir de agora, toda referência a ciclo de vida de um experimento científico será referente a este ciclo estendido, que contempla a etapa de planejamento do experimento.

2.1.2.2 Trabalhos Relacionados

Dada a motivação e definição de sistemas de gerência de experimentos científicos, as próximas subseções irão detalhar alguns trabalhos desta área relacionados com esta dissertação.

CHEN & MARKOWITZ (1995b) propuseram um modelo de dados derivado do modelo orientado a objetos (OO) para a representação de experimentos científicos. O modelo OPM (*Object Protocol Model*) combina a riqueza semântica do

modelo OO para a modelagem de protocolos científicos produzindo uma estrutura única para representação de experimentos e dados científicos.

A proposta OPM consiste em estender o modelo de dados OO, permitindo o uso de classes, herança e atributos derivados, através da criação de classes-protocolo. Uma classe-protocolo pode ser associada com atributos (artefatos) convencionais, parâmetros de entrada, resultados experimentais (parâmetros de saída) e sub-experimentos. Através de ferramentas que implementam o modelo OPM sobre um banco de dados relacional (CHEN & MARKOWITZ, 1995a), um projetista pode criar um novo esquema (novo experimento), utilizando o Editor de Esquemas OPM. A heterogeneidade de dados científicos também é considerada quando a ferramenta Rel2OPM provê estruturas para montagem de visões sobre bancos de dados científicos relacionais já implementados (CHEN, KOSKY et al., 1997).

Os mesmos autores ainda apresentam uma série de trabalhos interessantes sobre o modelo OPM, com destaque para o gerenciamento de versões de objetos que representam as observações científicas, que evoluem ao longo do tempo (CHEN, MARKOWITZ et al., 1996). A arquitetura de ferramentas OPM se encontra na Figura 2.2.

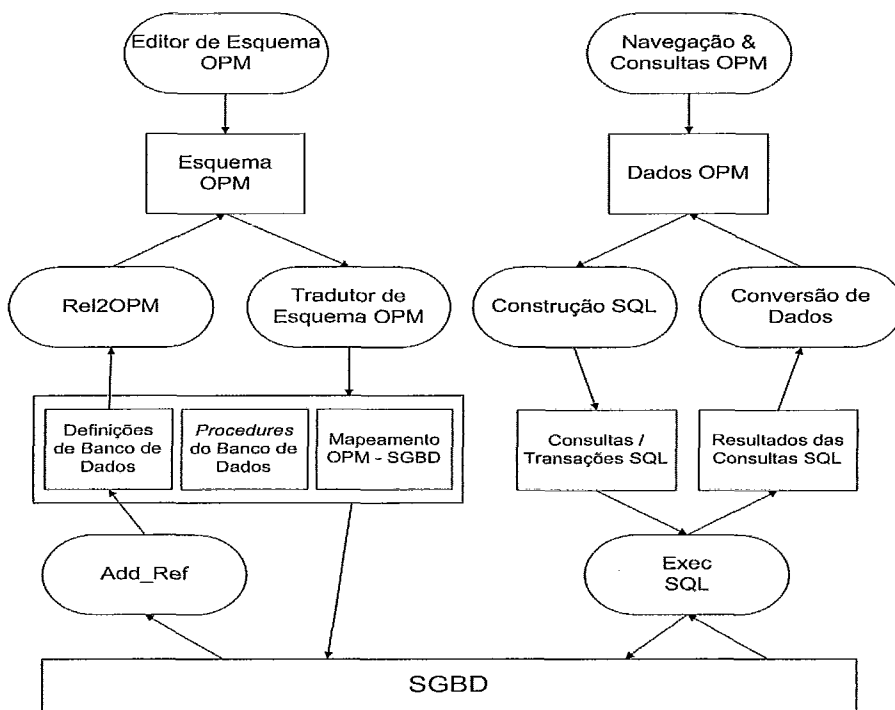


Figura 2.2 - Ferramentas e Arquitetura OPM (CHEN & MARKOWITZ, 1995a)

Seguindo a mesma linha de sistemas que o OPM, o ambiente de gerência de experimentos científicos ZOO (IOANNIDIS, LIVNY et al., 1996) prevê solução para o problema da enorme geração de dados científicos que se deu sem o desenvolvimento de ferramentas de bancos de dados apropriadas. O sistema ZOO foi desenvolvido a partir de colaboração com pesquisadores da área de solos e bioquímica.

Os principais objetivos do ambiente ZOO são dar apoio ao ciclo de vida da realização de experimentos e ser um ambiente integrado de software. Com relação ao ciclo de vida, mais especificamente tem-se que o ambiente deve (a) dar suporte as etapas 1, 3 e 4 do ciclo de um experimento permitindo transições suaves entre elas, (b) gerenciar transparentemente todos os resultados experimentais gerados no processo, e (c) ocultar os detalhes de qualquer software de apoio que venha a ter sido usado. Outro objetivo chave consiste em permitir alguma interseção entre as etapas de coleta e análise de dados, uma vez que uma análise pode eventualmente envolver novas coletas de resultados experimentais.

Tabela 2.1 - Módulos do Ambiente ZOO

Módulo	Descrição
EMU	Gerente de experimentos
FOX	Linguagem de consulta orientada a objetos declarativa
FROG	Ferramenta gráfica para especificar mapeamentos entre objetos MOOSE e arquivos ASCII
HORSE	Banco de dados orientado a objetos baseado no MOOSE e no FOX
MOOSE	Modelo de dados orientado a objetos
OPOSSUM	Gerenciador visual de esquemas
SQUID	Gerenciador visual de consultas
TURTLE	Tradutor entre objetos MOOSE e arquivos ASCII

Os componentes do ambiente estão resumidamente descritos na Tabela 2.1. No núcleo do sistema está o banco de dados HORSE, que implementa um modelo de dados próprio denominado MOOSE, que, como no ambiente OPM, também é derivado do modelo OO, e uma linguagem de consulta chamada FOX. O gerenciador de esquema OPOSSUM e a ferramenta de consulta SQUID compõem o ambiente de visualização dos dados. O gerenciador de experimentos EMU é responsável por transformar requisições de usuários em ações que são realizadas em sistemas externos, preparando-os para execução. A funcionalidade das classes-protocolo do OPM, que capturam o

fluxo de execução dos experimentos, é reproduzida através de relacionamentos no MOOSE que são processados pelo EMU.

A iniciativa ZENTURIO (PRODAN & FAHRINGER, 2002) busca estabelecer um sistema de gerência de experimentos (basicamente a execução de simulações matemáticas) para arquiteturas de supercomputação. Os principais objetivos do sistema são o estudo de parâmetros³, análise de desempenho e testes de softwares em arquiteturas *clusters* e GRID.

Através da linguagem de marcação de diretivas ZEN, é permitido ao pesquisador especificar faixas para parâmetros de entrada de experimentos, incluindo variáveis, nomes de arquivos, parâmetros de compilação, estratégias de agendamento, distribuição de dados etc. A diretiva ZEN é implementada através de comentários nos programas-fonte e desta formam são ignoradas por outros sistemas que desconhecem sua semântica. No momento em que o sistema ZENTURIO efetua a leitura dos arquivos de controle que utilizam diretivas ZEN, são geradas as diversas execuções de experimentos com seus respectivos resultados experimentais.

Este sistema lida com as etapas 3 e 4 do ciclo de vida do experimento científico, assumindo que o protocolo e planejamento do experimento são etapas já completadas.

CAVALCANTI, MATTOSO et al. (2002) propõem uma arquitetura baseada em mediadores (*Le Select Extended Architecture* - LSEA) que venha a facilitar a troca, reuso e disseminação de informação, gerenciando artefatos científicos distribuídos em bancos de dados heterogêneos. Os artefatos científicos passíveis de compartilhamento previstos nesta arquitetura são dados científicos, modelos e programas.

Nesta proposta, a arquitetura *Le Select* de mediação será estendida com uma camada de metadados, de forma que definições de modelos fiquem isoladas das suas implementações. Esta funcionalidade é bastante útil no cenário de planejamento do

³ Consiste na execução de uma mesma computação (modelo matemático por exemplo) diversas vezes com a utilização de diferentes parâmetros de entrada.

experimento, pois um cientista pode considerar diversas alternativas, contando com modelos similares encontrados na arquitetura, para a execução de sua pesquisa.

A arquitetura do LSEA está descrita na Figura 2.3. A camada de serviços *web* disponibiliza facilidades de busca, seleção, instanciação, monitoramento e publicação de recursos científicos. A camada de gerência de artefatos científicos é subdividida em duas partes: módulo de operação de recursos e módulo de descrição de recursos. O papel do módulo de operação de recursos é bem desempenhado pelo *Le Select* pois somente dados e programas são tratados. O módulo de descrição de recursos é um gerenciador de metadados que lida com definições de dados, programas, e ainda modelos, experimentos e *workflows*.

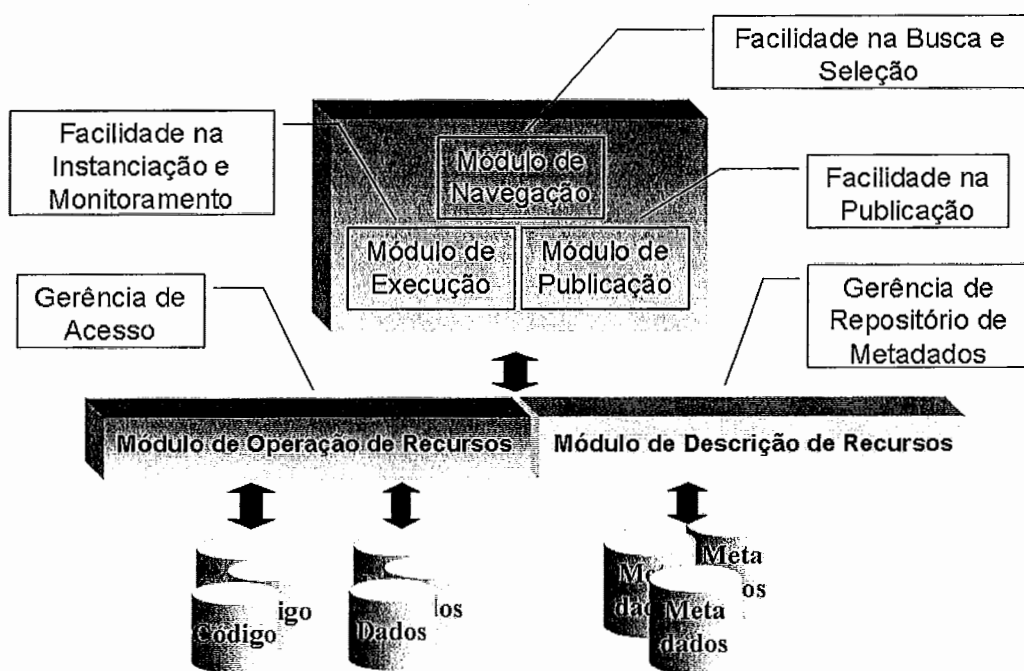


Figura 2.3 - Arquitetura do LSEA (CAVALCANTI, MATTOSO et al., 2002)

A camada de serviços *web* é composta de três módulos: publicação, execução e navegação. O módulo de navegação permite que cientistas busquem recursos científicos e suas descrições (previamente publicadas pelo módulo de publicação). Após este passo, o usuário seleciona um programa e um conjunto de dados específico para submeter como parâmetro de entrada. Se tudo correr bem, o módulo de execução interage com o módulo de operação de recursos, submetendo uma execução do programa com os dados selecionados, e monitorando sua execução. Claramente percebe-se atuação do LSEA nas etapas 2 e 3 do ciclo de vida do experimento.

A proposta da arquitetura CHIMERA (FOSTER, VOCKLER et al., 2002) é semelhante a LSEA em virtude de existir um enfoque destacado para a gerência de metadados, especialmente no monitoramento de dados derivados ou virtuais (*data provenance, data lineage*). No que diz respeito ao planejamento do experimento, o sistema pretende resolver algumas questões de forma eficiente:

- Uma vez que se obteve dados interessantes para uma pesquisa, é importante saber quais tipos de correções foram originalmente aplicadas ao conjunto de dados para verificar sua utilidade prática em outros trabalhos.
- Se já existe um programa que desempenha determinada tarefa, não é necessário construir outro.
- Se for detectado um erro de configuração em um sensor científico, quais dados derivados necessitam de recálculo?

Em resumo, é importante rastrear como dados experimentais são derivados. Um grau de precisão elevado é requerido para que se possa reproduzir e eventualmente reconstruir resultados experimentais através deste conhecimento. A arquitetura CHIMERA (Figura 2.4) contém dois componentes principais: catálogo de dados virtuais (*Virtual Data Catalog - VDC*) e interpretador da linguagem de dados virtuais (*Virtual Data Language Interpreter*).

O catálogo de dados virtuais provê uma representação compacta e expressiva de procedimentos computacionais utilizados para a criação de dados derivados (resultados experimentais ou dados virtuais). O módulo ainda é capaz de representar as execuções dos procedimentos e os dados gerados em cada uma delas, participando da etapa 3 do ciclo de vida do experimento. O interpretador da linguagem de dados virtuais tem a funcionalidade de executar requisições para construção e consulta de objetos no banco de dados.

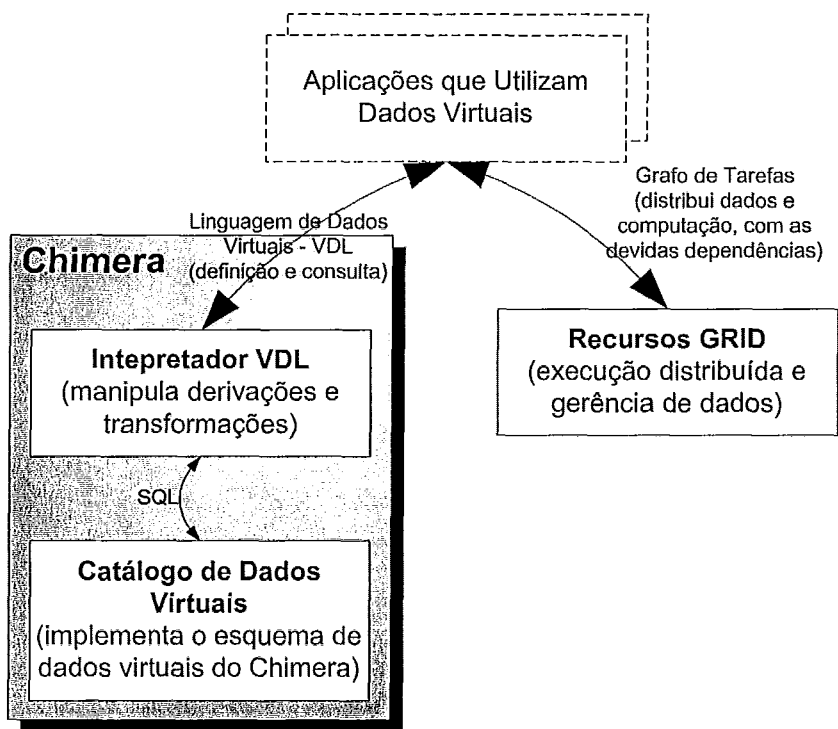


Figura 2.4 - Arquitetura do Sistema CHIMERA (FOSTER, VOCKLER et al., 2002)

Esta seção procurou mostrar sistemas que atuam em pelo menos uma etapa no ciclo de vida de um experimento científico (vide resumo na Tabela 2.2). Nenhum sistema apresentado implementa todas as etapas em sua plenitude e de maneira integrada, até mesmo porque a diversidade de plataformas, programas e dados não permitem que esta seja uma tarefa fácil. A possibilidade de distribuição, compartilhamento e reuso de artefatos científicos é uma característica importante para o aprimoramento dos trabalhos científicos que envolvem experimentos, pois com a *internet* tendo como função a disseminação de artefatos científicos, abre-se um potencial caminho para o aperfeiçoamento da pesquisa como um todo.

A ferramenta BOE conta com um banco de dados de experimentos e protocolos científicos e, conforme será visto com detalhes no capítulo três desta dissertação, dá apoio às etapas 1 e 2 do ciclo de vida de um experimento. Em virtude da citada complexidade presente em grandes bancos de dados científicos, a alternativa seguida neste trabalho foi a de representar somente a parte abstrata de um experimento ou protocolo, ou seja, sua definição. Entretanto, esta representação de metadados é feita de maneira detalhada, englobando os processos, etapas, fluxos de dados e artefatos científicos pertinentes ao experimento ou protocolo que se deseja armazenar.

É importante ressaltar ainda que, embora não atue na gerência da execução do experimento (etapas 3 e 4 do ciclo de vida do experimento) o modelo de dados empregado para a definição de protocolos científicos é comum para todas as ferramentas do ambiente SPeCS (vide seção 1.3), o que permite que experimentos ou protocolos definidos pelo BOE tenham suas execuções futuramente acompanhadas e analisadas por outras ferramentas do ambiente.

Tabela 2.2 - Sistemas x Atuação nas Etapas do Ciclo de Vida de um Experimento

	PROJETO	PLANEJAMENTO	COLETA	ANÁLISE
OPM (1995)	Sim	Não	Não	Não
ZOO (1996)	Sim	Não	Sim	Sim
ZENTURIO (2002)	Não	Não	Sim	Sim
LSEA (2002)	Não	Sim	Sim	Não
CHIMERA (2002)	Não	Sim	Sim	Não
BOE (2003)	Sim	Sim	Não	Não

2.2 CSCW e Workflow

Esta seção irá apresentar conceitos e discussões relacionados à área de trabalho cooperativo, principalmente o apoiado por computador, denominado *Computer Supported Cooperative Work* (CSCW). A tecnologia de *workflow* é bastante empregada nesta classe de sistemas e sua abordagem no contexto científico freqüentemente vem sendo encontrada na literatura. Tanto a parte conceitual de CSCW como sistemas de *workflow* científico são pertinentes ao estudo desenvolvido neste trabalho, daí a razão da abordagem destes assuntos.

2.2.1 CSCW, Cooperação e Colaboração

Segundo GRUDIN (1994), o conceito de CSCW foi introduzido em meados da década de 1980, em um seminário que reuniu pesquisadores de várias disciplinas. O principal interesse do encontro foi o estudo de como as pessoas realizavam suas atividade de trabalho e como a tecnologia existente poderia auxiliá-las. A partir de então houve um crescente interesse pelo assunto nas áreas de suporte a grupos de trabalho, sistemas de informação gerencial, educação, ciências sociais e outras disciplinas.

FLUCKIGER (1995) define CSCW como o campo de pesquisa relativo ao projeto de sistemas baseados em computador para o suporte e melhoria do trabalho em

grupo de usuários envolvidos em tarefas ou objetivos comuns, e ao entendimento dos efeitos da utilização de tais sistemas.

A definição do grupo inglês DTI (*Department of Trade and Industry*) traz alguma interseção com a de FLUCKIGER. CSCW consiste no desenvolvimento e uso de técnicas e tecnologias que promovem e facilitam um trabalho efetivo entre indivíduos, grupos, e empresas, que podem estar separadas por distância, tempo e até mesmo sistemas (DTI/CSCW, 1997).

Estas e outras definições da mesma linha estão sujeitas a críticas por algum segmento do ramo de pesquisa em CSCW. Entretanto, elas conseguem retratar um grande consenso existente nesta área de pesquisa.

Algumas confusões têm sido geradas entre os termos CSCW e *groupware*. O termo CSCW focaliza a área de pesquisa e é abrangente, conforme as definições expostas acima. Já *groupware* é um tipo de software designado para dar suporte a mais de uma pessoa trabalhando em uma mesma tarefa. Pode-se concluir que *groupware* é uma ferramenta que dá apoio às tarefas de CSCW.

A distinção entre os termos **cooperação** e **colaboração** é importante no contexto deste trabalho, para que não haja falta de entendimento da proposta da ferramenta BOE de atuar de forma colaborativa.

A partir do estudo de definições dos dois termos feitas por diversos autores, TIJIBOY & MAÇADA (1997) chegaram a uma conclusão bastante aceitável e que será utilizada neste trabalho. Segundo as autoras:

*“A diferença fundamental entre os conceitos de colaboração e cooperação reside no fato de que para haver **colaboração** o indivíduo deve interagir com o outro existindo ajuda - mútua ou unilateral. Para existir **cooperação** deve haver interação, colaboração mas também objetivos comuns, atividades e ações conjuntas e coordenadas”* (TIJIBOY & MAÇADA, 1997).

2.2.2 Tecnologia *Workflow*

Diversas tecnologias de software procuram incrementar a interação entre usuários de forma que estes tenham algum tipo de vantagem na utilização do computador para a realização de seus trabalhos em conjunto. Alguns exemplos são:

- Editores de texto e planilhas para uso em grupos (de forma síncrona ou assíncrona).
- Vídeo-Conferência.
- *Internet* (contemplando *web*, correio eletrônico, listas de discussão, fóruns etc).
- Ferramentas CASE (*Computer Aided Software Engineering*).
- Sistemas de *Workflow*.

Este trabalho se concentra na utilização da tecnologia de sistemas de *workflow* para promover colaboração e reuso entre cientistas. Devido a esta razão, esta classe de sistemas será melhor examinada.

Segundo GEORGAKOPOULOS, HORNICK et al. (1995), *workflow* é uma coleção de tarefas organizadas para realizar um processo de negócios. Uma tarefa pode ser desempenhada por um ou um grupo de indivíduos, um ou um grupo de sistemas, ou até mesmo uma combinação destes. As tarefas desempenhadas por humanos envolvem uma interação com os computadores que pode ser forte (quando a tarefa é realmente realizada com o auxílio do computador) ou fraca (quando somente o andamento e conclusão da tarefa são informados ao computador). Exemplos de tarefa incluem a atualização de um arquivo ou banco de dados, a geração e envio por correio eletrônico de uma nota de pagamento, e a colocação de estruturas de cabos para a montagem de uma rede. O *workflow* ainda define a ordem da invocação das tarefas e eventuais condições que elas apresentam para execução, o sincronismo que pode vir a existir entre as tarefas, e o fluxo de dados ou informações (denominado *dataflow*) que é inerente ao processo.

A WFMC (*Workflow Management Coalition*), organização internacional que reúne vendedores, usuários, pesquisadores e consultores da área de *workflow*, define esta tecnologia como sendo a automação de um processo de negócios, completa ou

parcialmente, na qual documentos, informações e tarefas são passadas de um participante a outro de acordo com uma série de procedimentos descritos em regras (WFMC, 1999).

PLESUMS (2002) define *workflow* através de exemplos e argumenta que nos últimos quinze anos a tecnologia criou não somente ferramentas que auxiliem no trabalho cotidiano ou até mesmo o façam, mas também que gerenciem o fluxo de trabalho existente: “o processo de trabalho é gerenciado por um programa de computador que delega tarefas, as passa adiante, e rastreia seu progresso.”

Sistemas gerenciadores de *workflow* nasceram a partir da concepção destas idéias e trouxeram consigo algumas vantagens importantes para a realização do trabalho no dia a dia:

- Um trabalho dificilmente é perdido ou protelado, uma vez que tarefas posteriores dependem da eficiente realização das anteriores.
- Os gerentes podem se concentrar em assuntos empresariais de mais alto nível, como desempenho individual, melhores práticas, e casos especiais; ao invés da atribuição rotineira de tarefas.
- Os procedimentos são documentados formalmente e seguidos exatamente. Assegura-se que o trabalho é executado do modo planejado pela administração, satisfazendo todas as exigências empresariais.
- Os profissionais e máquinas mais adequados são selecionados para cada tarefa, e as tarefas mais importantes têm prioridade sobre as demais. Colaboradores não desperdiçam seu tempo decidindo qual tarefa devem realizar, porém às vezes se atrasam nas mais complexas.

Sistemas de *workflow* se inserem no contexto geral de software cujo objetivo é o suporte ao trabalho cooperativo (envolve objetivos bem definidos), em que se enfatiza a interação entre usuários, e não apenas a interação usuário e sistema. O que torna a área mais interessante é a grande variação apresentada na prática pelos processos, devido a diferenças de cultura empresarial, a existência de casos especiais e a evolução natural dos processos no decorrer do tempo.

2.2.3 Classificação de *Workflows*

Uma classificação para *workflows* bastante aceita na literatura, divide os fluxos de trabalho em três tipos básicos: *ad-hoc*, administrativo e produção (GEORGAKOPOULOS, HORNICK et al., 1995; PLESUMS, 2002).

2.2.3.1 *Workflows ad-hoc*

Também chamado de colaborativos, estes fluxos de trabalho executam processos de negócios, tais como documentação ou venda de produtos, em que não existe um padrão pré-determinado de fluxo de informação entre os participantes. Um novo *workflow* tende a ser definido para cada uso.

Tarefas do tipo *ad-hoc* envolvem coordenação humana, colaboração ou a tomada de decisão em conjunto. A ordenação e a coordenação de tarefas em um *workflow* do tipo *ad-hoc* podem não ser automatizadas, e simplesmente controladas por humanos. Esta classe de *workflow* tipicamente envolve pequenos grupos de profissionais e tem a intenção de apoiar pequenas atividades que requerem uma solução rápida. Como os fluxos não são tão estruturados, não existe razão para uma interface de desenho. Os produtos *groupware* encaixam-se neste modelo.

2.2.3.2 *Workflows Administrativos*

Envolvem processos estruturados, com repetições, e que têm tarefas com regras de coordenação simples, tal como roteamento de um relatório de despesa ou requisição de viagem. A ordenação e coordenação de tarefas em *workflows* administrativos podem ser automatizadas. Esta classe de *workflow* não contempla um processamento complexo de informações e normalmente não requer acesso a sistemas de informação que auxiliem na execução das tarefas.

2.2.3.3 *Workflows de Produção*

Envolvem processos de negócios repetitivos e previsíveis, tais como empréstimos e seguros. Diferente dos administrativos, os de produção englobam um processamento de informações complexas envolvendo acesso a múltiplos sistemas de informação ou bancos de dados para o desempenho de suas tarefas. A ordenação e coordenação de tarefas neste tipo de *workflow* podem ser automatizadas. Entretanto, esta automatização apresenta um fator de complexidade que é lidar com a diversidade

de informações no processo, e acesso a múltiplos sistemas de informação para execução do trabalho e para a recuperação de dados para tomada de decisão.

De uma forma geral, os *workflows* de produção provêm um controle amplo de um processo pré-definido e uma substancial análise de produtividade.

2.2.4 Funcionamento de *Workflows*

De uma forma geral, os sistemas de *workflow* tendem a ter um funcionamento semelhante para que os seus objetivos possam ser alcançados. Alguns termos inerentes à área surgem e se relacionam para que seja possível um entendimento completo da eficiência dos sistemas que implementam esta tecnologia.

A WPMC (1999) propõe um esquema de relacionamento entre os termos existentes (Figura 2.5).

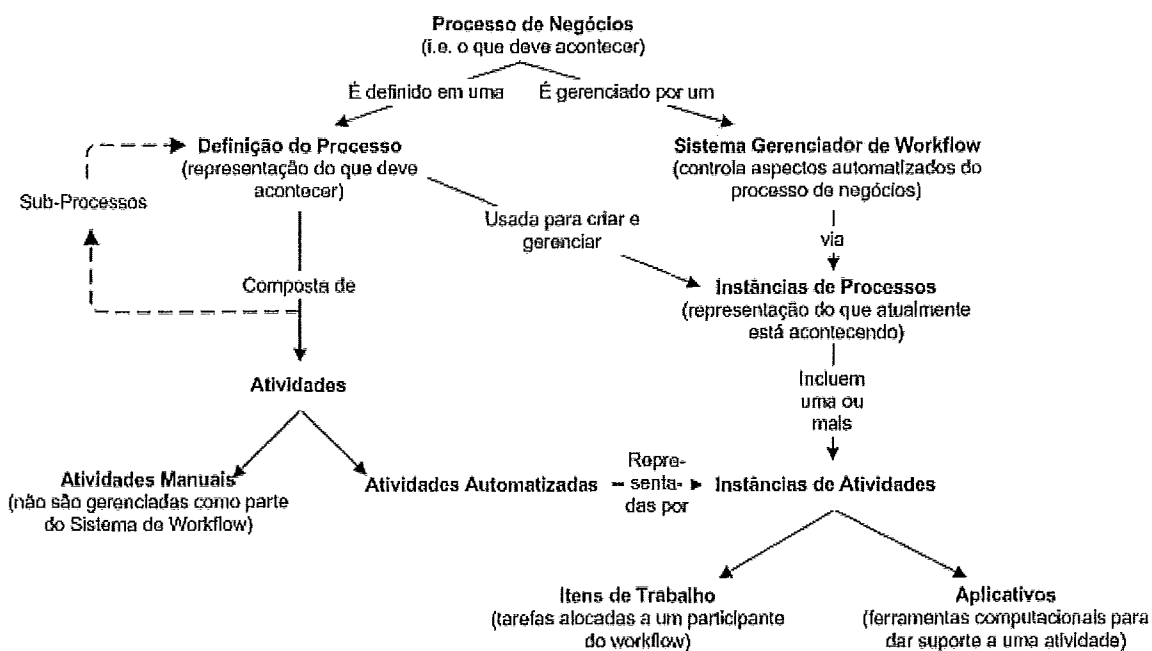


Figura 2.5 - Relacionamentos entre a terminologia básica de *workflows* (WPMC, 1999)

Uma crítica relevante ao esquema de relacionamentos exposto na Figura 2.5 é o aspecto das tarefas não automatizadas ou manuais. Embora não diretamente, esta categoria de atividades pode ser gerenciada em um sistema apropriado, desde que exista suporte à informação de início, andamento e conclusão de cada instância de atividade.

A realização de diversas instâncias do mesmo processo (definido apenas uma vez) atesta a capacidade de monitoramento de eficiência e produtividade que são alcançados com a utilização de sistemas *workflow*.

2.2.5 Workflows Científicos

Nas seções anteriores, um conjunto de funcionalidades interessantes foi abordado a respeito de sistemas de *workflow*. Em um nível de abstração bem alto, estes sistemas controlam atividades de um fluxo de trabalho inseridas em um processo. Aproveitando-se desta característica torna-se possível a associação da tecnologia *workflow* não somente ao trabalho de escritório ou de negócios, mas também para a realização de experimentos científicos.

No contexto da definição de um fluxo de tarefas, tanto os trabalhos de escritório como os científicos envolvem a execução de atividades, algumas manuais e outras automatizadas, o relacionamento entre elas, e o fluxo de dados que trafega entre um passo e outro.

No momento em que é preciso trabalhar com um nível de abstração mais baixo, o trabalho de escritório e o científico apresentam sérias diferenças. WAINER, WESKE et al. (1996) tratam primordialmente das diferenças de objetivos existentes nos dois tipos de trabalho.

Segundo os autores, os objetivos do trabalho de escritório devem ser alcançados, não importando a maneira como isso é feito. A existência de mecanismos de tratamento de exceções e de planejamento ad-hoc em sistemas de *workflow* destinados ao trabalho de escritório atestam que a flexibilidade deve ser incentivada para que o trabalho fique pronto de forma mais eficiente.

No trabalho científico, normalmente os objetivos são a coleta, geração e análise de dados. A obtenção e junção de dados advindos de vários experimentos, a geração de dados a partir de modelos matemáticos, e a análise estatística sobre conjuntos previamente gerados são atividades que podem ser perfeitamente adequadas em sistemas de *workflow* (WAINER, WESKE et al., 1996).

Conclui-se que a principal motivação do uso da tecnologia *workflow* no ambiente de trabalho científico, além de melhorar a eficiência dos experimentos, é

tornar disponível a informação de como os experimentos foram conduzidos, quais dados foram gerados, e quais conclusões foram obtidas.

Além disso, o grau de flexibilidade que os cientistas necessitam é bem maior do que o encontrado no trabalho de escritório. Enquanto no escritório um gerente pode permitir que um crediário seja aberto sem a análise de crédito (pulando uma etapa importante para atingir um grau de eficiência maior), no trabalho científico um pesquisador pode resolver filtrar um conjunto de dados através de algum parâmetro mesmo que a atividade de filtragem não tenha sido prevista anteriormente (simplesmente colocando os dados filtrados figurando como advindos de um dispositivo). Em outras palavras, é importante o rastreamento de cada execução de um trabalho científico, para que seja possível a correta identificação dos dados gerados, e as origens dos resultados (*data lineage*).

Uma outra diferença relevante entre os dois tipos de *workflows* é a necessidade da previsão de *workflows* parciais no trabalho científico. Ocasionalmente, um experimento ou protocolo científico não está totalmente definido no momento de seu início, sendo necessária a criação de novas atividades (ou a remoção delas) na medida em que o trabalho vai sendo realizado. Um sistema eficiente de *workflows* científicos deve permitir a definição de fluxos de trabalho parciais, de forma que eles possam sofrer manutenções mesmo depois de instanciados.

O tratamento de insucessos no ambiente científico serve como um mecanismo para que se evitem futuros erros. No ambiente de escritório, um insucesso pode ser fruto de um caso isolado e nem sempre é importante. Finalmente, a tarefa de definição de *workflow* nos dois ambientes é delegada a categorias distintas: os próprios usuários a fazem no meio científico enquanto que no ambiente de escritório um administrador ou um gerente será o responsável. A Tabela 2.3 resume as principais diferenças encontradas nos dois tipos de *workflows*.

Tabela 2.3 - Quadro comparativo entre *workflows* científicos e de escritório

	Objetivos	Flexibilidade	Definição Parcial	Tratamento de Insucessos	Responsável pela Definição
Workflow Científico	Eficiência na geração de resultados e disponibilidade de informação de condução, geração e conclusão	Indispensável	Aplicável	Evitam erros futuros	Próprio usuário
Workflow de Escritório	Eficiência na realização do trabalho	Desejável	Não Aplicável	Nem sempre útil	Administradores ou Gerentes

A noção de gerência de processos, introduzida nos sistemas de *workflow*, quando aplicada ao trabalho científico, tradicionalmente associado a experimentos (MEDEIROS, VOSSEN et al., 1995; WESKE, VOSSEN et al., 1996), consegue trazer a idéia de se estabelecer uma associação entre o ciclo de vida de experimentos, definido na seção 2.1.2.1, e a tecnologia *workflow*:

1. **Projeto do Experimento (*Experiment Design*)** – é apoiado por ferramentas, normalmente gráficas, de definição de *workflows*. Tais ferramentas dão uma maior facilidade e flexibilidade para o projetista do experimento científico, permitindo que se consultem definições de dados, modelos, sub-experimentos, e outros artefatos científicos.
2. **Planejamento do Experimento (*Experiment Planning*)** – consiste no planejamento da instanciação do *workflow*. É nesta etapa que os bancos de dados são consultados para se conhecer a disponibilidade dos dados iniciais necessários para o experimento ser executado. Dados que não estão disponíveis devem ser obtidos.
3. **Coleta de Dados (*Data Collection*)** – é a instanciação do *workflow*, com a geração de resultados experimentais. O sistema de *workflow* deve invocar as etapas na seqüência correta e ir passando dados intermediários de uma atividade para outra, até que se atinja o resultado experimental desejado.

4. **Análise de Dados (*Data Exploration*)** – como citado nos objetivos acima, a análise dos dados também deve estar presente após a instanciação efetiva do *workflow*. O sistema deve acoplar as conclusões aos dados gerados para que os resultados experimentais sejam aperfeiçoados e reproduzidos.

Outros trabalhos também buscam a correlação entre a realização de experimentos científicos e o ciclo de vida do trabalho científico (AILAMAKI, IOANNIDIS et al., 1998; WESKE, GOTTFRIED et al., 1998).

A ferramenta BOE disponibiliza um assistente gráfico para a definição de *workflows* científicos e auxilia o pesquisador no planejamento do experimento quando os dados base não estão totalmente disponíveis para uso. Como já citado na seção 2.1.2.2, o BOE não atua nas etapas 3 e 4 do ciclo de vida de experimentos quando este é associado a tecnologia *workflow*.

No capítulo 4, encontram-se exemplos de trabalhos científicos descritos como *workflows*, dentre eles destacam-se o zoneamento agro-climático, como exemplo na agricultura, e a citometria de fluxo⁴, na área biológica.

2.2.6 Trabalhos Relacionados

A arquitetura WASA (*Workflow-based Architecture to support Scientific Applications*) (MEDEIROS, VOSSSEN et al., 1995; WESKE, GOTTFRIED et al., 1998; WESKE, VOSSSEN et al., 1996) propõe uma integração da tecnologia de bancos de dados com ferramentas existentes no ambiente científico, de forma que seja possível dar suporte a realização de experimentos científicos representados através de *workflows*. A arquitetura WASA é composta por quatro camadas: interface, ferramentas internas, serviços de banco de dados e banco de dados.

A camada de interface é responsável pela comunicação dos usuários com o sistema. Ela é subdividida em quatro módulos que permitem a especificação e reuso de *workflows* científicos. Na camada de ferramentas internas reside o núcleo da arquitetura. Um sistema gerenciador de *workflows* é acompanhado de módulos de documentação, análise e planejamento. A camada de serviços de bancos de dados provê

⁴ Termo conhecido em inglês como “*flow cytometry*”.

funcionalidades não existentes em sistemas gerenciadores de bancos de dados tidos como “legados”, por exemplo: características temporais, versionamento e regras ativas. Finalmente, a camada de banco de dados representa as definições de *workflows* e dados específicos dos domínios de aplicação de interesse.

O trabalho de AILAMAKI, IOANNIDIS et al. (1998) estabelece uma proposta de utilização do próprio banco de dados científicos para que se controle a execução de um *workflow* científico. Tanto as definições como as informações de execução são apoiados pelo ambiente ZOO (IOANNIDIS, LIVNY et al., 1996) e residem no banco de dados, permitindo que se abra mão de um software específico de controle de processos. O estudo é ilustrado com um exemplo prático, o *cranberry*⁵ *workflow*. Em trabalho preliminar (SOUZA, CARDOSO et al., 2002), a ferramenta BOE se mostrou capaz de lidar com o exemplo deste trabalho, demonstrando melhores alternativas para a realização hipotética do *cranberry workflow*.

Com relação à organização dos bancos de dados científicos presentes nos sistemas de *workflow* apresentados, é possível encontrar duas abordagens distintas:

1. **Bancos de Dados Específicos** – geram um esquema de dados para cada *workflow* que necessita ser representado. Nesta abordagem um possível controle de execução pode ficar no banco de dados, simplificando a tarefa de gerenciamento de execuções de experimentos. A desvantagem é em relação à flexibilidade, pois uma inclusão de uma atividade leva a criação de novas definições de objetos dentro do banco de dados (tabelas caso o modelo relacional seja utilizado). São exemplos (AILAMAKI, IOANNIDIS et al., 1998; CHEN & MARKOWITZ, 1995a).
2. **Banco de Dados Geral** – apenas um único esquema existe e deve ser capaz de representar todos os *workflows* existentes. Esta abordagem é bastante robusta e suporta bem as alterações nas definições pré-existentes. Por outro lado, é necessária a existência de uma camada de software que gerencie o andamento da

⁵ Fruta silvestre muito popular nos EUA, conhecida em português como “oxicoco”.

execução. São exemplos (CARDOSO, SOUZA et al., 2002a; PALMA, 2002; WESKE, VOSSEN et al., 1996).

Utilizando um banco de dados geral, a ferramenta BOE é capaz de atingir os requisitos de flexibilidade e definição parcial exigidos para a descrição dos *workflows* científicos. O esquema geral foi definido no ambiente SPeCS (PALMA, 2002).

2.3 Gestão do Conhecimento

Esta seção irá apresentar alguns conceitos relacionados à área de Gestão do Conhecimento. Este trabalho utiliza conceitos advindos da área em questão para embasar sua discussão a respeito do reuso e explicitação de *workflows* e artefatos científicos. A partir da definição gradual de termos pertinentes à disciplina, será possível entender em que ponto a ferramenta BOE dá apoio tecnológico à gestão do conhecimento científico.

2.3.1 Conhecimento

Um bom ponto de partida para a discussão a respeito do tema gestão do conhecimento é o entendimento do significado do termo **conhecimento**. O que é conhecimento? Embora o problema esteja em estudo há séculos, infelizmente não existe uma resposta amplamente aceitável para esta pergunta. As definições dadas pelos autores que se aventuraram no tema foram sendo suplantadas uma após a outra ao longo da história, e mesmo a definição usual é contestada. Os próximos parágrafos apresentam uma breve discussão a respeito deste termo.

A primeira idéia para a busca da definição do termo conhecimento é a consulta ao dicionário. A lista é extensa, mas consegue ser extremamente esclarecedora. No dicionário Houaiss, a palavra conhecimento apresenta as definições abaixo. “*Substantivo masculino. Ato ou efeito de conhecer. O ato ou a atividade de conhecer, realizado por meio da razão e / ou da experiência. Ato ou efeito de apreender intelectualmente, de perceber um fato ou uma verdade; cognição, percepção. Domínio, teórico ou prático, de um assunto, uma arte, uma ciência, uma técnica etc.; competência, experiência, prática.*” (HOUAISS, 2001).

Em sua tese de doutorado, o cientista da computação Lethbridge dá uma definição bastante simplificada para o conceito de conhecimento. Em geral, qualquer

formato de informação que uma pessoa é capaz de manipular no seu cérebro é considerado conhecimento (LETHBRIDGE, 1994).

NETO (2002) apresenta um atualizado trabalho a respeito de definições do termo conhecimento. O objetivo do estudo, exclusivamente bibliográfico, foi propor ou axiomatizar uma nova definição ou conceito de conhecimento.

O estudo dos fundamentos filosóficos do conhecimento é conhecido como “epistemologia” ou Teoria do Conhecimento. Suas questões principais são: O que é conhecimento? Como ele é alcançado?

BARROS & LEHFELD (1986) definem o conhecimento como sendo "*a manifestação da consciência-de-conhecer; é a consciência de conhecimento*". As duas autoras apresentam uma interpretação do conhecimento como proveniente das experiências vividas através das percepções dos seus sentidos. Para elas é através do conhecimento que se faz possível ao homem adaptar-se, interpretar e assimilar o mundo exterior e o seu interior.

Para NONAKA & TAKEUCHI (1995) a resposta para a pergunta “o que é conhecimento?” guia a história da filosofia desde o período grego. Esta história, da epistemologia ocidental, divide-se em duas correntes opostas, uma racionalista e outra empirista, consideradas estas complementares pelos autores acima citados.

- Racionalismo → afirma ser possível adquirir conhecimento por dedução, através do raciocínio.
- Empirismo → afirma ser possível adquirir conhecimento por indução, a partir de experiências sensoriais.

Segundo NONAKA & TAKEUCHI (1995), os racionalistas argumentam que o verdadeiro conhecimento não é produto da experiência sensorial, mas sim de um processo mental ideal. Segundo essa visão, existe um conhecimento *a priori* que não precisa ser justificado pela experiência sensorial. A verdade absoluta é deduzida a partir de uma argumentação racional baseada em axiomas. Para os racionalistas o paradigma de conhecimento é a matemática e a lógica, em que verdades necessárias são obtidas por intuição e inferências racionais.

Por outro lado, os empiristas alegam não existir conhecimento *a priori* e que a única fonte de conhecimento é a experiência sensorial. Por esta visão, todas as pessoas

têm uma existência intrinsecamente objetiva, mesmo quando se tem uma percepção ilusória. O próprio fato de se perceber alguma coisa já é significativo. O paradigma desta escola é a ciência natural, onde as observações e os experimentos são cruciais à investigação.

Por exemplo, imagine que uma criança pequena foi passar o dia na casa de sua tia e lá, devido a um descuido do responsável, queimou a mão em um ferro quente de passar roupas. Ao chegar em casa, sua mãe pergunta à criança se a estadia na casa da tia foi agradável e o que a criança aprendeu durante o período. A criança responde que não aprendeu nada interessante, mas provavelmente não mais irá tentar novas brincadeiras com ferros quentes.

Na filosofia contemporânea, a definição usual tem o conhecimento como sendo “crença verdadeira justificada”. Apesar de ser aceito, de um modo geral, pelos filósofos ocidentais, este conceito foi introduzido primeiramente por Platão. GRAYLING (1996) chama esta definição de “*definição padrão*”, enquanto NONAKA & TAKEUCHI (1995) referem-se a ela como sendo a “*definição tradicional de conhecimento*”.

Para GRAYLING (1996) “*cada uma das três partes da definição parece expressar uma condição necessária para o conhecimento, e a reivindicação é a de que, tomadas em conjunto, elas são suficientes*”. Segundo ele, esta definição parece aceitável pelo menos no que diz respeito ao conhecimento empírico.

Esta definição está embasada na ocorrência de três condições necessárias para que o sujeito (A) tenha conhecimento do objeto, que aqui passa a ser tomado como proposição (P). Estas condições são: (a) P é verdadeiro (condição de verdade); (b) A precisa acreditar que P é verdadeiro (condição de crença) e (c) a crença de A em P verdadeiro necessita ser justificada (condição de justificação).

Para se chegar a uma definição formal do conhecimento, é importante que se tenha clara a distinção entre dado, informação, conhecimento e ação. Estes quatro elementos juntos constituem uma “escala de valor” ascendente que vai desde o dado até a ação, passando pela informação e pelo conhecimento (SANTOS & CERANTE, 2000). Tal escala de valor proposta em (DAVENPORT & PRUSAK, 1998) e ilustrada em (SANTOS & CERANTE, 2000) se encontra na Figura 2.6.

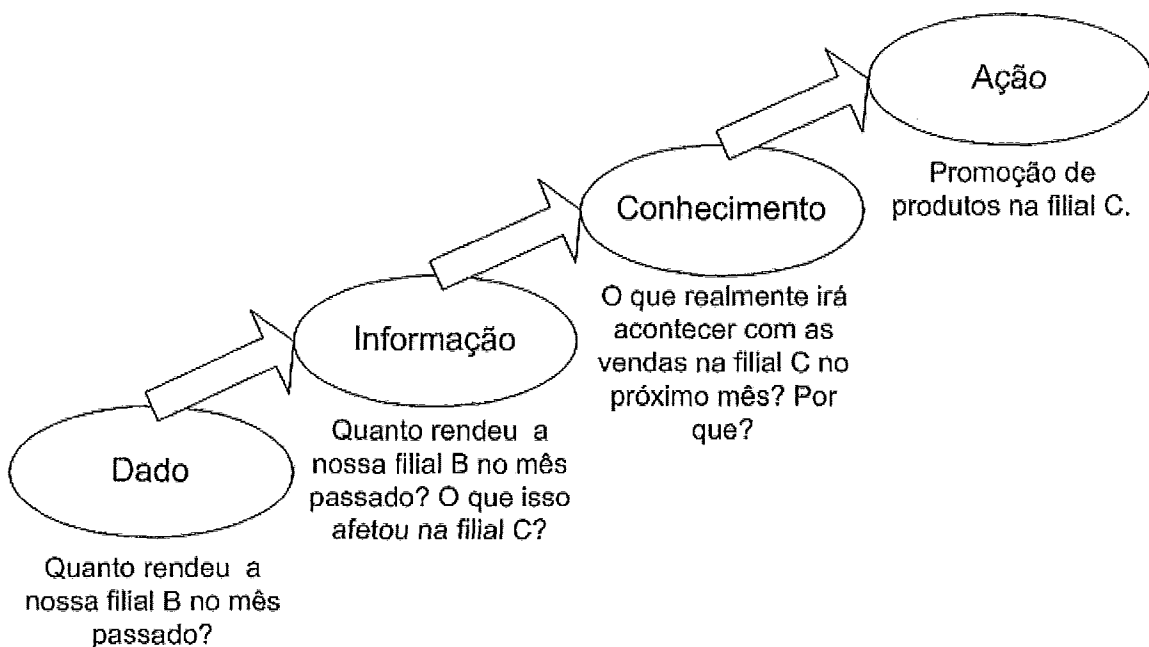


Figura 2.6 - Escala de valores de dado, informação, conhecimento e ação (SANTOS & CERANTE, 2000) – modificado.

Os dados “são um conjunto de fatos distintos e objetivos, relativos a eventos” (DAVENPORT & PRUSAK, 1998) que constituem a base para a criação da informação, sendo obtidos, por exemplo, de registros ou transações. Eles não possuem significados inerentes, descrevem apenas parte do acontecido e não fornecem condições para qualquer julgamento, interpretação ou base para tomada de ação.

As informações são “dados dotados de relevância e propósito” (DAVENPORT & PRUSAK, 1998), constituindo-se em um fluxo de mensagens que tem por objetivo exercer algum impacto sobre o julgamento e o comportamento do seu receptor.

Segundo DAVENPORT & PRUSAK (1998) o conhecimento pode ser definido como: “...uma mistura fluida de experiência condensada, valores, informação contextual e insight experimentado, a qual proporciona uma estrutura para a avaliação e incorporação de novas experiências e informações.”

O conhecimento possui um valor maior que o dado e a informação estando mais próximo da ação nesta escala podendo embutir os conceitos de “sabedoria” e “insight”. Está presente no indivíduo, nos grupos ou nas rotinas organizacionais sendo desenvolvido através da “experiência” e possibilitando, assim, uma definição de um

histórico que permite o reconhecimento de padrões para uma tomada de ação mais rápida. O conhecimento possui uma “verdade fundamental” que está relacionada com a identificação, através de experiências reais, do que realmente funciona ou não. Estes dois fatores citados, experiência e verdade fundamental, auxiliam o tratamento de assuntos complexos.

Tabela 2.4 - Características de dado, informação e conhecimento SANTOS & CERANTE apud (DAVENPORT & PRUSAK, 1998)

Dado	Informação	Conhecimento
<ul style="list-style-type: none"> • Simples observações sobre o estado do mundo • Facilmente estruturado • Facilmente obtido por máquinas • Frequentemente quantificado • Facilmente transferível 	<ul style="list-style-type: none"> • Dados dotados de relevância e propósito • Requer unidade de análise • Exige consenso em relação ao significado • Exige necessariamente a mediação humana 	<ul style="list-style-type: none"> • Informação valiosa na mente humana • Inclui reflexão, síntese, contexto • De difícil estruturação • De difícil captura em máquinas • De difícil transferência • Frequentemente tácito

Outro fator característico, é que o conhecimento possui “discernimento”. Isto quer dizer que ele é capaz de avaliar novas situações tendo como base o que já é conhecido, podendo fornecer julgamentos que possibilitam seu próprio aprimoramento e conseqüente evolução. Atua seguindo normas práticas, desenvolvidas por meio da tentativa e erro, que enxergam padrões e permitem aumento da velocidade na solução de problemas. Por fim, como foi dito anteriormente, o conhecimento opera segundo valores e crenças, pois reside nas pessoas e estas, portanto, podem ter diferentes interpretações para uma mesma situação. Um resumo das características apresentadas nos últimos parágrafos está esquematizado na Tabela 2.4.

E por último, a ação, que resumidamente pode ser considerada “algo que se faz com o conhecimento” (DAVENPORT & PRUSAK, 1998) e está na mais alta “posição” desta escala.

Não é objetivo deste trabalho esgotar o assunto ou estabelecer uma profunda discussão filosófica a respeito do termo conhecimento. A partir de agora o conhecimento será tratado de forma que este se materializa na cabeça das pessoas, e estas podem estabelecer interpretações de maneiras diferenciadas.

2.3.2 Tipos de Conhecimento

Diversos autores embasaram as suas definições de conhecimento através de tipificações, ou seja, subdivisões em níveis. Alguns destes tipos de conhecimento, existentes na literatura serão apresentados a seguir. Primeiramente discutem-se quatro tipos principais de conhecimento (LAKATOS & MARCONI, 1991): popular, religioso, filosófico e científico; tendo este último maior destaque neste trabalho. Em seguida, uma outra dimensão com relação a tipos de conhecimento, importante para este trabalho, e que se enquadra nestas quatro categorias, é discutida.

2.3.2.1 Conhecimento Popular

Também denominado conhecimento vulgar, conhecimento sensível (senso comum), ou ainda empírico, provém da experiência do dia a dia, fruto do acaso, obtido através de, segundo CERVO & BERVIAN (1996), "*... investigações pessoais feitas ao sabor das circunstâncias da vida ou então sorvido do saber dos outros e das tradições da coletividade ou, ainda, tirado da doutrina de uma religião positiva.*"

O conhecimento vulgar é o modo comumente encontrado de conhecimento, adquirido no trato direto com objetos e os seres humanos (NETO, 2002).

2.3.2.2 Conhecimento Religioso

É o conhecimento revelado, aceito pela fé teológica. Este conhecimento é apoiado em doutrinas de proposições sagradas e direcionado à compreensão do mundo em sua totalidade. O conhecimento religioso caracteriza-se por ser valorativo, advindo de inspiração, sistemático, não verificável, falível e aproximadamente exato (LAKATOS & MARCONI, 1991).

2.3.2.3 Conhecimento Filosófico

A filosofia é apoiada basicamente na reflexão, suas hipóteses advêm da experiência e não da experimentação. Esta por sua vez fornece um amplo leque de interpretações e por conseguinte impressões, opiniões e conclusões diversas.

O conhecimento filosófico pode ser caracterizado como valorativo, racional, sistemático, não verificável, infalível e exato (LAKATOS & MARCONI, 1991). A filosofia está dividida em dois grandes grupos, a saber: prática ou normativa e especulativa ou teórica.

2.3.2.4 Conhecimento Científico

O conhecimento científico é o que é produzido pela investigação científica, através de seus métodos. Surge não apenas da necessidade de encontrar soluções para problemas de ordem prática da vida diária, mas do desejo de fornecer explicações sistemáticas que possam ser testadas e criticadas através de provas empíricas.

A investigação científica se inicia quando se descobre que os conhecimentos existentes, originários do senso comum ou do corpo de conhecimentos existentes na ciência, são insuficientes para explicar os problemas surgidos. O conhecimento prévio que nos lança a um problema pode ser tanto ordinário quanto científico.

O conhecimento científico transpõe o empírico, procurando conhecer não só um fenômeno, mas suas relações de causa e efeito. Pressupõe um ou mais problemas a serem resolvidos, ou hipóteses a serem confirmadas, através de processos de pesquisa norteados por métodos. A própria definição de ciência se confunde com conhecimento, como descrito em (HOUAISS, 2001): *“conjunto de conhecimentos teóricos, práticos ou técnicos voltados para determinado ramo de atividades.”* Além disso, o conhecimento científico deve ser aceito pela comunidade científica antes de se tornar efetivo.

ANDER-EGG apud LAKATOS & MARCONI (1991) conceitua a ciência como *“um conjunto de conhecimentos racionais, certos ou prováveis, obtidos metodicamente sistematizados e verificáveis, que fazem referência a objetos de uma mesma natureza.”*

Atualmente, a ciência não possui mais a posse de tudo aquilo que é considerado verdade, isto é, não se considera mais como algo definitivo, pronto ou acabado, mas sim em constante contestação e aperfeiçoamento. A ciência é composta de objetivo ou finalidade, função e objeto formal ou material. São características do conhecimento científico, o fato dele o ser, real, contingente, sistemático, verificável, falível e aproximadamente exato (LAKATOS & MARCONI, 1991).

É importante enfatizar que o conhecimento científico e os processos através dos quais este conhecimento é desenvolvido são essenciais para o aprendizado científico. O processo científico, conforme o nome esclarece, é aquele utilizado em relação ao assunto ciência. Na definição anterior, o termo conhecimento científico é utilizado com significado mais amplo do que um mero conhecimento de fatos, nomes,

condições e termos científicos. Ela inclui entendimento de conceitos científicos fundamentais, as limitações de conhecimento científico e a natureza de ciência como uma atividade humana.

2.3.2.5 Conhecimento Tácito e Explícito

POLANYI (1967) estabeleceu uma classificação do conhecimento humano em dois grupos básicos: tácito e explícito. NONAKA & TAKEUCHI (1995) se basearam nos estudos de POLANYI e desenvolveram definições que são amplamente aceitas e referenciadas na literatura.

O conhecimento tácito é pessoal, específico ao contexto e, desta forma, difícil de ser formulado e comunicado (NONAKA & TAKEUCHI, 1995). Como exemplo de conhecimento tácito, pode-se descrever a habilidade do jogador Pelé para a cobrança de uma penalidade máxima em um jogo de futebol. O conhecimento de Pelé a respeito da cobrança é pessoal, pois só o Pelé bate pênalti como o Pelé. Também é totalmente enquadrado num contexto, que consiste na partida de futebol. Finalmente, será bastante difícil para qualquer pessoa que nunca bateu um pênalti na vida passar a cobrá-los como o Pelé, mesmo que o próprio esteja tentando passar o seu conhecimento.

O conhecimento explícito ou codificado refere-se ao conhecimento transmissível em linguagem formal e sistemática (NONAKA & TAKEUCHI, 1995). É fácil de ser comunicado e compartilhado pois pode ser expresso em palavras, números, dados brutos, fórmulas científicas, procedimentos codificados ou princípios universais. Documentos, manuais, bancos de dados, métodos, modelos e diagramas são alguns exemplos de conhecimento explícito.

Dadas as duas dimensões de conhecimento descritas acima, uma proposta de interseção de ambas está esquematizada na Tabela 2.5.

Pelas suas características de casualidade, experiência do dia a dia e empirismo, o conhecimento popular é mais frequentemente encontrado na forma tácita. Já o conhecimento religioso, apoiado em doutrinas, se apresenta primordialmente na forma explícita, como se pode comprovar com os livros bíblicos. O conhecimento filosófico frequentemente se dá em ambas as dimensões, visto as separações inerentes à própria área filosófica (especulativa ou teórica e prática o normativa). Finalmente, o conhecimento científico, tem características tácitas muito importantes, tais como: idéias

ainda não propostas em trabalhos, correlações de conceitos e intuições em investigações. Quando um conhecimento científico criado é aceito pela comunidade, este frequentemente se torna explícito através de publicações em jornais, revistas, congressos, relatórios e outros.

Tabela 2.5 - Interseções das dimensões de tipificações de conhecimento

Conhecimentos	Popular	Religioso	Filosófico	Científico
Tácito	Muito Freqüente	Pouco Freqüente	Freqüente	Freqüente
Explícito	Pouco Freqüente	Muito Freqüente	Freqüente	Freqüente

2.3.3 O Processo de Gestão de Conhecimento

O conceito de gestão do conhecimento surgiu no início da década de 90, principalmente no meio empresarial. Alguns autores chegaram a utilizar o termo em estudos técnicos na área de banco de dados (STONEBRAKER, ROWE et al., 1990) (vide seção 2.3.4), mas o termo foi consagrado no âmbito empresarial, com correntes da teorias de administração e de tecnologia da informação.

A grande competição do mundo que se tornava globalizado, a informatização, a necessidade de diferenciais competitivos e, ao mesmo tempo, o desapontamento com modismos e soluções ineficientes fez com que as empresas percebessem a importância do conhecimento e da sua gestão.

A importância da informação para as organizações é universalmente aceita, constituindo, senão o mais importante, pelo menos um dos recursos cuja gestão e aproveitamento estão diretamente relacionados ao sucesso desejado (MORESI, 2001). A informação pode ser classificada segundo a sua finalidade para uma organização, conforme relata MORESI (2001), em **sem interesse** (sem utilidade), **potencial** (acarreta vantagem competitiva), **mínima** (fundamental para a gestão da organização) e **crítica** (essencial para a sobrevivência da organização). Os esforços principais de uma organização devem priorizar a busca e manutenção de informação crítica, mínima e potencial, respectivamente. É preciso evitar desperdício de recursos para a obtenção da informação sem interesse.

DAVENPORT & PRUSAK (1998) afirmam que neste novo contexto de negócios as organizações reconhecem que o conhecimento é a única fonte capaz de

gerar uma vantagem competitiva sustentável. Os autores constatam que atividades baseadas no conhecimento, como o desenvolvimento de novos processos e produtos, estão se tornando primordiais para as empresas e que estas estão se diferenciando umas das outras pelo que sabem.

Este fato vem tornando-se mais notável na medida em que: as novas possibilidades técnicas e o conhecimento de mercados determinam as inovações nos produtos; operações funcionais advêm do conhecimento combinado entre como as coisas funcionam e como poderiam funcionar e a participação no mercado cresce com um melhor conhecimento dos clientes atuais e potenciais e de como melhor atendê-los.

Foi relatado (BARROSO & GOMES, 1999), com base nos estudos de SVEIBY (2001) que a nova fortuna das organizações são os ativos intangíveis. O surgimento desta classe de ativos justifica que algumas empresas tenham suas ações muito mais valorizadas do que o seu valor contábil. No mercado de capitais, ações têm sido negociadas acima de seus valores patrimoniais e este ágio tem crescido muito para certas ações. Em muitas empresas, a importância de seus ativos intangíveis supera o de seus ativos contábeis. Mais ainda, a relação de valor entre os ativos intangíveis e os ativos contábeis tem se tornado cada vez maior.

Todos estes fatos levam a conclusão que o conhecimento se tornou fundamentalmente um diferencial no mercado competitivo e globalizado do século XXI. Entretanto, ainda não é fácil entender como lidar com esta riqueza recentemente descoberta. A área de Gestão do Conhecimento veio exatamente com o propósito de decifrar o conhecimento dentro da organização, principalmente o que existe dentro da cabeça das pessoas, e dele tirar proveito.

Neste ponto é importante rever e discutir definições a respeito do tema Gestão do Conhecimento. A princípio, a definição pode parecer evidente, visto que as pessoas têm um entendimento próprio do que significa “conhecimento” e “gestão”. A próxima seção demonstra que esta tarefa de definição é um pouco mais complexa do que parece.

2.3.4 Definições de Gestão do Conhecimento

STONEBRAKER, ROWE et al. (1990) definiram tecnologicamente a gestão de conhecimento. Na referência citada acima, que descreve a implementação de

um sistema gerenciador de banco de dados relacional-objeto, os autores recomendam que os bancos de dados deveriam passar a oferecer dois tipos serviços que não vinham oferecendo: a gestão de objetos e a gestão de conhecimento. A gestão de objetos serve para o tratamento eficiente de tipos de dados não convencionais, como imagens, textos e polígonos. A gestão do conhecimento seria a capacidade de armazenar e impor uma coleção de regras que são parte da semântica de uma aplicação. Tais regras descrevem restrições de integridade sobre a aplicação além de permitir derivações de dados que não estão diretamente armazenados no banco de dados.

É fácil perceber que o termo gestão de conhecimento utilizado por STONEBRAKER, ROWE et al. é meramente tecnológico e não retrata completamente a área de gestão do conhecimento como esta é entendida atualmente.

Ainda no mundo da ciência da computação, LETHBRIDGE (1994) define a gestão do conhecimento como o processo de aquisição, representação, armazenamento e manipulação de categorizações e caracterizações, bem como definições a respeito de objetos e seus relacionamentos.

Segundo EDWARDS & MAHLING (1997), gestão do conhecimento consiste no esforço organizado para captura, organização e compartilhamento de conhecimento dos empregados de uma organização para o alcance de um objetivo estratégico comum. Neste trabalho, já é perceptível que os cientistas da computação começam a tratar o termo gestão do conhecimento como é utilizado nos dias atuais.

Para DAVENPORT & PRUSAK (1998), qualquer organização que deseje priorizar a gestão do conhecimento deve gerenciar os processos de geração, codificação e transferência do conhecimento.

TERRA (1999) caracteriza o termo “Gestão do Conhecimento” como interdisciplinar e abrangente. É relativamente difícil encontrar um denominador comum ou mesmo estabelecer limites para a forma como os termos conhecimento, competência e habilidade, criatividade, capital intelectual, capital humano, tecnologia, capacidade inovadora, ativos intangíveis e inteligência empresarial, entre outros, são utilizados e definidos na literatura. Esta dificuldade, contudo, ao invés de ser um problema, aponta antes para a riqueza do tema em questão. A Gestão do Conhecimento nas organizações passa, necessariamente, pela compreensão das características e demandas do ambiente

competitivo e, também, pelo entendimento das necessidades individuais e coletivas associadas aos processos de criação e aprendizado (TERRA, 1999).

Segundo LIEBOWITZ (2000) gestão do conhecimento é a ação de criar um processo de valoração dos ativos intangíveis da organização de maneira a melhor alavancar conhecimento interno e externo. A Gestão do Conhecimento é responsável por criar, segurar, capturar, coordenar, combinar, recuperar e distribuir conhecimento.

MORESI (2001) define a gestão do conhecimento argumentando que ela “pode ser vista como o conjunto de atividades que busca desenvolver e controlar todo tipo de conhecimento em uma organização, visando à utilização na consecução de seus objetivos”.

Para SVEIBY (2001), conceitos são bem definidos quando são constituídos a partir das pessoas que os usam. Deste modo, o autor define a gestão de conhecimento através das atividades dos praticantes da área e identifica duas correntes de atividades: i) gestão da informação e ii) gestão de competências, em dois níveis: a) perspectiva individual e b) perspectiva organizacional.

Na **gestão da informação**, o conhecimento é um objeto que pode ser identificado e tratado através de sistemas de informação. Os profissionais que atuam nesta corrente normalmente têm sua formação em ciência da computação e da informação e estão empenhados na construção de sistemas de informações baseados em inteligência artificial, reengenharia, *groupware* etc. Esta visão tem crescido rapidamente no momento, acompanhada pela evolução da tecnologia da informação (TI).

Na **gestão de competências** ou de pessoas, o conhecimento é um processo, um complexo conjunto de habilidades, *know-how* etc, que está em constante evolução. Os profissionais atuantes têm sua formação em filosofia, psicologia, sociologia ou administração de empresas e estão envolvidos em avaliar, modificar e melhorar as habilidades e comportamentos dos indivíduos. Esta abordagem é bastante antiga e não tem apresentado um crescimento muito acentuado.

O nível de perspectiva individual tem o indivíduo como principal foco de pesquisa e prática. No nível de perspectiva organizacional, este foco muda para a organização.

HENDRY (2002) concorda parcialmente com SVEIBY (2001) pois também divide a gestão do conhecimento em duas “gerações”. A primeira geração tem o foco em aplicações tecnológicas e de gestão de documentos. A crítica relevante feita por HENDRY (2002) é que muitas empresas investem dinheiro em sistemas e ferramentas para captura de informações sem qualquer análise ou organização.

A segunda geração utiliza os progressos advindos da primeira, e procura expandir os aspectos sócio-organizacionais do conhecimento, preocupando-se com o aprendizado corporativo. Seu objetivo se aproxima da abordagem de gestão de competências de SVEIBY (2001) e consiste em disseminar conhecimento através da geração de diálogos entre pessoas com conhecimentos diferenciados (por exemplo, criando equipes de múltiplas funções).

Segundo HENDRY (2002), somente a segunda geração da prática da gestão do conhecimento leva a criação de conhecimento efetivamente novo dentro da organização, enquanto que a primeira geração, quando obtém sucesso, apenas alavanca a disseminação do conhecimento já existente.

Pioneiros na gestão do conhecimento na forma como esta é aceita atualmente, NONAKA & TAKEUCHI (1995) definem indiretamente o termo através do sucesso de companhias japonesas justificado pelas habilidades de *know-how* de criação de conhecimento corporativo (*organizational knowledge creation*). Os autores afirmam que uma empresa não é capaz de criar conhecimento por si só, pois necessita da iniciativa de cada indivíduo e a interações de pessoas em grupos.

Ainda segundo NONAKA & TAKEUCHI (1995), a criação de conhecimento organizacional deve ser uma interação contínua e dinâmica entre conhecimentos tácito e explícito, como uma espiral. A criação do conhecimento começa no nível individual e vai subindo, ampliando comunidades de interação que cruzam fronteiras entre seções, departamentos, divisões e organizações. A partir deste fato, o conhecimento humano é criado e expandido através de interações entre conhecimento tácito e explícito, nos “quatro modos de conversão de conhecimento”.

A Figura 2.7 ilustra os quatro modos de conversão do conhecimento: do conhecimento tácito para o conhecimento tácito (socialização); do tácito para o explícito

(externalização); do explícito para o explícito (combinação) e ainda do explícito para o tácito (internalização), atribuindo um caráter cíclico ao processo.

Para implementar a **socialização** deve ser sugerido que os indivíduos tenham interações entre si objetivando o compartilhamento de experiências associadas às emoções, modelos mentais, intenções e visões. Assim, é possível a transferência dos conhecimentos tácitos entre estes indivíduos e a associação de um mesmo tipo de conhecimento a diferentes contextos individuais.

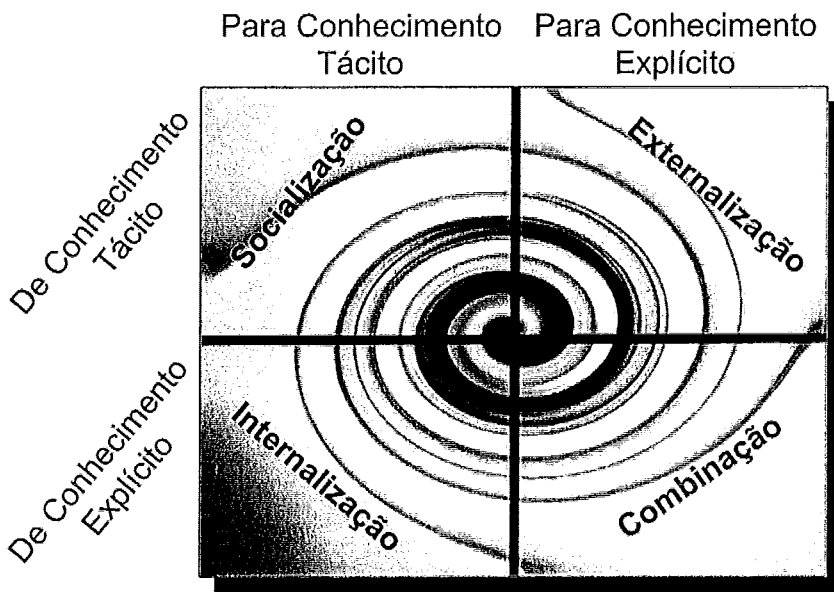


Figura 2.7 - Espiral do Conhecimento

Na **externalização**, o conhecimento tácito manifesta-se em novos conceitos capazes de serem justificados, categorizados e contextualizados uma organização. Somente quando contextualizado coerentemente, o conhecimento subjetivo passa a fazer sentido para a organização e pode vir a se tornar fonte de inovação futura. Normalmente, grandes esforços são aplicados neste modo de conversão, pois existe crescente necessidade de se formalizar o conteúdo abstrato do conhecimento tácito.

O papel da **combinação** é identificar dentre os conceitos que foram extraídos pela externalização, aqueles que possuem alguma relação entre si e agrupá-los em conjuntos de conhecimento explícito. Cada conjunto de conhecimento é parte da base de conhecimento organizacional e está diretamente relacionado a um tipo específico de informação ou modelo.

O que é observado no processo de **internalização** é a captação individual do conhecimento que foi extraído para a organização. É o modo pelo qual o conhecimento explícito se torna ferramenta de aprendizagem através de manuais ou documentos e volta a assumir um contexto abstrato e subjetivo para cada indivíduo na empresa.

A PONTE S.A., empresa que administra a ponte que faz a ligação rodoviária entre as cidades do Rio de Janeiro e Niterói, sobre a Baía de Guanabara (Estado do Rio de Janeiro), realizou um trabalho estratégico de gestão de conhecimento aplicado. Para ilustrar, seguem abaixo, alguns casos que detalham e demonstram como acontecem as conversões do conhecimento nos diversos ambientes da PONTE S.A. (BRAGA, 2002).

A socialização é implementada através do DSO (Diálogo Semanal de Operações). Nestas reuniões de DSO são apresentados novos conhecimentos adquiridos no decorrer de uma semana. A conversão do conhecimento se dá no nível tácito e com o compartilhamento das experiências vividas por cada indivíduo da equipe.

O programa de monitoramento da ponte contribui para o processo de externalização. Através de recurso tecnológico capaz de monitorar todo o trecho da ponte, composto de 20 câmeras, algumas atividades de socorro são gravadas. Estes filmes são usados para a realização de análise de falhas e sucessos dos procedimentos executados. A conversão do conhecimento ocorre quando os procedimentos documentados de salvamento (conhecimento explícito) são alterados em função das experiências adquiridas pelos indivíduos (conhecimento tácito).

A combinação se dá no Serviço de Atendimento ao Cliente. Questões explicitadas pelos usuários da ponte e registradas no "Sistema de Atendimento ao Cliente" são analisadas pelos empregados da PONTE S.A. e, quando aplicável, combinadas com outras questões de modo a desenvolver um novo conhecimento. Por exemplo: reclamações sobre o grande fluxo de carros que usam o sistema onda livre são combinadas com os horários em que incidem as reclamações.

Finalmente a internalização é promovida no Programa de Exploração da Ponte (PEP). A manutenção física da ponte é realizada com base no PEP mais a documentação de engenharia (conhecimentos explícitos). Esta conversão do conhecimento ocorre na fase de conclusão dos projetos, quando o estado da arte é

registrado nos relatórios e disseminado para as pessoas envolvidas. Nesta fase, acontece a difusão interativa desse novo estado da arte.

Dois termos são importantes quando são analisados os lados da espiral do conhecimento: **cultura e tecnologia**. Para promover a conversão de qualquer forma de conhecimento em tácito deve haver uma cultura corporativa para tal, de forma que as os indivíduos tenham motivação para participar dos processos, por exemplo. A tecnologia atua no lado da criação do conhecimento explícito, trazendo facilidades e agilidade para a tomada de decisão, por exemplo. Antes da utilização de qualquer tecnologia é necessária uma alteração na cultura da empresa. Além disso, o estabelecimento de maravilhosos processos de gestão de pessoas tende ao fracasso sem o devido aparato tecnológico.

Nenhuma abordagem geral da gestão do conhecimento foi aceita de forma pacífica, embora diversas noções estejam avançando de modo isolado. Uma dessas noções diz respeito à gestão do conhecimento explícito usando-se abordagens técnicas. Elas focalizam o conhecimento adquirido pelo indivíduo através do uso de banco ou sistemas de conhecimento computadorizados, *groupware*, ferramentas de apoio a decisão e outros instrumentos.

Os trabalhos mais relevantes para esta dissertação (NONAKA & TAKEUCHI, 1995; SVEIBY, 2001) foram mais detalhados nesta seção. Conforme será visto mais adiante e como não poderia deixar de ser, a ferramenta BOE atua no lado tecnológico da espiral de conhecimento, promovendo externalização e combinação. Além disso, acredita-se que uma ferramenta tecnológica não atua por si só em um ambiente de gestão do conhecimento, devendo ser devidamente “promovida” por processos de gestão adequados.

2.3.5 Ferramentas de Gestão do Conhecimento

Como o escopo deste trabalho é fundamentalmente tecnológico, esta seção irá apresentar alguns conceitos a respeito de ferramentas tecnológicas para a gestão do conhecimento. Além disso, algumas categorias de ferramentas existentes serão discutidas para haver uma melhor análise a respeito do tópico.

Dada a já discutida amplitude do termo gestão do conhecimento, percebe-se que alguns fornecedores de software tendem a embutir interpretações erradas do

conceito em seus produtos, ou até mesmo usar a gestão do conhecimento apenas como um novo modismo ou apelo de marketing. Desta forma, a definição de categorias para classificar este tipo de tecnologia faz-se necessária com a finalidade de orientar as empresas na sua correta utilização. As ferramentas de gestão do conhecimento pretendem auxiliar no processo de captura e estruturação do conhecimento de grupos de indivíduos, disponibilizando este conhecimento em uma base compartilhada por toda a organização (CARVALHO & FERREIRA, 2000).

Algumas propostas para a representação do conhecimento, como os diagramas de estrutura do conhecimento e os mapas do conhecimento, têm sido implementadas por algumas ferramentas. Implementações destas ferramentas incluem repositórios de conhecimento, sistemas de apoio à decisão baseada em casos, mapas do conhecimento, softwares de modelagem de processos ou até mesmo sistemas especialistas e redes neurais. Esta diversidade de abordagens frequentemente confunde as organizações no momento da escolha da solução mais adequada para o seu tipo de problema.

RUGGLES (1997) define ferramentas para gestão do conhecimento como tecnologias amplamente definidas que aumentam e ativam a criação, codificação e transferência do conhecimento.

CARVALHO & FERREIRA (2000) concordam e definem ferramenta de gestão de conhecimento como um tipo específico de software que oferece apoio a pelo menos uma das atividades de geração, codificação ou transferência de conhecimento definidas por DAVENPORT & PRUSAK (1998). Os autores argumentam ainda que a amplitude da definição tem a finalidade de englobar não somente as ferramentas que se auto-intitulam como sendo de gestão do conhecimento, mas também softwares de GED (Gerenciamento Eletrônico de Documentos), sistemas de *Business Intelligence*, ferramentas de *workflow*, produtos *groupware* e sistemas especialistas.

Segundo a mesma linha de pensamento, PIROLA (2002) argumenta que uma ferramenta específica no apoio à gestão do conhecimento pode ser definida como um *software* idealizado, projetado e implementado para dar suporte à gestão de pelo menos um dos capitais do conhecimento. No entanto, segundo este mesmo autor, uma abrangência maior é necessária, de forma que ferramentas de apoio à gestão do

conhecimento são definidas como *softwares* que apóiam a gestão de qualquer um dos capitais do conhecimento, mesmo que não tenham sido projetados inicialmente com este intuito.

CORMICAN & O'SULLIVAN (2000) publicaram uma definição mais prática para ferramenta de gestão de conhecimento. Segundo os autores o objetivo de uma ferramenta de gestão de conhecimento não é gerenciar conhecimento por si só, mas facilitar a automação e implementação de elementos no processo de conhecimento como um todo. O trabalho destes autores propõe a criação de uma ferramenta colaborativa para a gestão do conhecimento objetivando a inovação produtiva.

Para DAVENPORT & PRUSAK (1998), o objetivo das ferramentas de gestão do conhecimento é modelar parte do conhecimento que existe nas cabeças das pessoas e nos documentos corporativos, disseminando-os através de toda a organização. A existência de conhecimento na empresa é de pouco valor se este não estiver acessível. Com tais ferramentas pretende-se que o conhecimento possa fluir através de redes de comunidades, com a tecnologia desempenhando o papel de meio e o conhecimento o de mensagem.

Neste ponto, torna-se importante investigar alguns temas acerca do uso de ferramentas de gestão do conhecimento utilizando uma visão mais geral, ou seja, a natureza dessas ferramentas. Além disso, o enorme número de ferramentas que potencialmente podem apoiar o processo de gestão do conhecimento, estas freqüentemente com constantes evoluções e extensões, leva a opção de uma análise desta natureza.

WENSLEY (2000) assume a definição de RUGGLES e busca correlacionar a natureza das ferramentas com as características de cada etapa presente na definição. Primeiramente, para permitir a geração de conhecimento (*knowledge generation*), uma ferramenta deve permitir o compartilhamento de conhecimento pré-existente e com algum tipo de estrutura. É apenas através do compartilhamento de conhecimento que se torna possível a descoberta de pontos em que novos conhecimentos são necessários.

No que diz respeito à codificação e refinamento do conhecimento (*knowledge codification and refinement*), WENSLEY (2000) cita a pesquisa em inteligência artificial como um importante delimitador de potenciais e limitações para a

representação e codificação de conhecimento. Uma abordagem popular e bastante aceita segundo o autor consiste na utilização de sistemas regras para a codificação do conhecimento. Tais regras garantem que deduções que partem de proposições válidas também são válidas. Ainda é discutida a amplitude do uso de regras para a representação de qualquer conhecimento. É sabido que, apesar da flexibilidade, sua utilização não é suficiente para uma representação completa e satisfatória, e é este o cuidado que uma ferramenta de gestão de conhecimento deve tomar.

Finalmente, para a transmissão de conhecimento, a *web* é citada como um meio de transmissão que independe do tempo e de plataformas computacionais. Uma vez que o conhecimento foi gerado, codificado e refinado, este pode ser disponibilizado para qualquer pessoa com acesso a *web*.

Este trabalho se enquadra na abordagem de natureza de ferramentas de WENSLEY (2000) pois procura estimular a geração de conhecimento (*knowledge generation*) através de processos colaborativos que envolvam conhecimentos explícitos de especialistas. A codificação e o refinamento também são promovidos com a utilização de um sistema de regras de inferência apoiado por bancos de dados auxiliares. A característica *web* da ferramenta BOE a credencia a participar do processo de transmissão de conhecimento de maneira eficiente.

2.3.6 Gestão do Conhecimento Científico

A condução da pesquisa científica está ficando criticamente dependente de um gerenciamento efetivo dos dados científicos e informações técnicas. Os rápidos avanços nos instrumentos científicos, computadores e comunicação permitiram aos cientistas a coleta, geração, processamento e compartilhamento de um volume de dados sem precedentes.

O dinamismo e a competitividade que marcam o ambiente empresarial também se mostram presentes no meio científico. Um dos temas abordados por uma revista especializada em medicina foi o monopólio dos resultados do projeto Genoma, que efetuou um mapeamento do código genético do ser humano. No campo da genética, como no da informática (o caso Microsoft é exemplar), podem ser criadas situações de monopólio que são um obstáculo à competição científica, à concorrência produtiva e à difusão universal dos benefícios (BERLINGUER, 2000).

Por esta razão, as instituições de pesquisa estão abrindo os seus olhos para a importância da captura, armazenamento e disseminação do conhecimento existente dentro delas, o que levanta a importância da aplicação da gestão do conhecimento na ciência.

Qualquer que seja o cenário futuro, da agricultura à indústria, da mineração aos serviços, o desenvolvimento brasileiro e sua inserção competitiva no século XXI dependerão das tecnologias da informação, que representam a convergência das tecnologias de computação, comunicação e informação. E é através da pesquisa, educação e infra-estrutura adequada que se conseguirá dominar e aplicar as tecnologias da informação a setores de crítica importância para o Brasil.

O tradicional modelo linear de desenvolvimento científico e tecnológico, que vai, passo a passo, da pesquisa básica à geração de produtos, não mais atende à competitividade necessária aos tempos atuais (ALMEIDA, 1998). Essa observação fica ainda mais evidente no caso da informática, onde o ciclo de geração de novos produtos e tecnologias é cada vez mais curto. Atualmente, nos EUA e outros países altamente industrializados, as várias fases do processo de pesquisa e desenvolvimento ocorrem concomitantemente, com a participação de universidades, empresas e institutos de pesquisa. Algumas áreas do saber têm sido denominadas "*strategic enabling research areas*" devido à amplitude do impacto dos resultados de suas pesquisas nas demais áreas do conhecimento científico e tecnológico. Tecnologias da informação, biologia e ciência dos materiais são exemplos claros de "*strategic enabling areas*" para o século XXI. Assim, frente a essas observações sobre as tecnologias da informação, é oportuno analisar o papel da gestão do conhecimento no apoio à pesquisa e ao desenvolvimento científico.

No atual contexto da gestão do conhecimento, o escopo empresarial tem tido um grande destaque. Entretanto, não é aconselhável esquecer de quem é a responsabilidade pela produção do conhecimento científico. Universidades e instituições de pesquisas estão cada vez mais preocupadas em recuperar dados de maneira eficaz e encontrar colaboradores ou experimentos para suas áreas de atuação. Pesquisadores criam e trocam informações com mais rapidez e num volume muito maior do que se poderia imaginar no passado.

Assim, torna-se necessário um crescente investimento na integração entre cientistas de diferentes áreas. Algumas propostas de integração pregam que as próprias universidades deveriam permitir que professores mantenham vínculos com dois ou mais departamentos. As instituições superiores ainda têm de incentivar a criação de núcleos e centros de pesquisa e dar a esses grupos uma estrutura flexível.

Alguns processos da gestão de conhecimento podem ser aplicados com coerência no ambiente científico. A gestão de competências é importante, pois pode vir a facilitar a integração de grupos de pesquisa e, conseqüentemente, incentiva o compartilhamento de conhecimento. A gestão do capital intelectual, na medida em que promove explicitação e reuso de conhecimento científico, pode vir a facilitar a criação de novos conhecimentos, às vezes através de combinações de outros conceitos previamente conhecidos.

Conforme a classificação discutida na seção 2.3.2, é natural a associação do conhecimento científico ao processo de gestão de conhecimento quando se fala em gestão do conhecimento científico. O entendimento dos tipos de conhecimento envolvidos em um processo de gestão é importante para que não surjam dificuldades decorrentes de interpretações problemáticas.

Por exemplo, é sabido (WENSLEY, 2000) que organizações evoluíram em diversas “formas de vida” relacionadas. A criação de disciplinas funcionais resultou na existência de diferentes tipos de conhecimento no interior das organizações. Alguma parte deste conhecimento certamente tem características científicas. Frequentemente, departamentos de pesquisa e desenvolvimento têm uma cultura científica forte e dispõem de ferramentas que são típicas da comunidade científica. O conceito de “universidade corporativa” também é um bom exemplo. Por outro lado, o entendimento do comportamento de consumidores feito por um departamento de marketing pode ser apoiado por argumentos científicos, mas pode ser nada mais do que intuição ou conhecimento tácito de seus gerentes.

2.3.6.1 Trabalhos Relacionados

Tomando os fatos discutidos acima como motivação para a aplicação da gestão do conhecimento na ciência, alguns trabalhos relacionados a esta área específica serão debatidos nos próximos parágrafos.

HARS (1998) discute de quais formas os avanços da tecnologia da informação afetam os processos de pesquisa e publicação científica. Para estudo destes aspectos, foi desenvolvida uma infra-estrutura *web* de gestão do conhecimento que permite a criação, síntese e aplicação de conhecimento científico, denominada *Cybrarium*. Além da apresentação da infra-estrutura, o trabalho também enfatiza os aspectos principais para a organização do conhecimento científico em meios eletrônicos.

No Banco de Teses e Dissertações (BTD) proposto por PACHECO & KERN (2001) para a UFSC e sua biblioteca digital, além dos produtores de conhecimento (alunos e professores), são atendidos tomadores de decisão e gestores do sistema de pós-graduação, subsidiando-os com informações sobre nível de interesse e intercâmbio de conhecimento entre cada uma das áreas de concentração do programa.

Desde seu surgimento, em 1995, o BTD contempla trabalhos em texto completo entre as dissertações de mestrado e doutorado já defendidas no programa desde 1970. Além disso, o BTD acrescenta o acesso ao conteúdo dos trabalhos uma série de estatísticas de busca. Estas estatísticas têm embasado a tomada de decisão de gestores do PPGE/UFSC (PACHECO & KERN, 2001).

Em (NAKAYAMA, SUMITA et al., 2000), um sistema baseado em conhecimento e informação sob demanda foi implementado objetivando a estruturação de textos em linguagem natural como uma forma de conhecimento. Usuários podem, através de linguagem natural, submeter consultas a respeito de conhecimentos específicos. No experimento relatado, somente as experiências individuais (*know-how*), como, por exemplo, idéias e dicas a respeito de pesquisas, são estruturadas e potencialmente compartilhadas. A exclusão de artigos técnicos, documentação de projetos etc, se deveu a já existir implementado um sistema de Gerência Eletrônica de Documentos. A validação do trabalho foi efetivada através da análise do esforço prático de compartilhamento de conhecimento no Centro de Pesquisa e Desenvolvimento da Toshiba.

Seguindo a mesma concepção do trabalho acima, RENAUX, STRAUHS et al. (2001), descrevem a ferramenta KNOW-LIT, de gestão do conhecimento científico em um laboratório de pesquisa. O laboratório exemplificado no trabalho é o LIT, da

área de informática e telecomunicações, localizado no CEFET-PR. Os recursos humanos disponíveis no LIT são principalmente professores de diversos departamentos e alunos de pós-graduação (mestrado e doutorado) e graduação. Em função desta característica, o fenômeno de *turnover* está presente, influenciando na perda de conhecimento quando, por exemplo, um aluno se forma e deixa o laboratório.

A ferramenta KNOW-LIT é capaz de criar mapas de conhecimento com a dupla função de mapear competências e também de representar uma ontologia pertinente ao ambiente aplicado. O enfoque utilizado nesta iniciativa é auxiliar as atividades de gestão do conhecimento promovendo informações como: as áreas de conhecimento presentes no laboratório, as pessoas que dominam o conhecimento nas áreas determinadas, os relacionamentos entre as diversas áreas e palavras-chave, os passos a serem seguidos em um treinamento que objetiva o aumento de conhecimento em uma determinada área, e o nível de documentação disponível relacionado a cada área.

Saindo da linha da gestão de competências acadêmica, no trabalho de ALLEN, BOSCH et al. (2001), a abordagem ISKM (*Integrated Systems for Knowledge Management*) é apresentada para ilustrar de que forma as abordagens de GC baseadas em aprendizado podem ser utilizadas para auxiliar comunidades a desenvolver, aplicar e refinar informações técnicas dentro de um amplo contexto de conhecimento compartilhado. Neste trabalho é apresentado um estudo de caso de controle de doenças em animais, em que a tomada de decisão ambiental é aperfeiçoada.

O projeto *ScienceDesk* (KELLER, 1996), implantado na NASA, provê uma solução de ferramentas *web* para a colaboração, organização, execução e administração de projetos científicos. Dentre as funcionalidades previstas no projeto, as de maior relevância para este trabalho são:

- Gestão do Conhecimento Científico – captura, preservação e rastreabilidade do conhecimento científico em grupos de pesquisa;
- Acesso a Informação – indexação inteligente e técnicas de visualização;
- Colaboração – uso de métodos síncronos e assíncronos; e

- Experimentos Remotos – controle e monitoramento de experimentos de forma semi-autônoma.

O projeto *SCIx* (TURK, BJORK et al., 2002) tem como objetivo: i) permitir que cientistas tenham acesso eficiente a publicações de sua área através da existência de um repositório de publicações científicas, ii) tornar disponível o material existente no repositório para profissionais ditos “não-cientistas” ou da indústria, e iii) dar suporte a uma comunidade virtual de autores e leitores de publicações científicas.

A partir da arquitetura do SPeCS descrita na seção 1.3, através da sua instanciação, deu-se a criação do ambiente **Agromet** (PINTO, STRAUCH et al., 2002; SOUZA, STRAUCH et al., 2002), uma plataforma que dá suporte ao trabalho científico de maneira colaborativa para a área de agrometeorologia. Para atingir o seu objetivo, o **Agromet** é composto de cinco módulos que são responsáveis por:

- Integração de Dados – Agromet X-Arc;
- Suporte aos *Workflows* Científicos – Agromet BOE;
- Suporte a Decisão Colaborativa – Agromet SPeCS;
- Gestão do Conhecimento Científico – Agromet Epistheme;
- Gestão de Documentos e Conteúdo – Agromet DOC.

O **Agromet** disponibiliza as funcionalidades básicas do SPeCS, pois foi o esforço de integração das ferramentas que vem sendo desenvolvidas na COPPE / UFRJ, conforme a Figura 2.8. Para a validação do ambiente colaborativo em um domínio científico real foi escolhida a área de Agrometeorologia da EMBRAPA Solos. A heterogeneidade da área, bem como a explícita distribuição de conhecimento foram fatores de motivação para a escolha desta área científica. As maiores contribuições ao ambiente **Agromet** foram as reduções nos custos de aquisição de dados, a redução de duplicidade de dados, troca de conhecimento e otimização na seleção e tratamento de dados. O BOE, refletido na ferramenta Agromet BOE, foi de fundamental importância para a conquista dos objetivos iniciais propostos.

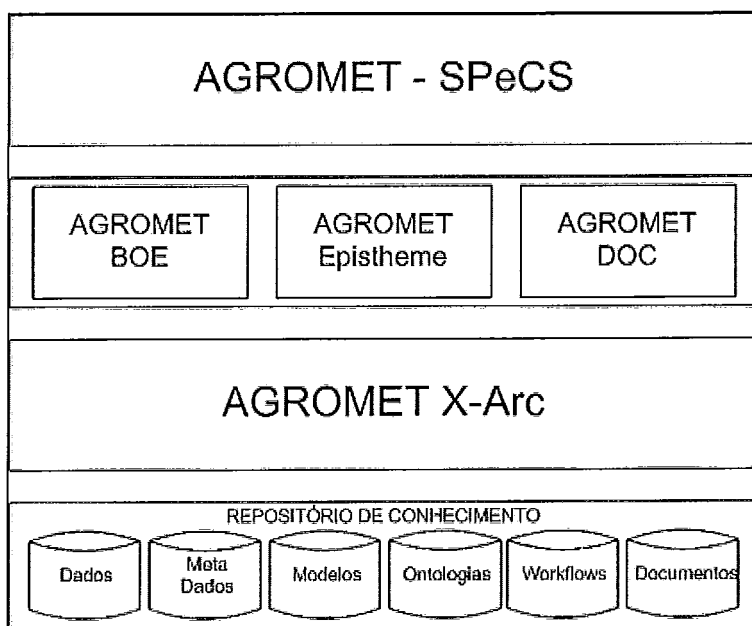


Figura 2.8 - Arquitetura do ambiente AGROMET

Capítulo 3 - Proposta da Ferramenta

Bill of Experiments

Visando a implementação de um ambiente de explicitação, reuso e planejamento de *workflows* científicos que atenda aos requisitos detalhados no capítulo dois, a ferramenta *Bill of Experiments* (BOE) foi projetada. Este capítulo apresenta os detalhes acerca da proposta da ferramenta BOE e discute sua arquitetura e módulos.

3.1 Introdução

Aproveitando as já citadas melhorias na área de comunicação (primordialmente a *internet*), e a crescente demanda por conhecimento que historicamente é característica dos cientistas do mundo inteiro este trabalho propõe um ambiente colaborativo para explicitação, reuso e planejamento de experimentos científicos. A ferramenta BOE é um sistema *web* dotado de um banco de dados de experimentos científicos (*workflows* ou protocolos científicos) capaz de apontar melhores alternativas para a realização destes através de técnicas de IA, tais como CLP⁶ e heurísticas (DASGUPTA, CHAKRABARTI et al., 2002). O planejamento do experimento, etapa adicional do ciclo de vida de experimentos científicos proposta neste trabalho, é promovido de forma intensa.

A natureza do ambiente proposto ainda favorece o processo colaborativo no sentido de permitir a explicitação, combinação, geração (*design*) e reuso de conhecimentos científicos inerentes aos experimentos e *workflows* científicos. Cabe ressaltar que o ambiente colaborativo é de fundamental importância para a implementação de uma métrica de qualidade subjetiva para os artefatos científicos presentes na ferramenta.

⁶ Sigla de *Constraint Logic Programming*. Uma mesclagem de dois paradigmas declarativos: programação lógica e solução de restrições.

O BOE ainda beneficia fornecedores de dados e modelos matemáticos bem como instituições que trabalhem com a realização de partes de protocolos científicos. Tal grupo de usuários, aqui denominados simplesmente “fornecedores” conseguirão melhorar o alcance de seus serviços e potencializar sua posição no mercado. Esta característica da ferramenta a aproxima de um paradigma *B2B*, que poderia ser denominado *B2Sci* (*Business to Scientists*).

Em resumo, enquanto pesquisadores e cientistas são responsáveis pela publicação de trabalhos, com detalhes a respeito dos processos do seu protocolo científico, instituições interessadas em vender ou simplesmente disponibilizar seus dados ou outros artefatos científicos contribuem com informações de custos e prazos de obtenção destes. Desta forma é possível obter colaboração de duas formas: (i) cientistas com cientistas, *Sci2Sci* (permitindo o reuso de experimentos) e (ii) cientistas com fornecedores, *B2Sci* (garantindo melhores opções para a realização dos experimentos).

O principal objetivo da ferramenta BOE é prover dados, modelos, programas, definições de *workflow* e metadados sobre experimentos científicos, e, na inexistência destes, com o auxílio de um ambiente colaborativo, orientar o pesquisador a como obtê-los, tornando-se uma ferramenta para melhorar o acesso, compartilhamento e conseqüente inovação do conhecimento científico inerente aos experimentos.

3.2 Analogia ao *Bill of Materials* (BOM)

Na literatura de Banco de Dados um dos exemplos mais comuns para expressar “auto-relacionamento”, ou relacionamento unário, tanto na modelagem conceitual quanto na modelagem lógica de dados é o “*Bill of Materials*” ou BOM. Este exemplo ocorre em (CHEN, 1976; DATE, 2000; ELMASRI & NAVATHE, 2000), entre outros, e possui um desdobramento natural, chamado de “agregado recursivo” (RUMBAUGH, 1991).

Trata-se do relacionamento entre peças cujos componentes, em geral, podem ser outras peças, e cada nova peça pode possuir outros componentes, estabelecendo-se o relacionamento recorrente.

A instância de BOM encontrada em (SULAIMAN, 1999) consiste em um aplicativo de planejamento de compra de peças (partes) para construção ou montagem

de componentes eletrônicos. Cada BOM é formado de componentes e cada componente é composto de itens. Um componente só pode ser considerado disponível quando todos os seus itens estiverem disponíveis também. Cada item simples é considerado disponível somente quando é escolhido a partir das opções existentes. Se um item é composto, pode ser expresso como um novo BOM (Figura 3.1).

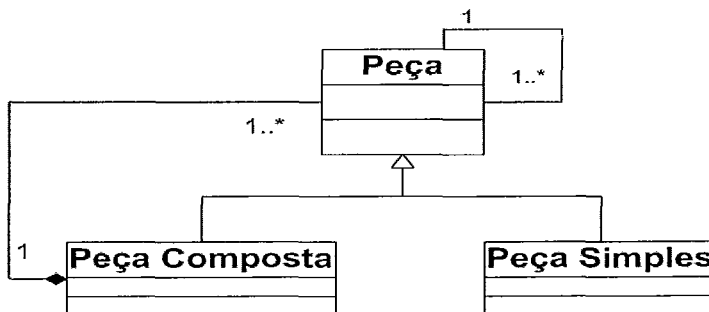


Figura 3.1 - Diagrama de classes exibindo o auto-relacionamento entre peças e seus componentes com o desdobramento em um agregado recursivo

Em (SULAIMAN, 1999) considera-se o segmento do mercado de eletrônicos e a existência de um agente de busca que encontra e armazena em um sistema de banco de dados a oferta de peças dos diversos fornecedores, examinando-se custos e prazos de disponibilidade. As transações entre o agente de busca e os fornecedores de peças constituem um sistema *web* fundamentado no conceito B2B, tendo sido concebido o código XML correspondente às bases de regras do modelo. A interação de um cliente potencial com um sistema *web* é fundamentada no conceito B2C e é planejada de forma que o comprador possa avaliar uma compra baseando-se em: melhores preços, prazos mais curtos, marcas, ou um balanceamento destas variáveis de forma que se atinja uma melhor relação custo-benefício. Assume-se ainda que um cliente tem o conhecimento de quais peças necessita comprar para a correta montagem de um componente.

Um componente é considerado como pronto para entrega a partir do momento em que todos os seus itens são instanciados e, conseqüentemente, cada peça é completamente identificada com o respectivo preço, prazo estimado de entrega, fornecedor de origem e disponibilidade.

As restrições do modelo são as seguintes: i) o prazo de entrega de um componente sempre é, no mínimo, igual ao maior prazo de entrega entre todos os itens

envolvidos; e ii) o menor preço de um componente consiste no somatório dos preços dos itens de menores preços, nos quais os agregados formem o componente. Explorando-se a natureza dedutiva do sistema, o usuário pode solicitar explicações sobre preços, prazos e montagem de algum componente. Em (SULAIMAN, 1999) encontram-se propostas de solução para o componente de melhor custo-benefício, com os usuários inserindo limites de tolerância para as variáveis de preço e prazo, e atribuindo pesos para estas grandezas. Entende-se peso como uma relação de importância de uma grandeza em relação à outra. Por exemplo, caso o prazo de entrega de um componente seja de maior relevância que seu custo, a variável “prazo” terá maior peso, alterando a escolha da melhor solução.

O conceito que embasou todo o trabalho desta dissertação surgiu da analogia com o BOM. Para uma melhor compreensão desta proposta, estabelece-se uma analogia entre a preparação do experimento científico e a compra de itens para montagem de componentes.

O processo de montagem de uma peça composta pode ser associado à aplicação de uma simulação ou experimento utilizando um ou mais conjuntos de dados científicos, gerando com isto um novo resultado experimental que pode vir a ser relevante para a comunidade pesquisadora. Da mesma maneira que uma peça é composta por diversos itens, um novo resultado é composto por dados, experiências utilizando tais dados e a execução de simulações. Assim como uma nova peça gerada pode servir como um item a ser utilizado na composição de outras novas peças, um novo resultado gerado no BOE pode servir de base para a aplicação de um outro modelo matemático ou experimento que gerará uma terceira informação e assim sucessivamente. Ou seja, cada descoberta utilizando dados do BOE, é incorporada à Base de Conhecimento e poderá ser utilizada futuramente como dado na geração de novos resultados.

A Tabela 3.1 esquematiza cada termo da analogia empregada no BOE. É importante citar que o enfoque adotado para dados é geral, ou seja, ele pode ser estendido para outros artefatos comuns à pesquisa científica, como modelos matemáticos, protocolos de sub-experimentos, ferramentas de automação etc.

Tabela 3.1 - Analogia BOM e BOE

BOM	BOE
Peça	Dados Iniciais
Componente	Dados Derivados
Empresa	Fornecedor
Preço	Custo
Prazo	Prazo
<<Não Existe>>	Qualidade

A métrica de qualidade não está presente no modelo proposto por SULAIMAN (1999). Qualidade é um conceito complexo, porque possui significados diversos para diferentes pessoas. Portanto, gerar uma medida de qualidade aceitável para todos não é trivial. O BOE utiliza uma avaliação subjetiva da qualidade dos dados em que especialistas julgam aspectos relevantes dos dados utilizados (esta avaliação será descrita na seção 3.3.4).

A partir da explicação da analogia, torna-se claro o propósito do BOE de armazenar os *workflows* científicos como regras de montagem de dados a partir de outros dados e artefatos. Este aparato será de fundamental importância para o módulo de inferência (seção 3.3.5), responsável pela etapa de planejamento de execução de um experimento científico.

3.3 Arquitetura da Ferramenta BOE

Distribuída em três camadas, a arquitetura do BOE está ilustrada na Figura 3.2. Os componentes principais da ferramenta estão distribuídos nas camadas de acesso *web*, controle e banco de dados.

A camada de acesso *web* é o ponto de partida da ferramenta. Ela é a responsável por permitir a captura, reuso e planejamento de *workflows* científicos. O módulo de projeto do *workflow* torna possível a representação dos *workflows* científicos de uma forma clara e concisa.

Ainda na camada de acesso *web*, existem os portais de pesquisadores e de fornecedores. No portal dos pesquisadores, reside toda a estrutura necessária para a busca de melhores alternativas para a realização dos experimentos, busca de experimentos alternativos, avaliação da qualidade de dados e interação com outros

cientistas. No portal dos fornecedores, há opções para cada fornecedor descrever os artefatos científicos oferecidos e acompanhar a utilização e avaliações de qualidades destes.

A camada de controle é composta de três módulos: inferência, busca e gerência de qualidade. O módulo de inferência utiliza heurísticas para inferir, a partir das definições dos experimentos e do banco de dados dos fornecedores, melhores alternativas para a realização dos experimentos. No módulo de busca da camada de controle reside a funcionalidade de pesquisa de artefatos científicos, como experimentos, dados e produtos científicos, pesquisadores e fornecedores de uma determinada área etc. A qualidade dos artefatos científicos tratados pelo BOE é gerenciada pelo módulo de gerência de qualidade.

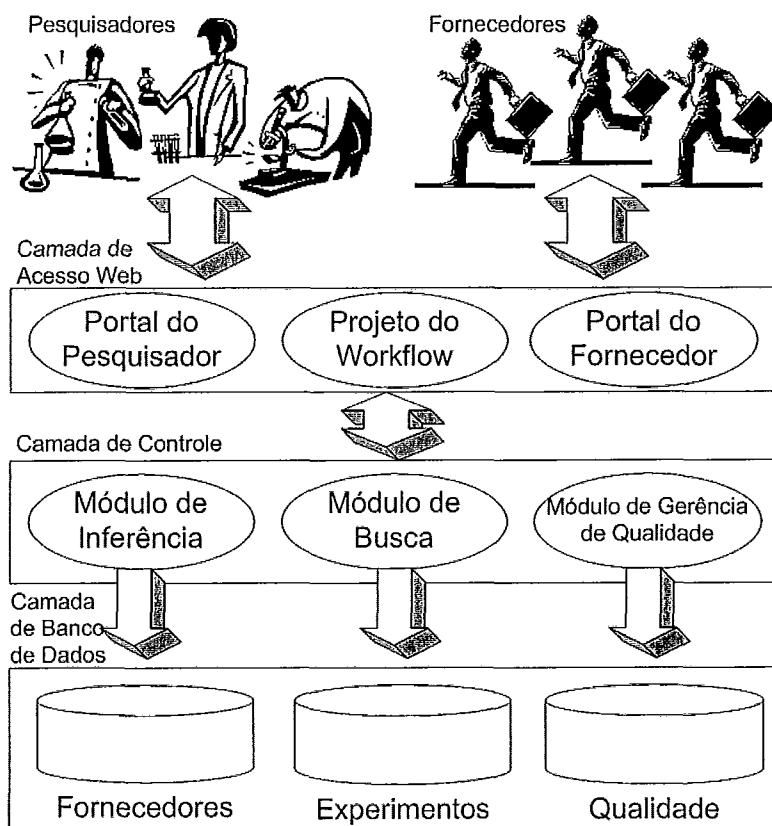


Figura 3.2 - Arquitetura da Ferramenta BOE

A camada de banco de dados é responsável por armazenar experimentos, fornecedores e as avaliações de qualidade existentes no ambiente.

As próximas seções trazem maiores detalhes sobre cada camada e módulo existentes na arquitetura da Figura 3.2.

3.3.1 Portal do Pesquisador

O portal do pesquisador é um ambiente *web* em que o cientista pode interagir com a ferramenta BOE. Nele se encontram as funcionalidades para o apoio ao trabalho científico propostas no sistema.

A partir do portal é possível: i) a criação de um novo experimento científico, com um título, a informação da área científica a que ele se refere e uma descrição detalhada do propósito do mesmo (Figura 3.3); ii) a definição de cada etapa e dos dados utilizados no *workflow* científico do experimento, através da utilização do módulo de projeto do *workflow*; iii) a consulta a outros experimentos relacionados, através de consultas direcionadas por palavras-chave (Figura 3.4); iv) a criação de experimentos combinados a partir de dois ou mais *workflows* previamente existentes no sistema; v) o preenchimento do formulário de critérios de qualidade de dados científicos, gerando uma métrica de qualidade colaborativa; e vi) o planejamento do experimento a partir da busca por melhores alternativas, implementado pelo módulo de inferência.

Workflow Inference	Click Here to Search for Best Alternatives
Workflow Definition	Click Here to Edit
WorkFlow Name	Zoneamento Agroclimatico
Description	Zoneamento Agroclimatico - EMBRAPA SOLOS
Sci Area	Agriculture

Submit Delete

UFRJ / COPPE - 2003

Figura 3.3 - Portal do Pesquisador Utilizado para Definir um *Workflow*

Workflow Search Fields

Scientific Area: Biology

Workflow Name: flow

Text Description:

Sort by: Select Field | Select Order

Records per page: Select Value

Search

Workflows

Workflow Name	Text Description	Scientific Area	Creation Date	Last Update Date	Create User	
CAL-LYSE Staining Protocol for Flow Cytometry	CAL-LYSE Staining Protocol for Flow Cytometry - MINI	Biology	19/11/02 15:08	01/02/03 00:00	manuel	c
Detection of Intracellular Antigens by Flow Cytometry	CALTAG Laboratories- Detection of Intracellular Antigens by Flow Cytometry	Biology	19/11/02 15:08	19/11/02 15:08	manuel	rr

Figura 3.4 - Portal do Pesquisador Efetuando uma Busca por Palavra-Chave

Como não podia deixar de ser, o portal do pesquisador é o ponto de partida da existência de toda a funcionalidade prevista nesta dissertação, a saber:

- explicitação (com a definição do *workflow* científico dos experimentos);
- reuso (através da combinação e até mesmo de buscas realizadas isoladamente); e
- planejamento (com a busca por melhores alternativas para a realização dos experimentos).

O apêndice A esquematiza as funcionalidades do portal do pesquisador na forma de diagramas de casos de uso.

3.3.2 Portal do Fornecedor

O portal do fornecedor também é um ambiente *web*. Através de seu uso, o fornecedor pode atuar na ferramenta BOE interagindo com cientistas para a efetiva realização de experimentos.

A partir do portal é possível: i) efetuar a cotação de custos e prazos para os artefatos científicos do ambiente (dados, modelos, experimentos simples etc) (Figura 3.5); e ii) acompanhar a cotação da qualidade dos dados pertencentes a cada fornecedor.

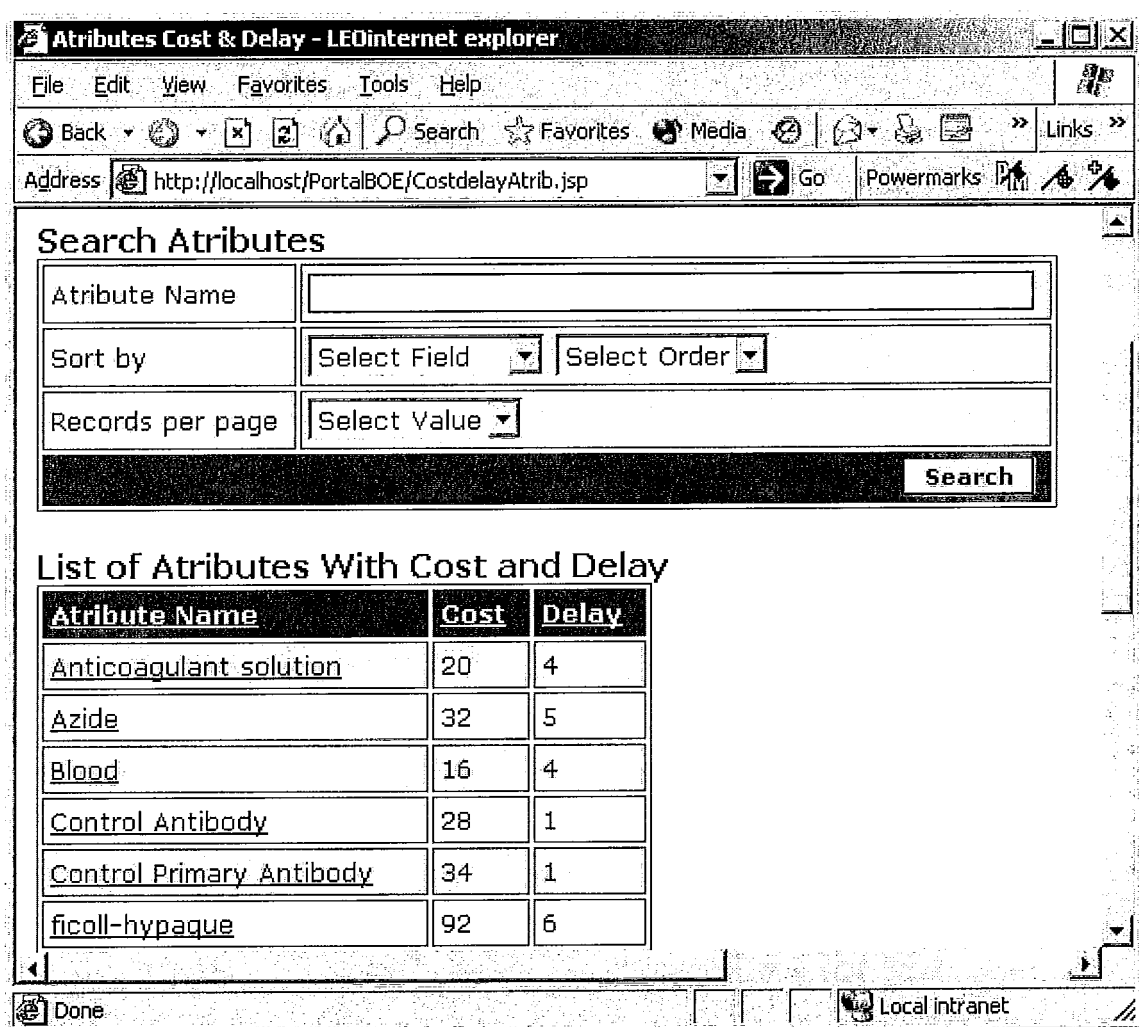


Figura 3.5 - Portal do Fornecedor Utilizado para Cotação de Dados Científicos

As cotações informadas no portal do fornecedor, junto com as informações geradas pelo módulo de gerência de qualidade, guiarão o módulo de inferência na busca por melhores alternativas para a realização do experimento. Este portal implementa a funcionalidade *B2Sci* proposta na introdução deste capítulo.

Assim como no portal do pesquisador, o apêndice A traz as funcionalidades do portal do fornecedor na forma de diagramas de casos de uso.

3.3.3 Projeto do *Workflow*

O módulo de projeto do *workflow* permite que os *workflows* científicos sejam representados (explicitados) graficamente, ocasionando simplicidade na sua utilização pelos pesquisadores. A idéia deste módulo consiste em dar motivação aos cientistas para que estes explicitem seus protocolos científicos de trabalho, permitindo posterior combinação e reuso.

Segundo a OMG (2001), criadora do padrão UML, o diagrama de atividades da UML é uma variação semântica de uma máquina de estados. Os chamados “estados” representam o desempenho de ações ou sub-atividades e as transições são executadas no momento em que uma ação ou sub-atividade se completa. O diagrama de atividades representa uma máquina de estados de um procedimento. No contexto da UML, recomenda-se o uso de diagramas de atividade quando os eventos de um procedimento ocorrem de maneira assíncrona.

Como neste trabalho um experimento é composto por uma série de atividades científicas interligadas e assíncronas, optou-se por utilizar uma extensão simplificada do diagrama de atividades da UML (OMG, 2001) para a representação dos fluxos de trabalho de cada cientista. A iniciativa de se estender o modelo de diagrama de atividades para a representação de *workflows* não é nova, sendo encontrada na literatura no contexto de processos de desenvolvimento de software (MURTA, 2002) e sistemas de produção (BASTOS & RUIZ, 2002).

O diagrama simplificado que esta dissertação propõe apresenta os seguintes objetos:

- **Atividade** – representada por uma elipse, é utilizada para modelar a execução de uma etapa no *workflow* científico;
- **Decisão** – representada por um quadrado com um ponto de interrogação, é utilizada para descrever uma decisão que deve ser tomada, utilizando como critério uma cláusula lógica;
- **Transição** – representada por uma seta que liga uma origem a um destino, é utilizada para que seja estabelecida uma seqüência de execução das atividades e decisões;

- **Início** – representado por um círculo preto, demarca o início da execução do *workflow* modelado; e
- **Término** – representado por um círculo cinza envolvido por uma circunferência maior, é o elemento em que se dará o fim da execução *workflow*.

No contexto dos *workflows* científicos representados nesta dissertação, a decisão permite que se faça uma escolha de um certo dado de entrada, dentre vários possíveis, para a realização de uma atividade. Cada transição leva consigo um conjunto de artefatos científicos que são necessários para a realização da atividade destino da transição (fluxo de dados), conforme a Figura 3.7.

O objeto de início é usado para representar artefatos simples, ou seja, aqueles que não são produzidos a partir de atividades (sendo obtidos via fornecedores, por exemplo). O objeto de término é empregado para representar artefatos derivados produzidos em alguma atividade, mas que não servem como entrada para outra, em outras palavras, transições que tenham como destino um “término” carregam artefatos que normalmente são os resultados experimentais de um *workflow* científico. Devido a essas razões, diversos objetos “início” e “término” podem existir num fluxo de trabalho científico modelado com este diagrama.

Optou-se por não usar as “raias” (*swim lanes*), existentes no diagrama de atividades da UML e com função de identificação dos responsáveis pelas execuções das atividades. Em seu lugar, um atributo em cada atividade modelada irá representar o responsável pela sua execução.

A Figura 3.6 mostra um exemplo do módulo de projeto do *workflow* em funcionamento para um experimento científico genérico e identifica os objetos descritos acima.

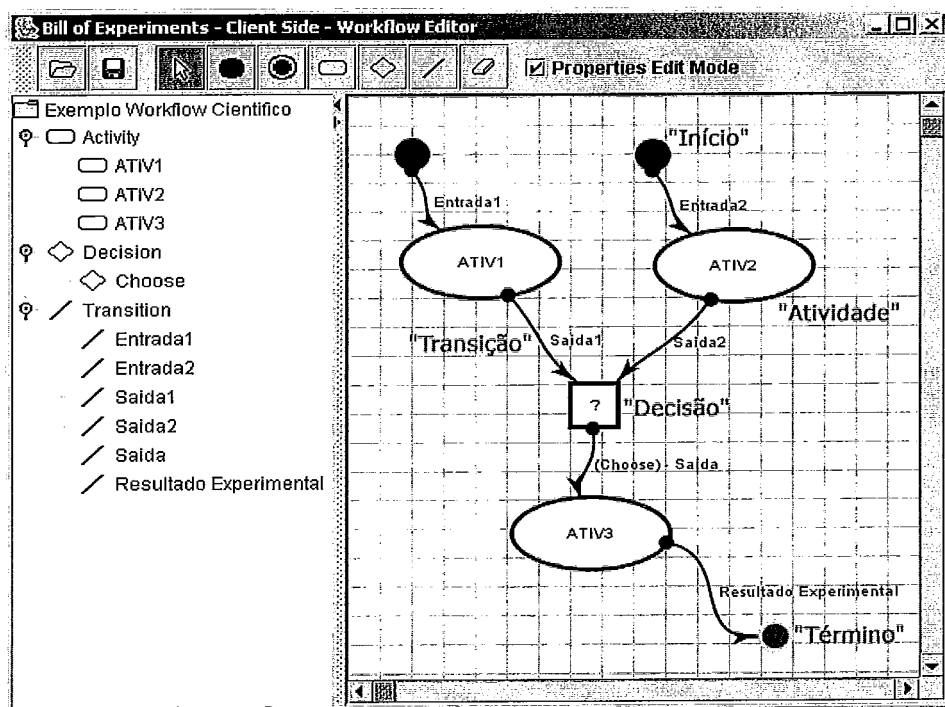


Figura 3.6 - Exemplo de *Workflow* Científico Representado no Diagrama

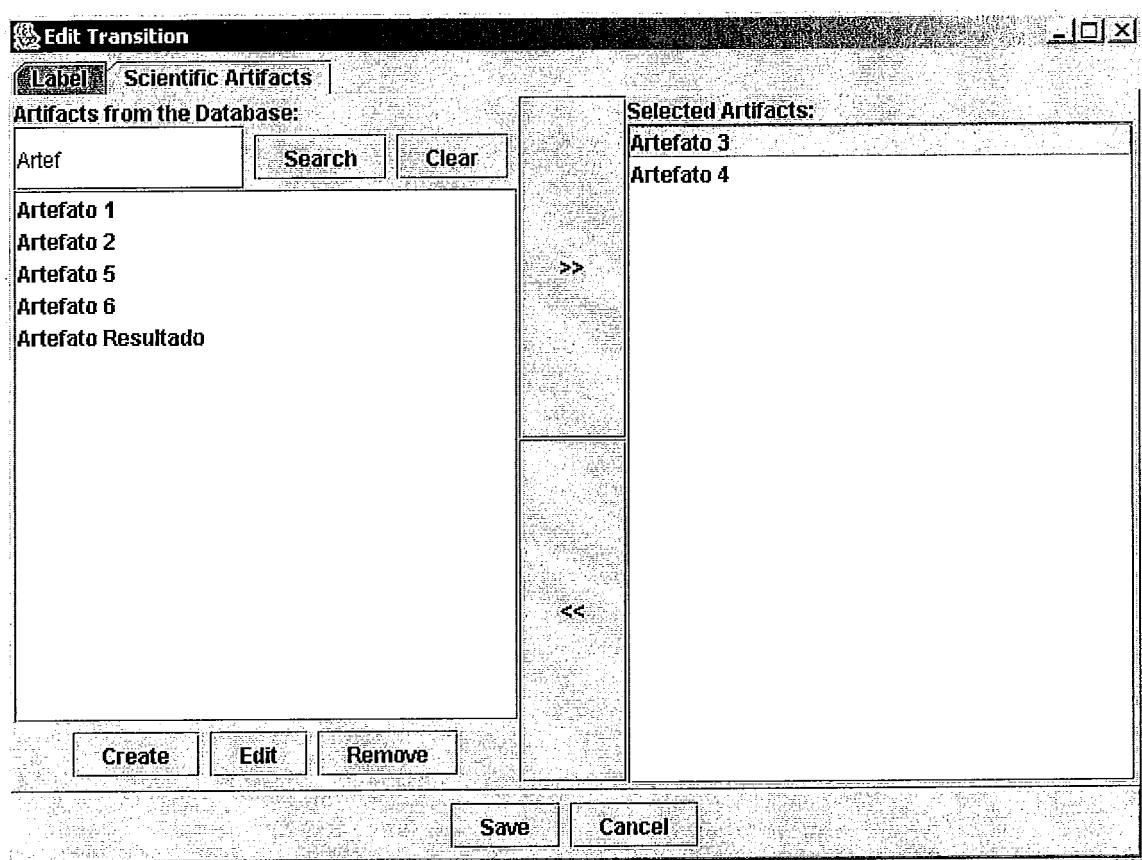


Figura 3.7 – Gerenciamento de Artefatos Ligados à uma Transição no Projeto do *Workflow*

3.3.4 Módulo de Gerência de Qualidade

Conforme citado na seção 3.2, a analogia que direciona o funcionamento da ferramenta BOE prevê a utilização de uma métrica de qualidade para os artefatos científicos presentes no ambiente. O módulo de gerência de qualidade é o responsável pela criação desta métrica para os dados científicos do BOE. Este módulo se baseia no método de avaliação da qualidade de dados pela não conformidade, proposto e validado por PINHO (2001).

Segundo WILLSHIRE & MEYEN (1997), existem dois tipos de avaliações que podem ser feitas em um conjunto de dados: avaliação **quantitativa** e avaliação **qualitativa**.

Normalmente, a avaliação **quantitativa** é realizada por ferramentas automatizadas que utilizam indicadores objetivos para mensurar a qualidade da representação e da estrutura dos dados. São observados, por exemplo, valores do domínio, presença de valor, regras de negócio, aferição de consistência, coerência de metadados etc. Este tipo de avaliação considera que a precisão e a objetividade são as duas dimensões primordiais para a avaliação da qualidade.

A avaliação **qualitativa** (subjetiva) pode ser realizada por especialistas ou usuários dos dados que utilizam indicadores subjetivos para julgar a qualidade atual do conjunto de dados em uso. Neste tipo de avaliação, os indicadores subjetivos são características de qualidade que foram definidas objetivando capturar a percepção e a expectativa do usuário sobre a qualidade de dados. Esta forma de avaliação pode ser feita com ou sem auxílio de ferramentas automatizadas. Cabe ressaltar que o ambiente colaborativo é de fundamental importância para a implementação de uma métrica de qualidade subjetiva, conforme apontado por TAYI & BALLOU (1998) e PINHO (2001) (Figura 3.8).

Não é objetivo dessa dissertação criar um novo método de avaliação subjetiva de qualidade de dados. A aplicação de um método já existente na literatura veio para confirmar a capacidade de extensão da analogia que criou o BOE. Assim, os próximos parágrafos descrevem resumidamente o método proposto por PINHO (2001) e apresentam a sua aplicação (com modificações) neste trabalho.

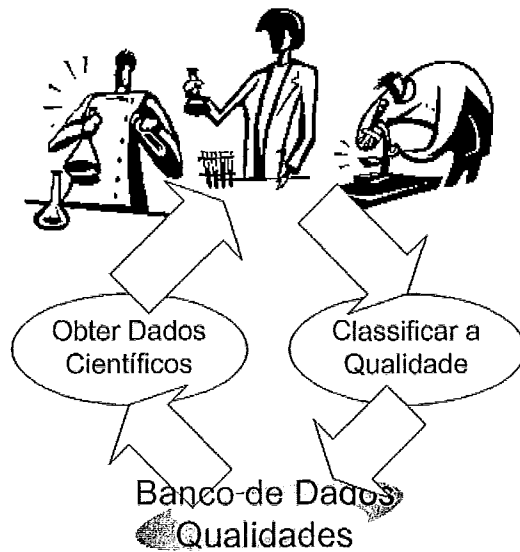


Figura 3.8 - Avaliação Colaborativa da Qualidade de Dados Científicos

Uma avaliação da qualidade de dados pela não-conformidade utiliza as não-conformidades (problemas) ocorridas em um conjunto de dados, considerando o peso de cada uma delas, a fim de verificar como a qualidade do produto avaliado foi degradada.

Através de pesquisas de campo, PINHO (2001) identificou as características e pesos mais importantes, bem como os problemas mais comuns que surgem durante o uso de um conjunto de dados. A partir deste levantamento, este trabalho selecionou as características julgadas mais importantes para dados exclusivamente científicos e simplificou o aparato matemático de estabelecimento de médias ponderadas. Todas as características selecionadas passam a ter pesos iguais. Além disso, a identificação de problemas se dá através do preenchimento de formulários, em que uma determinada pontuação é associada a um conjunto de dados científicos (Figura 3.9).

As características selecionadas para uso no BOE foram as seguintes:

- **Disponibilidade da Informação** – capacidade de se disponibilizar os dados em todos os locais onde poderão ser úteis, independentemente do local onde estão armazenados.
- **Idade dos Dados** – refere-se ao dado estar atualizado, expressando até que ponto a sua idade (tempo que está armazenado) é apropriada para a sua utilização.

- **Utilidade** – expressa a contribuição que os dados oferecem através da sua utilização. Por exemplo, um dado de grande utilidade pode ser usado em diversos experimentos.
- **Lucratividade** – refere-se ao aumento de produtividade conseguido através da utilização dos dados.
- **Benefício no Trabalho do Usuário** – característica da utilização dos dados que oferece simplificação e melhoria nas condições de trabalho dos seus usuários.
- **Competitividade** – capacidade da utilização dos dados proporcionar vantagens competitivas com relação a eventuais concorrentes, agregando mais valor às tarefas executadas.
- **Quantidade de Dados Adequada** – refere-se a quantidade ou volume de dados disponíveis ser apropriada para a realização das atividades.
- **Acurácia** – capacidade dos dados gerarem resultados corretos ou esperados através de manipulação.
- **Compleitude** – refere-se à existência de todos os dados para que sejam geradas informações úteis e necessárias.
- **Cobertura** – refere-se à característica que evidencia se o escopo da informação contida nos dados é suficientemente profundo, a fim de cobrir todo o escopo necessário exigido pelas atividades propostas.
- **Precisão do Dado** – capacidade do dado representar com exatidão, de forma perfeita, correta, clara e concisa o seu significado no mundo real.
- **Consistência** – refere-se ao valor de um dado ser o mesmo em toda uma eventual base de dados, não possuindo conteúdo contraditório ou significados diferentes.
- **Atributividade** – existência de mecanismos que registrem a identificação do autor de qualquer manipulação de dado, assim como a data e local de onde foi realizada a operação.

- **Flexibilidade** – característica que evidencia a capacidade dos dados serem manipulados e administrados facilmente, possibilitando a sua expansão, adaptação e agregação.
- **Interoperabilidade** – capacidade de interação com outros conjuntos de dados, com objetivo de complementar informações.
- **Facilidade de Compreensão da Informação** – capacidade da informação gerada ser objetiva, ou seja, que não coloque um potencial usuário em dúvida.
- **Uniformidade** – refere-se à capacidade dos dados estarem apresentados em um mesmo formato, estando eles consistentemente representados e formatados.
- **Disponibilidade de Documentação** – refere-se à existência de uma documentação atualizada e disponível dos dados, tornando-os assim bem documentados, verificáveis e facilmente associados a uma fonte.

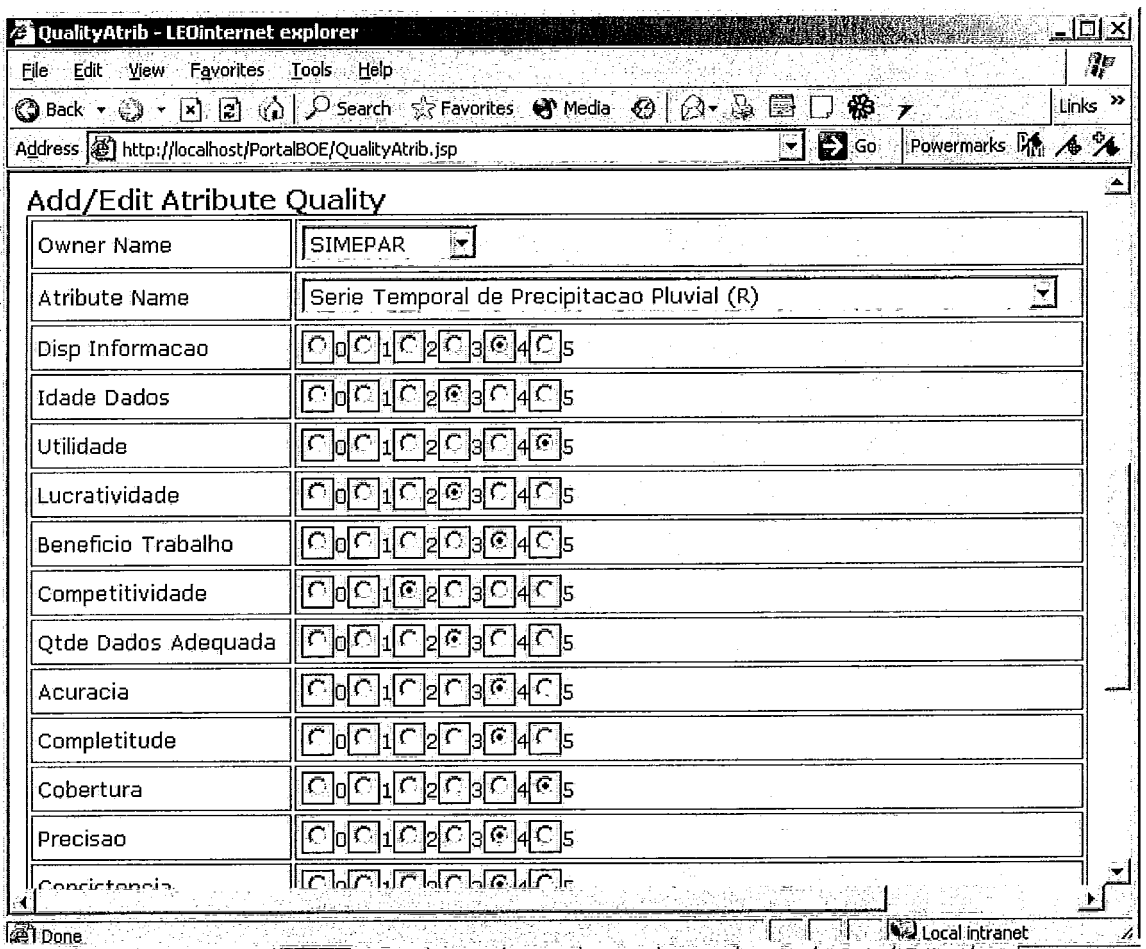


Figura 3.9 - Uso do Módulo de Gerência de Qualidade

A Tabela 3.2 descreve os principais problemas associados às características de qualidade selecionadas no estudo.

Tabela 3.2 - Principais Problemas Associados às Características de Qualidade

CARACTERÍSTICA	PROBLEMA ASSOCIADO
Disponibilidade da Informação	dado indisponível
Idade dos Dados	dados desatualizados
Utilidade	dado não é muito importante para a realização das atividades científicas
Lucratividade	dificuldade na utilização dos dados gerou queda de produtividade e perda financeira
Benefício no Trabalho do Usuário	utilização dos dados dificulta o trabalho dos usuários.
Competitividade	uso dos dados não trouxe vantagens competitivas
Quantidade de Dados Adequada	quantidade dados é insuficiente
Acurácia	manipulação do dado resulta em informação incorreta
Completitude	falta de campos e falta de informação
Cobertura	informação oferecida pelo dado é insuficiente
Precisão do Dado	o dado armazenado não representa corretamente o seu significado no mundo real
Consistência	mesmo dado com dois ou mais valores diferentes armazenados
Atributividade	ausência de registro de autoria dos dados
Flexibilidade	dificuldade na manipulação (expansão, adaptação ou agregação) dos dados
Interoperabilidade	impossibilidade de interagir com outras bases de dados.
Facilidade de Compreensão da Informação	informação pouco objetiva e gerando dúvidas
Uniformidade	apresentação dos dados insatisfatória
Disponibilidade de Documentação	falta de documentação

O resultado principal do módulo de gerência de qualidade é o estabelecimento de um valor numérico para a qualidade dos artefatos científicos. Este valor, em conjunto com os custos atribuídos e prazos de obtenção definidos pelos fornecedores, será utilizado pelo módulo de inferência para a indicação de melhores alternativas para a realização efetiva dos experimentos.

Este resultado é calculado com base na média aritmética da pontuação obtida em cada característica. A média aritmética foi escolhida pois, na indisponibilidade de um sistema de pesos apropriado, é o valor que melhor representa um conjunto de números. Quanto maior o número de avaliações feitas pelos usuários do BOE, mais precisa será a estimativa do valor da qualidade.

3.3.5 Módulo de Inferência

Conforme explicado na seção 3.2, conceitualmente o BOE é baseado no BOM. Para uma melhor compreensão desta proposta, estabelece-se uma analogia entre a preparação do experimento científico e a compra de itens para montagem de componentes.

Com base na Tabela 3.1 (seção 3.2) os próximos parágrafos descrevem termo a termo a analogia empregada.

- **Dados Iniciais (Item Simples ou Peça)** – no BOE ao invés de itens simples existem conjuntos de dados que poderão servir de base para a aplicação de algum modelo ou experimento científico. Pode-se entender por dados iniciais, constantes e medidas em seu menor *quantum* de informação, dados advindos de dispositivos de coleta, substâncias necessárias a um experimento químico etc. Exemplificando, cita-se o valor da gravidade, a tabela de elementos químicos, fórmulas, séries temporais de chuva em determinada região, amostras de sangue, dentre outros.
- **Dados Derivados (Item Composto ou Componente)** – consiste no conjunto de dados resultante da aplicação de algum modelo ou experimento científico sobre os dados iniciais ou sobre outros dados derivados.
- **Fornecedor (Empresa)** – assim como um fornecedor fabrica uma peça, os dados científicos podem ter uma origem ou um dono. Desta forma, instituições de pesquisa, empresas privadas e entidades governamentais podem ser representadas como “fornecedores” de massas de dados, artefatos ou recursos científicos. De maneira semelhante, vários fornecedores podem fabricar o mesmo item e várias organizações podem disponibilizar os mesmos dados científicos, ficando a cargo do “comprador” (pesquisador que realiza a procura dos dados) escolher qual “fornecedor” poderá entregar-lhe o “item” da maneira que melhor lhe convier, dependendo do custo, tempo de “entrega” e qualidade.
- **Custo (Preço)** – da mesma forma que o BOM, massas de dados científicos frequentemente apresentam custos, sejam eles de aquisição, levantamento, busca, descoberta ou disponibilização. A métrica de custo, então, tende a ser

mais rica no BOE que no BOM, ocasionando um problema ligeiramente mais complexo. No BOE as métricas de custo podem ser: o preço da disponibilização de um dado pertencente a uma instituição, pagamento de cotas de propriedade intelectual para o uso de uma informação, o custo de um levantamento ou pesquisa de campo, o custo do envio de dados ou o preço de consultoria de um pesquisador. Como no BOM, o menor custo de um componente consiste no somatório dos menores preços dos itens, nos quais os agregados formam o componente (Equação 1).

- **Prazo (Prazo)** – como no tópico anterior, a analogia é direta, pois da mesma maneira que um componente nem sempre se encontra disponível para envio imediato, massas de dados científicos podem não estar sempre disponíveis para consultas. Pode ser necessário aguardar o envio de informações via correio eletrônico ou postagem, a execução completa de um modelo ou o tempo da realização de um experimento para coleta de seus resultados. O prazo de disponibilidade de um dado derivado (análogo a componente) sempre é igual ao maior prazo de entrega dentre todos os itens envolvidos que sejam também dados, acrescido do prazo de execução do modelo ou experimento (Equação 2).
- **Qualidade** – métrica não presente no BOM e implementada neste trabalho através do módulo de gerência de qualidade. A qualidade de um dado inicial consiste na média aritmética dos critérios de qualidade associados pelos pesquisadores. Para dados derivados, novamente será utilizada a média, mas desta vez ela será composta pelas qualidades dos diversos dados de origem de um modelo ou experimento (Equação 3).

Equação 1 - Custo Calculado para um Dado Derivado

$$custo(DadoDerivado) = \sum_{\forall X \in DadoInicial, DadoDerivado, Modelo} \min[custo(X)]$$

Equação 2 - Prazo Calculado para um Dado Derivado

$$prazo(DadoDerivado) = \left(\max_{\forall X \in DadoInicial, DadoDerivado} [prazo(X)] \right) + \sum_{\forall Y \in Modelo} prazo(Y)$$

Equação 3 - Qualidade Calculada para um Dado Derivado

$$qualidade(DadoDerivado) = \frac{\sum_{i=1, \forall Xi \in DadoInicial, DadoDerivado, Modelo}^n qualidade(Xi)}{n}$$

Tal como em (SULAIMAN, 1999), o armazenamento dos dados científicos e seus relacionamentos é realizado em um ambiente dedutivo. O módulo de inferência é capaz de apontar melhores alternativas para a realização de experimentos através de técnicas de IA, abrangendo CLP (JAFFAR & MAHER, 1994) e heurísticas (DASGUPTA, CHAKRABARTI et al., 2002). Por melhores alternativas, entendem-se os menores preços e prazos de entrega e a melhor qualidade de um artefato científico. O BOE armazena os experimentos como *workflows* abstratos em que cada atividade é mapeada em regras de produção (HAYES-ROTH, 1985) da forma “**IF** (cond1 & cond2) **THEN** Resultado”.

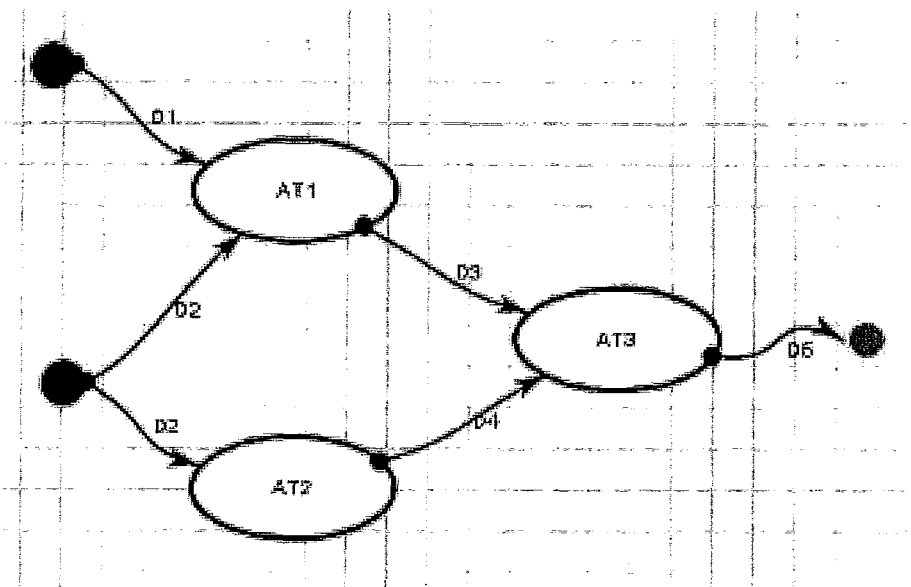


Figura 3.10 - Exemplo de Atividades de um *Workflow*

Para as atividades do *workflow* fictício descrito na Figura 3.10, as seguintes regras são obtidas (em notação Prolog simplificada):

- D5 :- D3, D4, AT3.
- D4 :- D2, AT2.
- D3 :- D1, D2, AT1.

Nesta modelagem, cada dado de saída depende dos dados de entrada e da correta execução da etapa. Por exemplo, o dado “D5” é obtido se existir disponibilidade dos dados “D3” e “D4”, e da atividade “AT3”. Por sua vez, “D4” depende da existência de “D2” e da atividade “AT2”. É perceptível que o algoritmo de “encadeamento para trás” possibilita a inferência da viabilidade da execução de um experimento acompanhada de um cálculo de melhores alternativas.

Em certos casos, o planejamento de utilização de um artefato científico inicial ou derivado não irá esvair o recurso, ou seja, certos recursos, uma vez obtidos, poderão ser reutilizados indefinidamente em novos planejamentos, com custo e prazo iguais a zero. Este comportamento normalmente não é observado no BOM.

Uma vez que os *workflows* abstratos são mapeados em regras de produção, o módulo de inferência utiliza uma máquina de inferência (por exemplo, qualquer ambiente PROLOG) para: i) verificar a viabilidade da execução de um experimento, garantindo que todos os artefatos necessários para a realização estejam presentes e sejam fornecidos por algum fornecedor; ii) gerar as melhores alternativas para a realização dos experimentos, incluindo as explicações sobre as decisões tomadas durante o processo de inferência. Neste momento, a representação de um *workflow* é feita de maneira declarativa e não mais procedimental.

Como vantagem desta técnica em relação a métodos de otimização matemática, tem-se: i) a utilização de regras é bastante intuitiva para o ser humano; ii) o processo de inferência gera explicações mais simples a respeito das decisões que está tomando; iii) como normalmente os protocolos científicos não tendem a apresentar centenas de atividades, o desempenho deste tipo de técnica não se torna um problema (vide exemplos de utilização do BOE no capítulo 4).

Os resultados do módulo de inferência são passados para os portais em formato XML. Esta iniciativa permite que a exibição dos resultados seja simples, pois a maioria dos navegadores atuais traz facilidades na apresentação de arquivos XML. Além disso, é interessante que alguma ferramenta que efetivamente instancie um *workflow* científico seja informada do planejamento definido no BOE. Assim, o arquivo XML criado com os resultados pode ser importado por alguma ferramenta de

acompanhamento de *workflows* científicos. Um trecho de resultado gerado pelo módulo está exposto como exemplo na Figura 3.11 e será melhor explicado na seção 4.3.

Para a implementação dos resultados em XML foi especificado o seguinte DTD:

```
<!-- BOE DTD -->
<!ELEMENT Workflow (#PCDATA | ControlFile | ObjectiveArtifact |
MinCost | MinDelay | MaxQuality)*>
<!ELEMENT MinCost (CompositeArtifact*, SimpleArtifact*, Cost, Delay,
Quality)>
<!ELEMENT MinDelay (CompositeArtifact*, SimpleArtifact*, Cost, Delay,
Quality)>
<!ELEMENT MaxQuality (CompositeArtifact*, SimpleArtifact*, Cost,
Delay, Quality)>
<!ELEMENT CompositeArtifact (Name, Composition)>
<!ELEMENT Composition (CompositeArtifact*, SimpleArtifact*, Cost,
Delay, Quality)*>
<!ELEMENT SimpleArtifact (#PCDATA | Institution)*>
<!ELEMENT Institution (Name, Cost, Delay, Quality)>
<!ELEMENT Name (#PCDATA)>
<!ELEMENT ObjectiveArtifact (#PCDATA)>
<!ELEMENT ControlFile (#PCDATA)>
<!ELEMENT Cost (#PCDATA)>
<!ELEMENT Delay (#PCDATA)>
<!ELEMENT Quality (#PCDATA)>
```

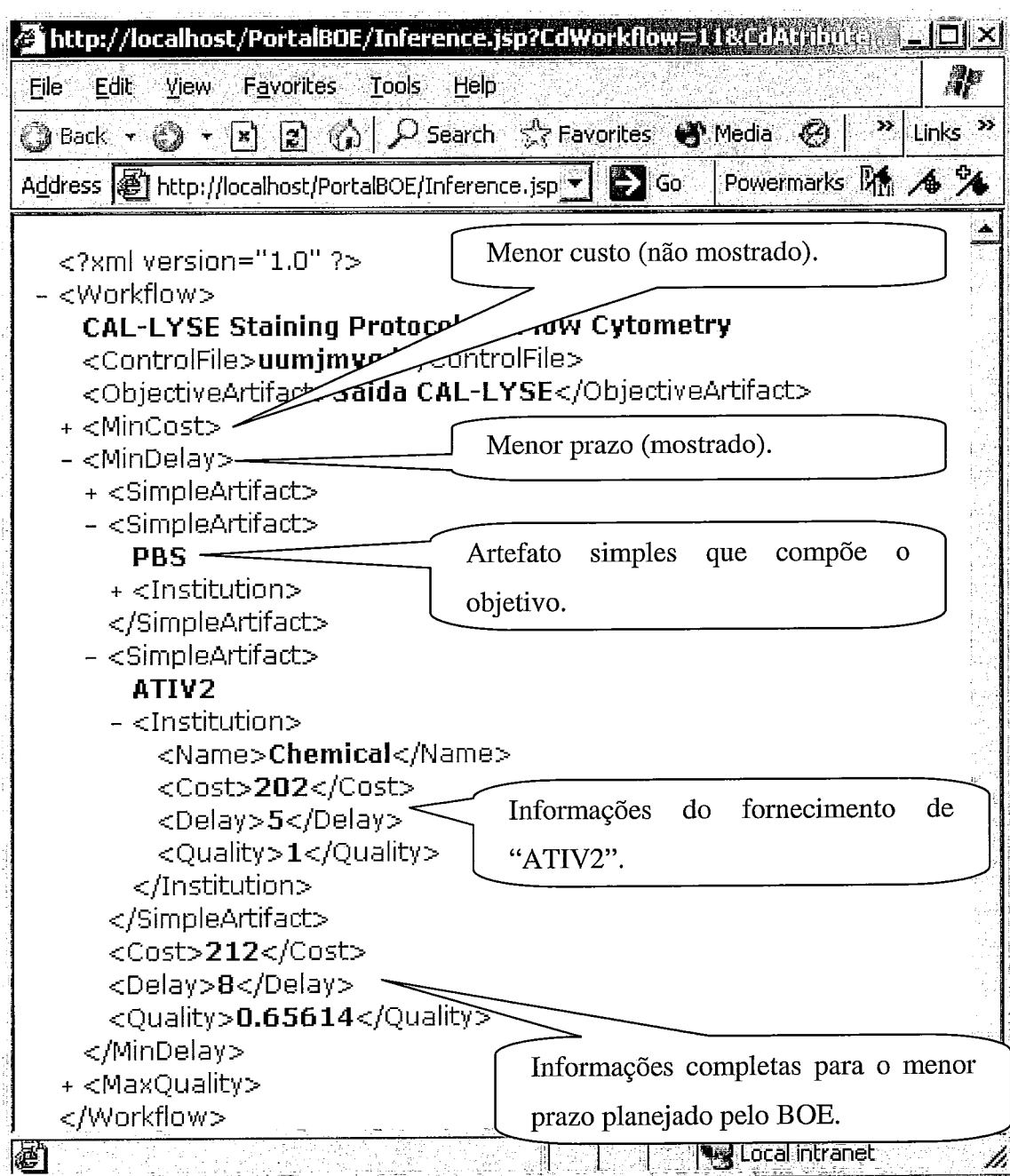


Figura 3.11 – Exemplo de Resultado do Módulo de Inferência em XML

3.3.6 Módulo de Busca

As buscas por artefatos científicos são implementadas pelo módulo de busca na arquitetura do BOE. Este módulo é fundamental para a promoção da colaboração como um todo no ambiente, uma vez que permite a identificação e interação dos participantes bem como a disseminação do conhecimento científico inerente aos experimentos científicos.

Para esta dissertação, este módulo está proposto de uma forma simples, ou seja, somente buscas por palavras-chave são possíveis. Não foram utilizados algoritmos de recuperação da informação (*information retrieval*).

No BOE, é possível buscar artefatos (dados, modelos, experimentos etc) científicos, protocolos científicos definidos em *workflows*, e pesquisadores e fornecedores específicos, dentro de determinada área de pesquisa (agricultura, biologia, química, etc).

3.3.7 Camada de Banco de Dados

A camada de banco de dados é responsável por armazenar os experimentos explicitados pelos pesquisadores, as informações dos artefatos disponibilizados pelos fornecedores e as avaliações de qualidade feitas pelos usuários destes artefatos. Ela está logicamente definida em três partes: fornecedores, experimentos e qualidade.

O banco de dados de fornecedores consiste no registro dos fornecedores do ambiente BOE. A partir deste banco de dados é possível a obtenção das cotações de custos e prazos dos artefatos científicos dadas por cada fornecedor participante. Ainda existe um histórico detalhando as cotações existentes ao longo do tempo.

O banco de dados de experimentos é um banco de dados científicos para gerência de experimentos (vide seção 2.1.1.1). Nele é possível o armazenamento de um *workflow* científico detalhando suas atividades, artefatos científicos envolvidos e responsáveis. Os usuários deste banco de dados são os próprios pesquisadores que têm uma fonte de informação para análise e planejamento de experimentos correlatos.

Finalmente, o banco de dados de qualidade tem a responsabilidade de armazenar as cotações das métricas de qualidade existentes no módulo de gerência de qualidade. A consulta eficiente deste banco de dados é importante para a rápida geração dos valores quantitativos de qualidade para um artefato científico de determinado fornecedor.

Os bancos de dados projetados e utilizados no BOE residem na camada de repositório de conhecimentos do ambiente SPeCS, introduzida no capítulo 1. Os bancos de dados de fornecedores e de qualidade fazem parte dos repositórios de dados convencionais. O banco de dados de experimentos está associado aos repositórios de

workflow, modelos e regras. Assim, a ferramenta **X-Arc** (PINTO, SOUZA et al., 2001) tem a capacidade de prover serviços básicos de integração dos dados do BOE para os outros componentes do ambiente.

3.3.8 Implementação da Arquitetura

Os portais foram implementados utilizando-se JSP (*JavaServer Pages*) (SUN, 2003b). Esta tecnologia de geração de páginas dinâmicas é de fácil utilização e é adotada nos componentes principais do ambiente Agromet.

Uma página JSP nada mais é do que uma página *web* normal que contém alguns trechos de códigos em Java, responsáveis por gerar conteúdo dinâmico, dentro de uma estrutura HTML, separando a lógica de uma aplicação da sua apresentação. Como componente da tecnologia Java, o JSP é amplamente suportado e desta forma não prende uma solução desenvolvida a um sistema operacional ou servidor *web* específico.

Através da criação de classe Java específicas, e da inclusão destas nas páginas JSP, foi possível que o portal desenvolvido tivesse capacidades de comunicação com os outros módulos que residem na camada de controle da arquitetura do BOE.

O módulo de projeto do *workflow* deve cumprir com um dos requisitos primordiais da ferramenta BOE, que é seu funcionamento completo em ambiente *web*. Para atingir este objetivo, optou-se pela implementação de um *applet* (SUN, 2003a). Um *applet* é um programa escrito em linguagem Java que pode ser incluído em uma página *web* da mesma forma que uma imagem, ou seja, através de *tags* específicas. No momento em que é utilizado um navegador capaz de interpretar Java (atualmente a maioria dos *browsers* é capaz), o código do *applet* é transferido para a máquina cliente e lá é executado pela máquina virtual Java.

Um outro requisito importantíssimo é que o módulo de projeto do *workflow* seja gráfico. Para implementação da parte gráfica do editor de *workflows* foi utilizada ferramenta Grace (KLEIN, 1999a; KLEIN, 1999b).

A ferramenta Grace é um gerador de código fonte Java para a criação do núcleo de editores gráficos para estruturas baseadas em grafos (contendo nós e arcos). Com o uso de uma linguagem de especificação, é possível definir o comportamento do editor gráfico de uma forma abstrata, e, após a compilação, as classes Java geradas

podem ser integradas a um projeto em execução. Esta linguagem de especificação é estruturada em seções: especificação do domínio do problema, controle de diálogo, apresentação e estilos de edição.

Após o estudo e entendimento da linguagem de especificação de editores do Grace, foi possível gerar o núcleo do editor de *workflows* do BOE, implementado na sua totalidade em um *applet* Java. O apêndice B descreve o modelo de classes do domínio do problema utilizado para o projeto do módulo de projeto do *workflow*.

O módulo de gerência de qualidade é implementado em classes Java que interagem com os portais JSP do BOE. Nestas classes, existem funcionalidades que permitem o correto funcionamento do módulo, tais como: i) cálculo da qualidade atual de um determinado dado científico (para o módulo de inferência); ii) criação de uma nova “entrada” de pontuação quando um usuário atribui pontos a um dado científico; e iii) acompanhamento, pelo fornecedor, das pontuações que vem recebendo ao longo do tempo.

O módulo de inferência teve como primeiro esforço de implementação a criação de uma máquina de inferência com as funcionalidades básicas utilizando a ferramenta *Microsoft Access* e a linguagem VBA (*Visual Basic for Applications*). Este protótipo visou a validação do conceito do BOE e conseguiu integrar regras e máquina de inferência em um único ambiente (SOUZA, CARDOSO et al., 2001).

Após a validação, surgiu a necessidade da utilização de recursos mais sofisticados existentes em uma máquina de inferência, tais como: adição e remoção de regras durante a execução de uma inferência, tratamento eficiente de *backtracking* e, velocidade de processamento. Assim, a integração obtida com o protótipo teve que ser refeita e optou-se por utilizar um ambiente Prolog completo, o *SWI-Prolog* (SWI, 2003). O esquema para a realização das inferências no BOE está descrito na Figura 3.12.

Neste novo esquema para a realização das inferências, utilizou-se uma modelagem de dados relacional para a representação dos *workflows* científicos. A tradução da representação relacional para a representação por regras, facilmente acoplável a uma máquina de inferência comum, foi feita via agentes de tradução (classes Java chamadas a partir do ambiente JSP) existentes no módulo de inferência.

Com essa iniciativa, os bancos de dados convencionais do BOE podem ser facilmente consultados por outras ferramentas do ambiente SPeCS.

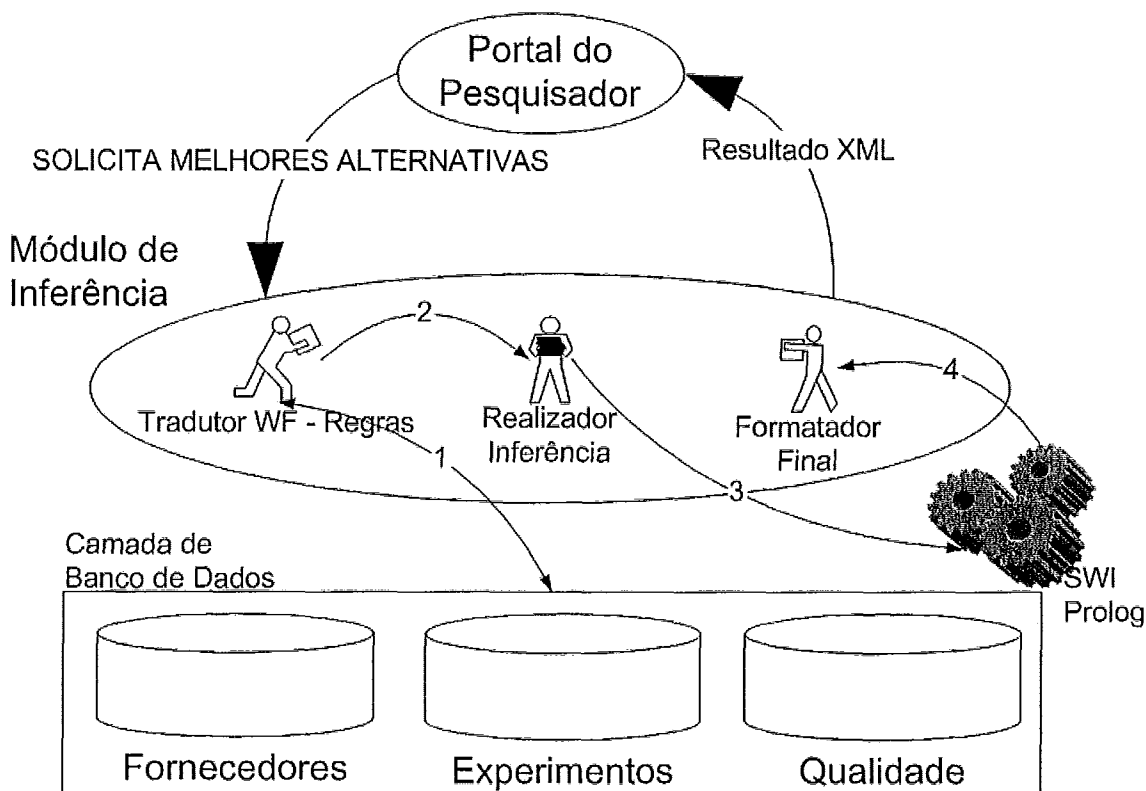


Figura 3.12 - Esquema Utilizado para Realização das Inferências

Na prática, inicialmente o portal do pesquisador irá solicitar ao módulo de inferência uma busca por melhores alternativas para um determinado artefato de um *workflow*. Tão logo o módulo de inferência receba a requisição, este irá acionar um tradutor que irá consultar a camada de banco de dados e criará um arquivo texto contendo a representação de regras do *workflow* desejado (1) e o passará para o agente de realização de inferências (2). De posse deste arquivo, o agente irá chamar o *SWI-Prolog* (3) e aguardará seu encerramento. Os resultados gerados pelo ambiente Prolog são repassados a um agente de formatação final (4), que efetua os últimos ajustes no XML antes do módulo de inferência retorná-lo para o portal solicitante.

É importante ressaltar que os usuários não necessitam da instalação do ambiente Prolog, todo o trabalho é realizado no servidor do BOE. Um exemplo de arquivo texto criado pelo módulo de inferência para execução pelo *SWI-Prolog* se encontra no apêndice C.

Conforme os outros módulos da camada de controle, a implementação do módulo de busca se deu com a criação de classes Java. Tais classes, invocadas pelos portais implementados em JSP, são responsáveis pela criação de consultas que são submetidas para a camada de banco de dados, retornando os resultados apropriados. Todas as funções de busca nos portais são implementadas desta forma.

Confirmando a proposta inicial simplificada deste módulo, as consultas por palavras-chave submetidas utilizam o operador “*LIKE*” da linguagem SQL. Assim, uma busca pela palavra-chave “série” irá gerar uma consulta com o operador “*LIKE* ‘%série%’” na tabela apropriada.

É importante observar que o módulo de projeto do *workflow* também utiliza o módulo de busca no momento da edição de uma transição, para encontrar determinados artefatos que estarão associados à transição em edição (Figura 3.7).

Os bancos de dados desta camada usam o modelo relacional e foram implementados primeiramente nas ferramentas MS Access e SQL Server 2000 da Microsoft. O esquema lógico dos bancos de dados da ferramenta BOE está no apêndice D.

Capítulo 4 - Estudo de Caso e Exemplos de Aplicação

Dada a definição da arquitetura da ferramenta BOE bem como detalhes referentes a sua implementação no capítulo anterior, surge a necessidade do entendimento das funcionalidades do sistema em situações reais.

Neste capítulo será apresentado um estudo de caso com um *workflow* científico da área de agrometeorologia e exemplos aplicados de utilização da ferramenta BOE em outras áreas da ciência.

4.1 Estudo de Caso na Agrometeorologia

Esta seção apresenta um estudo de caso das capacidades do BOE. As atividades do módulo de inferência serão descritas com base na realização de um zoneamento agroclimático. Para testes foram designados fornecedores genéricos que associam diferentes custos, prazos e qualidades aos artefatos utilizados neste exemplo.

A competitividade do setor agrícola é determinada pela produtividade, qualidade do produto e diversidade da produção. A qualidade ambiental na produção agrícola afeta diretamente a produtividade e a qualidade do produto final. Deste modo, nas próximas décadas, o desenvolvimento sustentado estará estreitamente relacionado com a competitividade do setor (CARDOSO, SOUZA et al., 2002b).

Desta maneira, tendo como objetivo possibilitar o conhecimento do potencial das terras para uma determinada cultura, visando a conciliação da produção e a conservação ambiental, é elaborado, pela Embrapa Solos, juntamente com instituições parceiras, o zoneamento agrícola que necessita, entre as bases de informações, de um zoneamento agroclimático de uma região. Este tem por finalidade identificar as potencialidades climáticas para uma cultura, considerando-se anos médios e anos extremos, com vista a uma melhor planificação da assistência técnica, pesquisa e experimentação agrícola, e, servir como instrumento básico de orientação para a formulação de políticas de desenvolvimento agrícola.

O *workflow* científico para a geração de um zoneamento agroclimático de uma determinada região foi levantado a partir de reuniões com uma especialista na área de agrometeorologia da Embrapa Solos (CARDOSO, SOUZA et al., 2002b).

Na fase inicial do processo de zoneamento existe uma busca por fontes de dados meteorológicos na região em estudo. Normalmente, esta consulta é feita junto às instituições que possuam estações de coleta de informações na região. Como exemplos de tais instituições, citam-se o INMET (Instituto Nacional de Meteorologia) e a ANEEL (Agência Nacional de Energia Elétrica).

Os dados meteorológicos que podem ser utilizados na criação do zoneamento são compostos por séries temporais de variáveis como: temperatura do ar (mínima, máxima e média, em °C); temperatura do solo (mínima, máxima e média, em °C); precipitação pluvial (mm); umidade relativa do ar (%); vento ($m \cdot s^{-1}$); insolação (horas de brilho solar); radiação solar ($W \cdot m^{-2}$). No momento em que há a necessidade de aquisição externa desses dados, em outra instituição parceira, o BOE indica as melhores alternativas para adquirir tais informações. Quando não se dispõe de todos esses dados meteorológicos, o zoneamento é realizado em função dos dados existentes.

As etapas do *workflow* do zoneamento agroclimático são as seguintes:

a) Tratamento dos Dados

Esta fase consiste inicialmente na adequação dos dados adquiridos na fase inicial. Tal adequação é composta de padronização de formato (conversão de unidades de medidas, seleção de séries históricas, e variáveis meteorológicas disponíveis) e de adequação de conteúdo (como por exemplo, o cálculo de grandezas a partir de outras).

O passo seguinte desta fase consiste na separação de séries homogêneas e da observação da consistência dos dados coletados.

b) Tratamento Estatístico dos Dados

Nesta etapa existe o processamento estatístico dos dados obtidos até então. Fundamentalmente é composto de cálculos estatísticos e matemáticos, como avaliação da distribuição, média, desvio padrão, coeficiente de variação, eventos isolados de máximos e mínimos, análise de frequência dos eventos e outras análises das séries temporais. A tarefa descrita pode ser desenvolvida com o auxílio de ferramentas, tais como, o CLIMLAB (TANCO, 2000) e o SAS (SAS INSTITUTE INC, 1999).

Como resultado desta etapa, têm-se os dados das variáveis meteorológicas com tratamento estatístico, o que é essencial para a criação de mapas de isolinhas em uma ferramenta de SIG.

c) Balanços Hídricos

Esta etapa consiste na aplicação de um modelo matemático para a criação dos balanços hídricos na região em estudo. Três tipos de modelagem podem ser utilizados: por cultura, seqüencial e normalizado.

Em todos os tipos, é necessária a entrada de variáveis como: chuva, temperatura média do ar, horas de brilho solar, coordenadas geográficas, CAD (capacidade de água disponível) e coeficientes das culturas, conforme a especificidade do modelo. Dependendo da disponibilidade de dados diários, considera-se a precipitação pluvial diária, estima-se a evapotranspiração potencial diária, coeficiente cultural, ciclo e fases fenológicas da cultura. De um modo geral, as saídas dos balanços hídricos, apresentam grandezas como: evapotranspiração potencial, evapotranspiração real, excedente hídrico e deficiência hídrica.

d) Análise Espacial

Até este momento, foi feita a análise temporal dos dados. A partir dessa fase se inicia a análise espacial dos dados. Nesta etapa são gerados os mapas de isolinhas da região tratada. Como entrada há os valores de variáveis necessárias à geração do mapa e as coordenadas de estações meteorológicas. Dependendo da cultura analisada, é gerado um mapa para de cada variável. Para a geração dos mapas de isolinhas, algumas ferramentas podem ser utilizadas, como softwares SIG (ILVIS) e o CLIMLAB.

e) Integração de Variáveis

Cada cultura possui necessidades agroclimáticas específicas, ou seja, em que situação seu crescimento é favorecido ou diante de quais condições climáticas ela pode sofrer *stress* térmico ou hídrico. A realização de um zoneamento agroclimático para uma determinada cultura, leva em consideração as peculiaridades inerentes ao seu sistema de cultivo. Realiza-se o zoneamento em função das condições em que a cultura se torna mais vulnerável as condições climáticas, por exemplo deficiência hídrica na fase de desenvolvimento ou excedentes hídricos na colheita, com base nos resultados das simulações dos balanços hídricos. Tais necessidades já são dados prontos, advindos de outros estudos, principalmente, da área de ecofisiologia vegetal.

Nesta etapa as informações dos mapas de isolinhas e as necessidades agroclimáticas das culturas são cruzadas para a criação do mapa de aptidão agroclimática, que indica as melhores regiões para o plantio de culturas, baseado nas informações climáticas. O mapa de aptidão agroclimática gera áreas com características semelhantes.

f) Espacialização de Variáveis em Planos de Informação

Com o auxílio de ferramentas de geoprocessamento os mapas de aptidão agroclimática são transformados nos chamados planos de informação da região estudada. Os planos de informação apresentam as zonas agroclimáticas.

A representação gráfica deste *workflow* no BOE se encontra nas Figura 4.1 e Figura 4.2. O produto final de interesse é o “plano de informação” (Figura 4.3). Entretanto, novamente os resultados intermediários, como os “mapas de isolinhas” e a “evapotranspiração” podem ser de grande utilidade para o projeto de outros trabalhos semelhantes. É interessante ainda frisar que todo o processo do zoneamento é explicitado pelo especialista da área e ocasionalmente poderá servir de base para a realização de novos processos semelhantes.

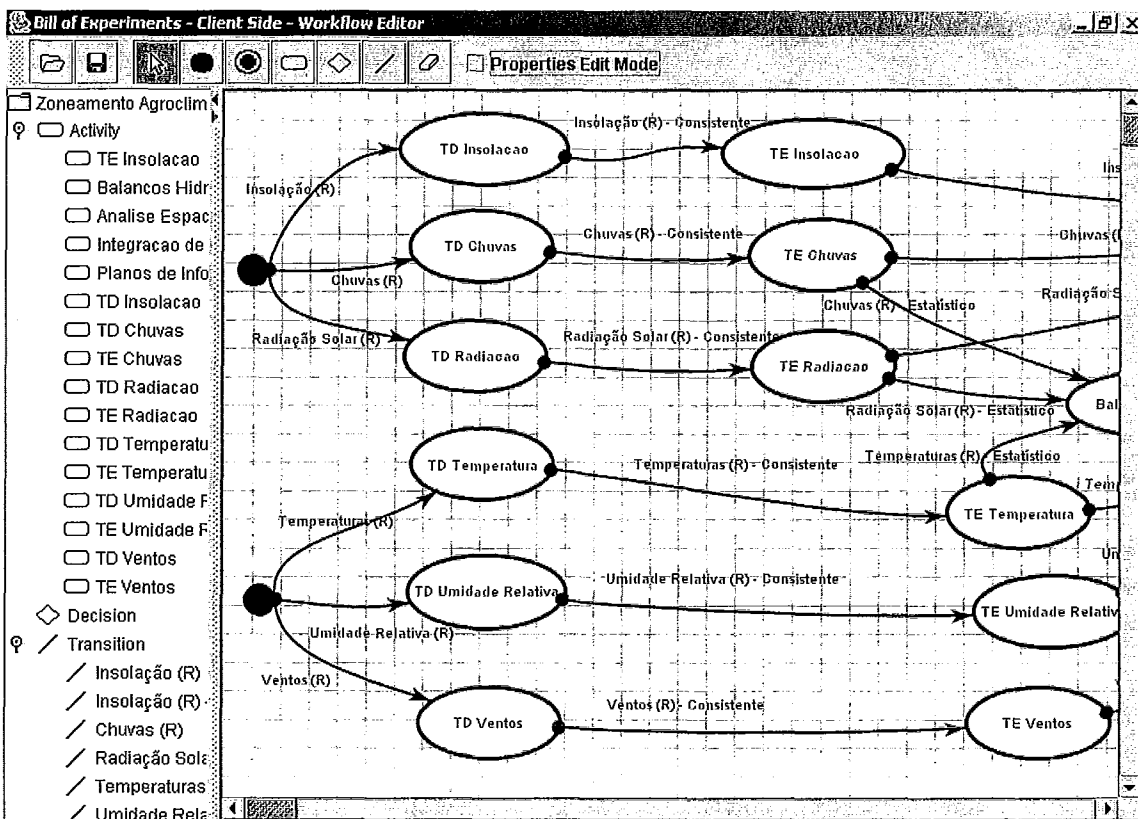


Figura 4.1 - Representação Gráfica do *Workflow* de Zoneamento Agroclimático (Parte 1)

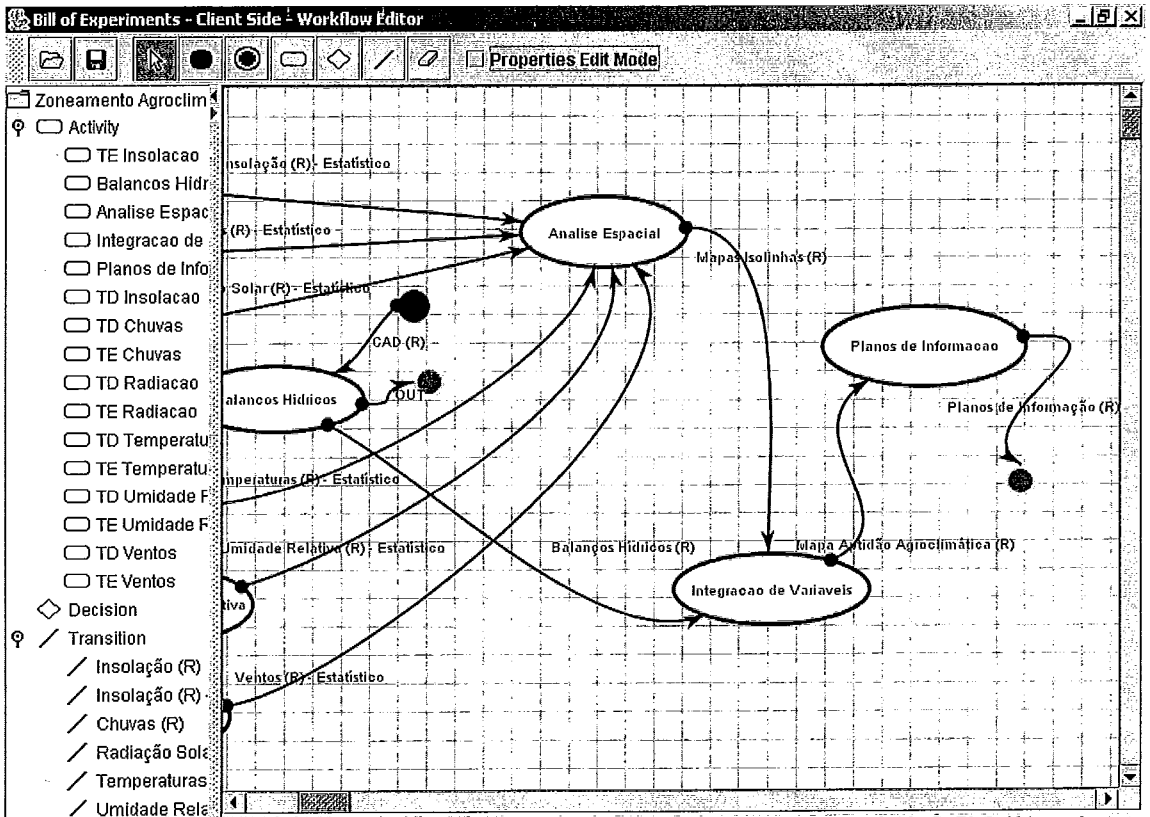


Figura 4.2 - Representação Gráfica do Workflow de Zoneamento Agroclimático (Parte 2)

```

http://localhost/PortalBOE/Inference.jsp?CdWorkflow=1&CdAttribute=68 - LEOinternet explorer
File Edit View Favorites Tools Help
Back Search Favorites Media Links Google
Address http://localhost/PortalBOE/Inference.jsp?CdWorkflow=1&CdAttribute=68 Go Powermarks
<?xml version="1.0" ?>
- <Workflow>
  Zoneamento Agroclimático
  <ControlFile>ggytxrff</ControlFile>
  <ObjectiveArtifact>Planos de Informacao (R)</ObjectiveArtifact>
  - <MinCost>
    - <CompositeArtifact>
      <Name>Mapa Aptidao Agroclimatica (R)</Name>
      - <Composition>
        - <CompositeArtifact>
          <Name>Mapas Isolinhas (R)</Name>
          - <Composition>
            - <CompositeArtifact>
              <Name>Serie Temporal ESTATISTICA de Insolacao (R)</Name>
              - <Composition>
                - <CompositeArtifact>
                  <Name>Serie Temporal CONSISTENTE de Insolacao (R)</Name>
                  - <Composition>
                    - <SimpleArtifact>
                      Serie Temporal de Insolacao (R)
                      + <Institution>
                        </SimpleArtifact>
                      + <SimpleArtifact>
                        <Cost>120</Cost>
                        <Delay>4</Delay>
                        <Quality>0.5</Quality>
                      </Composition>
                    </SimpleArtifact>
                  </Composition>
                </CompositeArtifact>
              </Composition>
            </CompositeArtifact>
          </Composition>
        </CompositeArtifact>
      </Composition>
    </MinCost>
  </Workflow>
  
```

Figura 4.3 - Resultado do Planejamento para o Zoneamento, mostrando explicações

Um pesquisador pode interpretar os resultados descritos na Figura 4.3 através da noção de derivação de dados científicos. Para o objetivo (“Planos de Informação (R)”) ser obtido com menor custo, é necessária a obtenção do artefato “Mapa Aptidão Agroclimática (R)”. Por sua vez, para a obtenção do mapa é necessária a construção do “Mapa Isolinhas (R)”. O mapa de isolinhas necessita de séries temporais de diversas variáveis climáticas. Percebe-se que quando se atinge o fim da composição de dados, existem apenas artefatos iniciais, que apresentam custo, prazo e qualidade bem definidos (ex.: “Série Temporal de Insolação (R)”, “Série Temporal de Chuvas (R)” e outras). No exemplo acima, o artefato inicial “Série Temporal de Insolação (R)” apresenta um custo de R\$ 120,00, um prazo de 4 dias para disponibilidade e uma qualidade estimada em 50%.

O fragmento de documento XML exibido na figura acima representa somente uma pequena parte do planejamento gerado pelo BOE para o estudo de caso. O documento inteiro é de grande valia para o pesquisador, para que ele consiga avaliar o custo-benefício do experimento que está prestes a realizar.

4.2 Exemplo de Aplicação 1 – Problemas Inversos

Um modelo matemático para problemas inversos é empregado para ilustrar uma simulação. Por problemas inversos entendem-se aqueles que estão relacionados com a obtenção de características ou parâmetros de modelos matemáticos com base no conhecimento reduzido das soluções das correspondentes equações.

Um exemplo típico de problema inverso na área de prospecção sísmica é:

“Determinação de dados de reservas petrolíferas, ou mais geralmente de dados geológicos, com base em medidas superficiais de dados de reflexão ou de espalhamento de ondas.” (ZUBELLI, 1999)

Observa-se que neste caso a aplicação de métodos distintos sobre duas massas de dados diferentes provê resultados cientificamente similares. Logo, este exemplo se encaixa perfeitamente na estrutura do BOE.

Desta forma, dois experimentos científicos simples dão origem a dois resultados experimentais semelhantes, ambos os protocolos estão demonstrados como um só *workflow* na Figura 4.4.

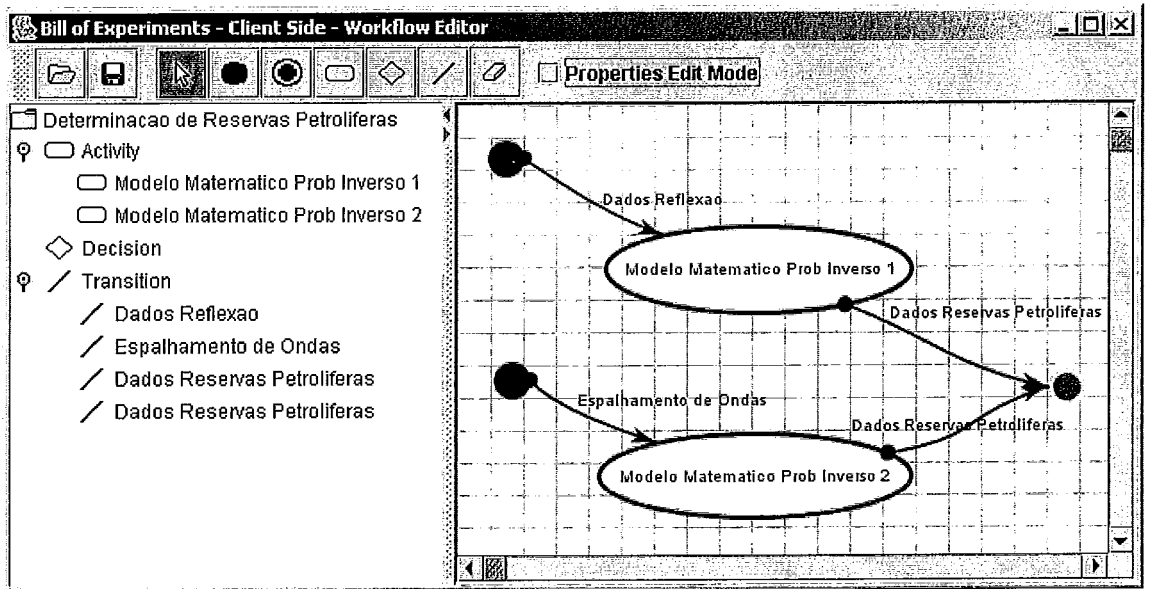


Figura 4.4 - Workflow para Determinação de Dados de Reservas Petrolíferas

Neste exemplo, um pesquisador interessado em dados sobre reservas petrolíferas poderá obter seus resultados através da aplicação de diferentes simulações, conforme demonstrado acima. A ferramenta BOE irá disponibilizar para o usuário o planejamento para a execução do experimento com mínimos custo e prazo e máxima qualidade no formato XML (Figura 4.5).

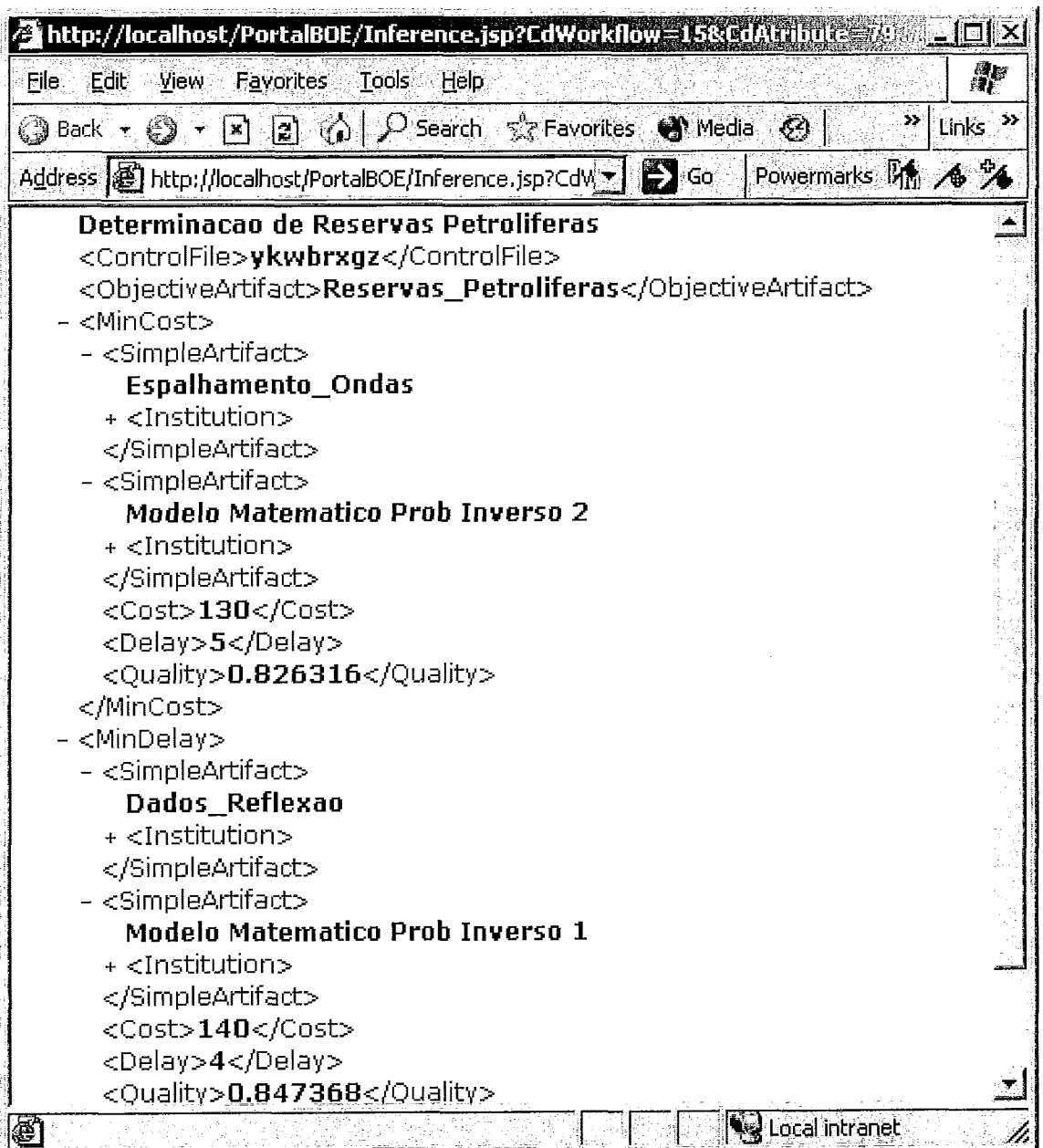


Figura 4.5 - Resultado do Planejamento do *Workflow* para Determinação de Dados de Reservas Petrolíferas

É interessante reparar que para obtenção dos dados de reservas petrolíferas com menor preço, o módulo de inferência optou pelo planejamento utilizando o “Modelo Matemático Prob Inverso 2” e os dados de espalhamento de ondas. Já para o menor prazo, a opção foi pelo o “Modelo Matemático Prob Inverso 1” e os dados de reflexão sísmica. A alternativa menos demorada é mais cara e apresenta uma qualidade ligeiramente maior. Os valores apresentados neste exemplo são fictícios.

4.3 Exemplo de Aplicação 2 – Área Biológica

A citometria de fluxo é uma metodologia de análise aplicável em diversas áreas de pesquisa, tais como botânica, zoologia, genética, microbiologia, medicina, além de outras. Ela permite o estudo da variabilidade genômica pelo perfil colorimétrico do conteúdo de DNA de uma célula. Essa metodologia é utilizada em pesquisas com aplicações práticas no melhoramento de plantas, principalmente na análise de protoplastos, grãos de pólen, núcleos e cromossomos (SARAIVA, 2003).

Células individuais, ou partículas em geral, passam seqüencialmente em um fluxo contínuo, em frente a um feixe laser com sensores para medir fluorescência e dispersão de luz. Um “citômetro de fluxo” pode medir estes parâmetros a uma taxa de milhares de células por minuto.

4.3.1 Consultando *Workflows* Alternativos

Existem diversos protocolos científicos que podem ser empregados para a preparação de um experimento que envolva citometria de fluxo. Consultando um *site* especializado (LABVELOCITY, 2003), foi possível a identificação de diversos métodos empregados para a utilização da metodologia.

Em um caso prático, seria o BOE a fonte de consulta utilizada por pesquisadores para encontrar *workflows* empregados para a citometria de fluxo. Assim, todos os protocolos referentes a esta pesquisa teriam sido explicitados e definidos através do módulo de projeto do *workflow*.

Dadas estas premissas, através de um cadastramento manual é possível demonstrar a funcionalidade do módulo de busca, na pesquisa por *workflows* para citometria de fluxo (Figura 4.6).

Workflow Name	Text Description	Scientific Area	Creation Date	Last Update Date	Create User
CAL-LYSE Staining Protocol for Flow Cytometry	CAL-LYSE Staining Protocol for Flow Cytometry - MINI	Biology	19/11/02 15:08	01/02/03 00:00	manuel
Detection of Intracellular Antigens by Flow Cytometry	CALTAG Laboratories- Detection of Intracellular Antigens by Flow Cytometry	Biology	19/11/02 15:08	19/11/02 15:08	manuel
Flow Cytometry Method Direct	Direct Method for Flow Cytometry	Biology	09/11/02 16:21	14/11/02 15:18	cardoso
Whole Blood Protocol for Intracellular Cytokines by Flow Cytometry	CALTAG Laboratories- Whole Blood Protocol for Intracellular Cytokines by Flow Cytometry	Biology	19/11/02 15:07	24/01/03 00:00	manuel
Zymed's Flow Cytometry	Flow Cytometry Methods from Zymed - Including Direct & Indirect	Biology	14/11/02 15:20	14/11/02 00:00	cardoso

Figura 4.6 - Resultado da Busca por *Workflows* para Citometria de Fluxo

Com o resultado da busca, o pesquisador pode estudar, interpretar e até mesmo reutilizar os diferentes protocolos existentes no banco de dados de experimentos.

4.3.2 Melhores Alternativas para a Citometria de Fluxo

Dentre os *workflows* pesquisados na seção anterior, um foi selecionado para um estudo mais detalhado e para a geração hipotética de melhores alternativas. As etapas de dois métodos para a citometria de fluxo (ZYMED LABORATORIES INC., 2000) são descritas em uma lista no próximo parágrafo. Ambos os métodos têm as mesmas etapas iniciais e finais e diferem na etapa de “Marcação de Células” (“*Cell Staining Procedure*”) que pode ser realizada direta ou indiretamente. Os termos científicos foram colocados com a finalidade ilustrativa, ou seja, não é intenção desta dissertação discutir a respeito de detalhes dos métodos de citometria de fluxo.

1. Preparação de Células (*Preparation of Cells*)
2. Isolamento de Células Mononucleares (*Mononuclear Cell Isolation*)

3a. Marcação de Células (Direta) (*DIRECT Cell Staining Procedure*)

3b. Marcação de Células (Indireta) (*INDIRECT Cell Staining Procedure*)

4. Análise da Citometria de Fluxo (*Flow Cytometric Analysis*)

A representação gráfica deste *workflow* no BOE está na Figura 4.7. Pode-se notar que o produto final de interesse é o “histograma de citometria de fluxo” (“*plotted histogram*”) (Figura 4.8). Entretanto, resultados intermediários, como “células preparadas” (“*prepared cells*”) (Figura 4.9) ou amostras de sangue (“*blood samples*”), podem ser empregados em estudos similares.

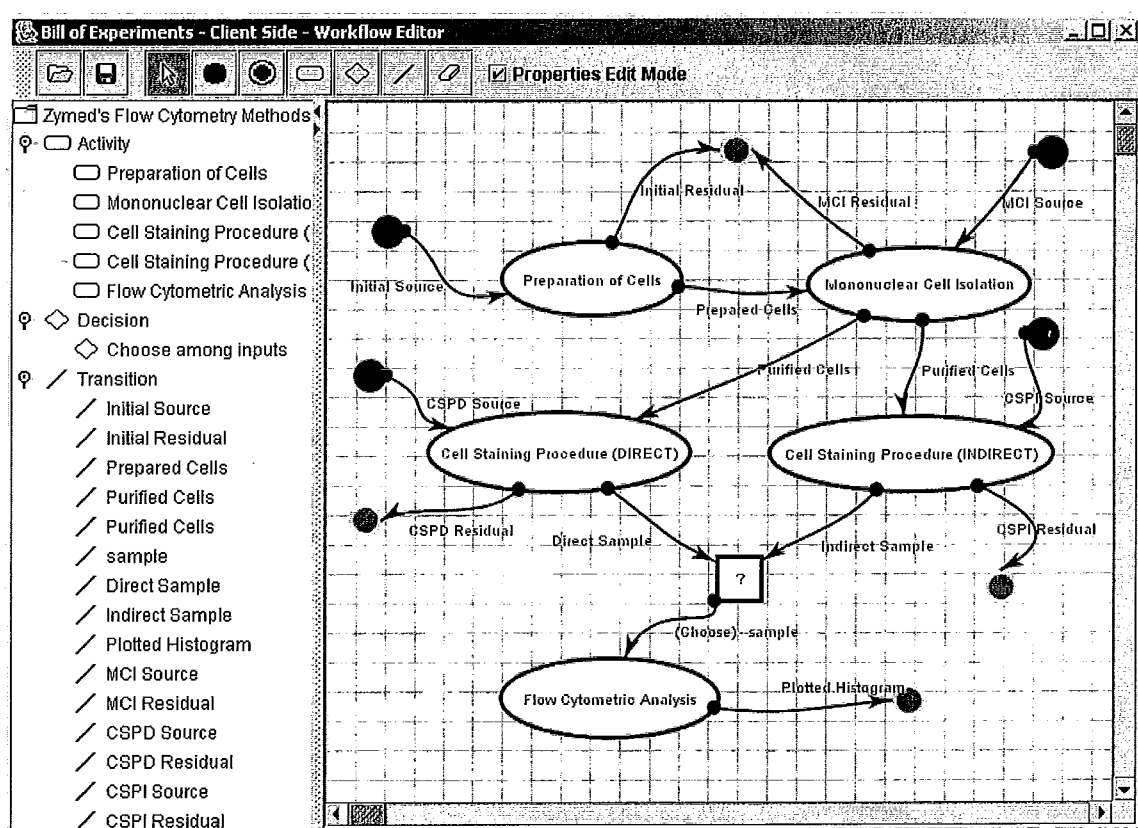


Figura 4.7 - Representação de um *Workflow* para Citometria de Fluxo

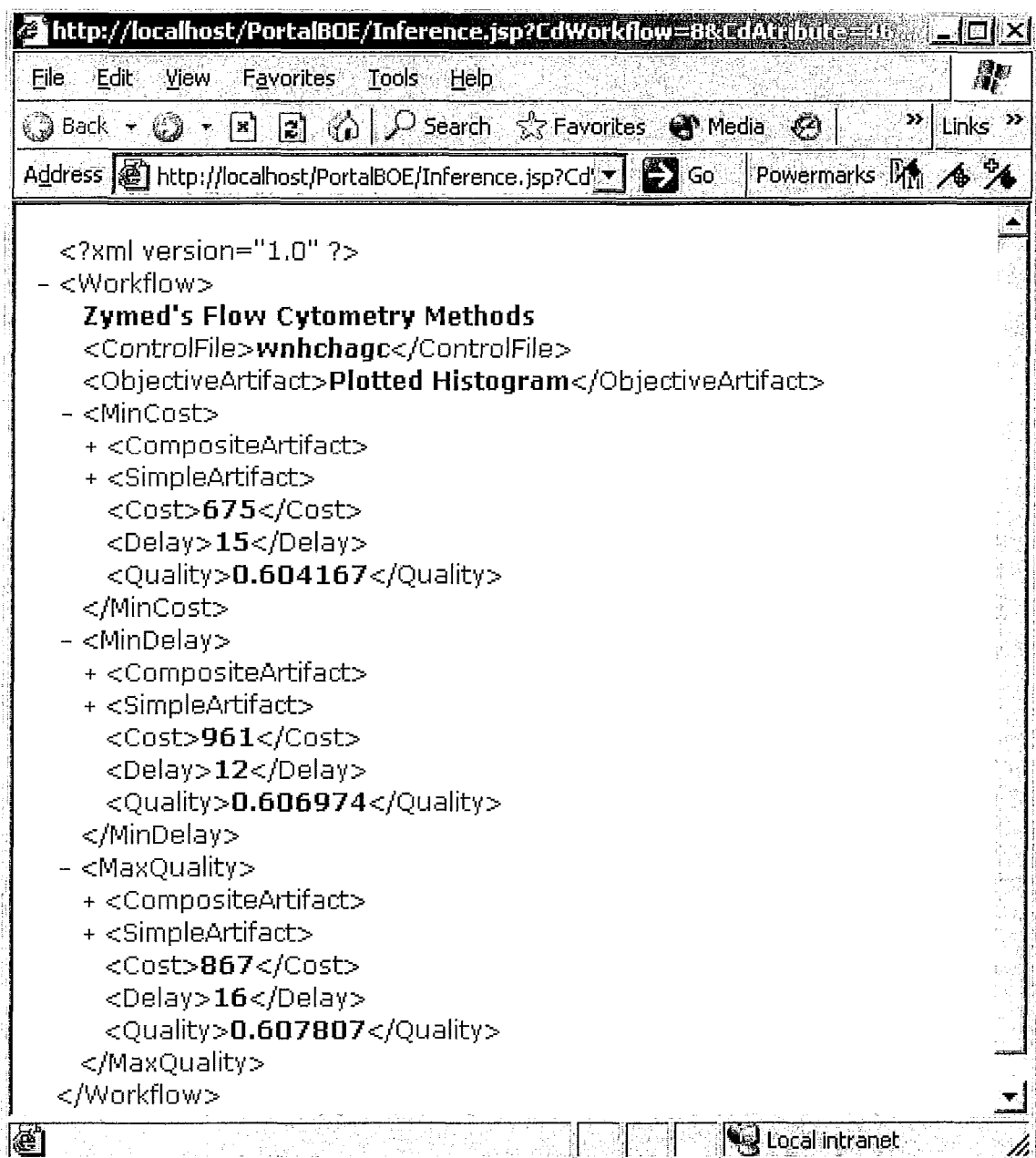


Figura 4.8 - Melhores Alternativas para o Histograma de Citometria de Fluxo

A figura acima demonstra os resultados obtidos para o artefato objetivo no *workflow*. Pode-se perceber a variação de custos, prazos e qualidades nos três planejamentos criados pela ferramenta. A composição de cada planejamento está disponível para o pesquisador, bastando para isso um clique nos elementos XML corretos.



Figura 4.9 - Melhores Alternativas para Células Preparadas

A figura acima traz os resultados obtidos para um artefato intermediário no *workflow*. Um cientista interessado em obter as “*Prepared Cells*”, mesmo que para uso em outro *workflow*, pode aproveitar este planejamento.

Capítulo 5 - Conclusões

Esta dissertação apresentou a ferramenta *Bill of Experiments* (BOE), um sistema colaborativo para explicitação, reuso e planejamento de *workflows* científicos. O conceito que embasou este trabalho foi criado a partir de uma analogia com o modelo clássico de *Bill of Materials* (BOM). O processo de montagem de um componente pode ser associado à aplicação de uma simulação ou experimento utilizando um ou mais conjuntos de dados científicos, gerando com isto uma nova informação relevante para a comunidade pesquisadora.

Na arquitetura proposta, a colaboração de pesquisadores e fornecedores é fundamental. Pesquisadores e cientistas são responsáveis pela explicitação de seus trabalhos de pesquisa, com detalhes referentes aos protocolos ou *workflows* científicos. Os fornecedores interessados em vender ou disponibilizar seus dados e experimentos contribuem com informações de preço e prazos de obtenção destes. Ao mesmo tempo em que temos um aumento na agilidade da obtenção de dados importantes para a realização de cada pesquisa, o custo envolvido em todo o processo tende a diminuir em consequência do planejamento adequado. A métrica de qualidade, obtida de maneira colaborativa, também pode ser considerada neste processo.

Para tornar capaz a **explicitação** de experimentos científicos em *workflows*, o BOE utiliza conceitos de Gestão do Conhecimento (NONAKA & TAKEUCHI, 1995), apoiando o processo de externalização, que consiste na conversão de conhecimento tácito em explícito. Este trabalho advoga que a representação de um experimento em um banco de dados é uma das formas de explicitação de conhecimento científico. Com a existência de conhecimento científico explícito no ambiente proposto, o BOE apóia o **reuso** de *workflows* e até mesmo a combinação deles através de um ambiente *web* que permite a colaboração entre pesquisadores e cientistas (*Sci2Sci*).

Fornecedores de artefatos científicos também são usuários da ferramenta BOE e podem descrever seus produtos, atribuindo custos e prazos de obtenção. Com o uso e adequação de uma métrica colaborativa de qualidade de dados científicos, o BOE, através de determinados critérios julgados pelos pesquisadores que utilizam determinado dado, estima quantitativamente a qualidade deste tipo de artefato científico.

Com as informações de fornecedores, custo, prazo e qualidade, é possível cogitar o melhor **planejamento** para execução de um experimento. Esta importante etapa também é apoiada pelo BOE, implementando colaboração entre fornecedores e cientistas (*B2Sci*) no momento em que aponta os menores custos e prazos de obtenção e a melhor qualidade para a obtenção dos artefatos científicos. Os resultados de planejamentos gerados pelo BOE seguem o padrão XML e podem ser utilizados por ferramentas que efetivamente controlem toda a realização de um experimento científico.

A proposta dessa dissertação é apresentada por meio da implementação de um protótipo voltado para a utilização via *web*. Aplicou-se o conceito do BOE numa área de aplicação real, a área de agrometeorologia, sendo o zoneamento agroclimático o experimento em questão. Duas outras áreas foram utilizadas para melhor exemplificar o funcionamento e uso da ferramenta.

Atualmente a ferramenta BOE faz parte do ambiente Agromet (PINTO, STRAUCH et al., 2002) que objetiva a aplicação da gestão do conhecimento na área de Agrometeorologia, em que cooperam a COPPE/UFRJ e a Embrapa Solos. No contexto do Agromet, o BOE atua como gerente de *workflows* científicos, explicitando, planejando e permitindo reuso e combinação dos experimentos da área.

5.1 Contribuições

Esta dissertação propõe a extensão do ciclo de vida de experimentos científicos proposto em (IOANNIDIS, LIVNY et al., 1996), contemplando a importante etapa de planejamento. Este consiste no planejamento dos artefatos científicos que se deseja utilizar em um experimento. Um dado base pode ser obtido a partir de diversos locais, e modelos matemáticos, criados por diferentes cientistas, podem ser executados em diferentes máquinas. A correta escolha destas fontes de artefatos pode levar a melhores resultados nas etapas seguintes, de coleta e análise de resultados experimentais.

Os sistemas de *workflows* científicos foram relacionados ao trabalho de banco de dados científicos de gerência de experimentos. Um sistema eficiente deve tratar as quatro etapas primordiais existentes no ciclo de vida proposto nesta dissertação. Não foi encontrado na literatura um trabalho que atuasse eficientemente em todas as etapas. O BOE atua claramente nas etapas de projeto e planejamento do experimento

(etapas 1 e 2) e deixa a cargo do ambiente SPeCS, através de trabalhos em andamento, a tarefa de implementar as outras.

No contexto da gestão do conhecimento científico, a ferramenta BOE atua no lado tecnológico da espiral de conhecimento, promovendo externalização, combinação e reuso, logo podendo ser vista como uma ferramenta tecnológica de apoio a gestão do conhecimento.

A efetiva utilização do BOE por um pesquisador o orientará a descobrir todos os dados e modelos que foram previstos em *workflows* científicos, com exposição de relações de custo-benefício apropriadas para cada coleção de dados de modo que este venha a ter um trabalho menos dispendioso e mais dinâmico.

5.2 Trabalhos Futuros

Algumas limitações são observadas no estágio atual de definição e implementação do BOE.

A criação de experimentos combinados a partir de dois ou mais *workflows* previamente existentes no sistema ainda não é automático e somente pode ser feito manualmente. Alguns trabalhos da literatura (KIEPUSZEWSKI, TER HOFSTEDE et al., 2000) já pesquisaram e confirmaram esta capacidade para *workflows* genéricos, desde que estes se encontrem estruturados.

Outras limitações dizem respeito à implementação. O histograma de cotações de qualidades para observação e aprimoramento dos dados científicos pelos fornecedores não está implementado até o momento. O módulo de busca ainda é muito limitado e necessita de diversas melhorias.

Seria interessante estender, sem perder a simplicidade, o modelo de diagrama de representação simples de *workflows* proposto neste trabalho (derivado do diagrama de atividades da UML), para que seja possível a tentativa de criação de um padrão de representação diagramática de *workflows* científicos.

Na avaliação de qualidade de dados, é importante um estudo sobre um sistema de pesos adequado ao contexto científico, semelhante ao proposto por PINHO (2001). Ainda tratando de qualidade, a criação de métricas de qualidade para outros

artefatos científicos que não sejam dados também é uma alternativa imediata de extensão deste trabalho.

Um estudo mais profundo sobre metadados necessários para a correta descrição e busca eficiente de artefatos científicos seria de grande valia para um ambiente como o BOE.

Referências Bibliográficas

- AILAMAKI, A., IOANNIDIS, Y. E., LIVNY, M., 1998, "Scientific Workflow Management by Database Management". In: *Proceedings of the 10th International Conference on Scientific and Statistical Database Management*, pp. 190-199, Capri, Italy, July.
- ALLEN, W., BOSCH, O., KILVINGTON, M., et al, 2001, "Benefits of collaborative learning for environmental management: Applying the Integrated Systems for Knowledge Management approach to support animal pest control", *Environmental Management*, v. 27, n. 2, pp. 215-223.
- ALMEIDA, V. F., 1998, "Pesquisa, tecnologias da informação e o futuro de Minas", *Revista Agrosoft*, v. 5.
- BARROS, A. J. P., LEHFELD, N. A. S., 1986, *Fundamentos de metodologia*, Sao Paulo, McGraw-Hill do Brasil.
- BARROSO, A. C. D. O., GOMES, E. B. P., 1999, "Tentando Entender a Gestão de Conhecimento", *Revista de Administração Pública*, v. 33, n. 2, pp. 147-170.
- BASTOS, R. M., RUIZ, D. D. A., 2002, "Extending UML Activity Diagram for Workflow Modeling in Production Systems". In: *35th Annual Hawaii International Conference on System Sciences (HICSS'02)-Volume 9*, pp. 291-292, Big Island, Hawaii, January.
- BERLINGUER, G., 2000, "Ciência, mercado e patentes do DNA Humano", *Revista Bio Ética*, v. 8, n. 1.
- BRAGA, E. V. D. C., 2002, "Gestão do Conhecimento na estrutura Organizacional da Ponte S.A.". In: <http://www.informal.com.br/artigos/a07121999001.htm>, Accessed in 08/10/2002.
- CARDOSO, L. F., SOUZA, J. M., MARQUES, C., 2002a, "A Collaborative Approach to the Reuse of Scientific Experiments in the Bill of Experiments Tool". In: *The Seventh International Conference on Computer Supported Cooperative Work in Design*, pp. 296-301, Rio de Janeiro, Brazil, September.
- CARDOSO, L. F., SOUZA, J. M., MARTORANO, L. G., 2002b, "Uma Abordagem Colaborativa para a Gestão do Conhecimento Científico na Ferramenta Bill of Experiments". In: *V International Symposium on Knowledge Management*, Curitiba, Brazil, August.
- CARVALHO, R. B., FERREIRA, M. A. T., 2000, "Análise de Softwares de Gestão do Conhecimento". In: *XXI SIMPÓSIO DE GESTÃO DA INOVAÇÃO TECNOLÓGICA*, Sao Paulo, SP, November.
- CARVALHO, R. B., FERREIRA, M. A. T., 2001, "Acelerando a Espiral do Conhecimento com a Tecnologia da Informação". In: *IV International Symposium on*

Knowledge Management/Document Management (ISKM/DM'2001), pp. 109-125, Curitiba, Brazil, August.

CASTRO, M., OLIVEIRA, J., STRAUCH, J. C. M., et al, 2002, "Decisio-Epistheme: An Integrated Environment to Geographic Decision-Making". In: *Fourth International Conference on Practical Aspects of Knowledge Management*, Vienna, Austria , December.

CAVALCANTI, M. C., MATTOSO, M., CAMPOS, M. L., et al, 2002, "An Architecture for Managing Distributed Scientific Resources". In: *14th International Conference on Scientific and Statistical Database Management (SSDBM'02)*, pp. 47-58, Edinburgh, Scotland, July.

CERVO, A. L.,BERVIAN, P. A., 1996, *Metodologia científica: para uso de estudantes universitários*. 3ª edição, São Paulo, Mc Graw - Hill.

CHEN, I.-M. A., KOSKY, A. S., MARKOWITZ, V. M., et al, 1997, "Constructing and Maintaining Scientific Database Views in the Framework of the Object-Protocol Model". In: *9th International Conference on Scientific and Statistical Database Management (SSDBM '97)*, pp. 237-248, Olympia, Washington, USA, August.

CHEN, I.-M. A.,MARKOWITZ, V. M., 1995a, "An Overview of the Object-Protocol Model (OPM) and OPM Data Management Tools", *Information Systems*, v. 20, n. 5.

CHEN, I.-M. A.,MARKOWITZ, V. M., 1995b, "Modeling Scientific Experiments with an Object Data Model". In: *Proceedings of the Eleventh International Conference on Data Engineering*, pp. 391-400, Taipei, Taiwan, Marchb.

CHEN, I.-M. A., MARKOWITZ, V. M., LETOVSKY, S., et al, 1996, "Version Management for Scientific Databases". In: *Advances in Database Technology - EDBT'96, 5th International Conference on Extending Database Technology*, pp. 289-303, Avignon, France, March.

CHEN, P. P., 1976, "The entity relationship model: toward a unified view of data", *ACM Transactions on Database Systems*, v. 1, pp. 9-36.

CORMICAN, K.,O'SULLIVAN, D., 2000, "A Collaborative Knowledge Management Tool for Product Innovation". In: *Proceedings of the Managing Innovative Manufacturing 2000 Conference*, Birmingham, UK, July.

DASGUPTA, P., CHAKRABARTI, P. P., DEY, A., et al, 2002, "Solving Constraint Optimization Problems from CLP-Style Specifications Using Heuristic Search Techniques", *IEEE Transactions on Knowledge and Data Engineering*, v. 14, n. 2, pp. 353-368.

DATE, C. J., 2000, *An introduction to database systems*. 7th ed, Reading, Mass, Addison-Wesley Pub. Co.

DAVENPORT, T. H.,PRUSAK, L., 1998, *Working Knowledge: How Organizations Manage What They Know*, Boston, Harvard Business School Press.

DTI/CSCW, 1997, "CSCW Glossary". In: <http://www.brainstorm.co.uk/CSCW/Glossary/Welcome.html>, Accessed in 21/12/2002.

EDWARDS, D. L., MAHLING, D. E., 1997, "Toward knowledge management systems in the legal domain". In: *Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge*, pp. 158-166, Phoenix, AZ USA.

ELMASRI, R., NAVATHE, S., 2000, *Fundamentals of database systems*. 3rd ed, Reading, Mass, Addison-Wesley.

FLUCKIGER, F., 1995, "Introduction to the Enabling Technologies", *Understanding Networked Multimedia: Applications and Technologies*, Prentice Hall.

FOSTER, I., KESSELMAN, C., TUECKE, S., 2001, "The anatomy of the grid: Enabling scalable virtual organizations", *The International Journal of Supercomputer Applications*, v. 15, n. 3, pp. 200-222.

FOSTER, I., VOCKLER, J., WILDE, M., et al, 2002, "Chimera: A Virtual Data System for Representing, Querying, and Automating Data Derivation". In: *14th International Conference on Scientific and Statistical Database Management (SSDBM'02)*, pp. 37-46, Edinburgh, Scotland, July.

FRENCH, J. C., JONES, A. K., PFALTZ, J. L., 1990, "Summary of the Final Report of the NSF Workshop on Scientific Database Management", *SIGMOD Record*, v. 19, n. 4, pp. 32-40.

GEORGAKOPOULOS, D., HORNICK, M. F., SHETH, A. P., 1995, "An Overview of Workflow Management: From Process Modeling to Workflow Automation Infrastructure", *Distributed and Parallel Databases*, v. 3, n. 2, pp. 119-153.

GRAYLING, A. C., 1996, "A Epistemologia". In: <http://www.filosofia.pro.br/textos/a-epistemologia.htm>, Accessed in 03/10/2002.

GRUDIN, J., 1994, "CSCW: History and Focus", *IEEE Computer*, v. 27, n. 5, pp. 19-27.

HARS, A., 1998, "Academic Knowledge Management Infrastructure: Information Systems Cybrarium". In: *Americas Conference on Information Systems 1998*, pp. 1119-1121.

HAYES-ROTH, F., 1985, "Rule-based systems", *Communications of the ACM*, v. 28, n. 9, pp. 921-932.

HENDRY, M., 2002, "Knowledge Management: One Technical Communicator's Odyssey", *News & Views Online*, v. 2001-2002.

HOUAISS, I. A., 2001, *Dicionário Houaiss da Língua Portuguesa*, Editora Objetiva.

IOANNIDIS, Y. E., LIVNY, M., GUPTA, S., et al, 1996, "Zoo: A Desktop Experiment Management Environment". In: *22nd International VLDB Conference*, pp. 274-285, Mumbai (Bombay), India, September.

JAFFAR, J., MAHER, M. J., 1994, "Constraint Logic Programming: A Survey", *Journal of Logic Programming*, v. 19/20, pp. 503-581.

KELLER, R., 1996, "ScienceDesk An integrating platform for collaborative science support". In: <http://ic.arc.nasa.gov/ic/projects/scidesk/>, Accessed in 10/12/2002.

KIEPUSZEWSKI, B., TER HOFSTEDÉ, A. H. M., BUSSLER, C. J., 2000, "On Structured Workflow Modelling". In: *12th International Conference CAiSE 2000*, pp. 431-445, Stockholm, Sweden, June.

KLEIN, G., 1999a, *Generating Graphical Editors for Graph-like Data Structures*, Tese de Diplomarbeit, Technische Universität München - Institut für Informatik, München, Germany.

KLEIN, G., 1999b, "Grace Project Page". In: <http://www.gerwin-klein.de/grace/>, Accessed in 17/02/2003b.

LABVELOCITY, 2003, "Protocols within Flow Cytometry". In: <http://researchlink.labvelocity.com/protocols/index.jhtml?parentId=28&id=356&path=0>, Accessed in 10/01/2003.

LAKATOS, E. M., MARCONI, M. D. A., 1991, *Metodologia Científica*. 3ª edição, São Paulo, Editora Atlas.

LETHBRIDGE, T. C., 1994, *Practical Techniques for Organizing and Measuring Knowledge*, Tese de PhD, University of Ottawa, Ottawa, Canada.

LIEBOWITZ, J., 2000, *Building organizational intelligence a knowledge management primer*, Boca Raton, Fla, CRC Press.

MEDEIROS, C. M. B., VOSSEN, G., WESKE, M., 1995, "WASA: A Workflow-Based Architecture to Support Scientific Database Applications". In: *6th DEXA Conference*, pp. 574-583, London, England, September.

MICHALEWICZ, Z., 1992, *Statistical and Scientific Databases*, New York, Ellis Horwood.

MORESI, E. A. D., 2001, "Gestão da Informação e do Conhecimento". In Tarapanoff, Kira, *Inteligência Organizacional e Competitiva*, chapter 5, Brasília, Editora UNB.

MURTA, L. G. P., 2002, *Charon: Uma Máquina de Processos Extensível Baseada em Agentes Inteligentes*, Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, Brasil.

NAKAYAMA, Y., SUMITA, K., SASAKI, K., et al, 2000, "Know-How Sharing Using a Knowledge Sharing System KIDS - A Knowledge Management Practice at a Research Laboratory". In: *PAKM 2000 Third International Conference on Practical Aspects of Knowledge Management*, pp. 21-1-21-7, Basel, Switzerland, October.

NELH, 2003, "KM principles and practices". In: http://www.nelh.nhs.uk/knowledge_management/km1/principles.asp, Accessed in 30/06/2003.

NETO, R. C., 2002, "O que é Conhecimento ? Sintetizando Epistemologia, Metodologia e Teoria de Sistemas em uma Nova Proposição", *Revista Eletrônica de Administração*, v. 8, n. 1.

NONAKA, I., TAKEUCHI, H., 1995, *The knowledge-creating company: how Japanese companies create the dynamics of innovation*, New York, Oxford University Press.

OLIVEIRA, J., SOUZA, J. M., STRAUCH, J. C. M., et al, 2002, "Epistheme: Scientific Knowledge Management Environment in the SPeCS Collaborative Framework". In: *The Seventh International Conference on CSCW in Design*, pp. 302-307, Rio de Janeiro, Brasil, September.

OMG, 2001, "Unified Modeling Language (UML), version 1.4". In: <http://www.omg.org/technology/documents/formal/uml.htm>, Accessed in 13/02/2003.

PACHECO, R. C. D. S., KERN, V. M., 2001, "Transparência e gestão do conhecimento por meio de um banco de teses e dissertações: a experiência do PPGEP/UFSC", *Revista da Ciência da Informação*, v. 30, n. 3, pp. 64-72.

PALMA, S. J. M., 2002, *SPeCS - Sistema de Suporte à Decisão Espacial Colaborativa*, Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, Brasil.

PALMA, S. J. M., SOUZA, J. M., STRAUCH, J. C. M., et al, 2001, "Coordination aspects in a spatial group decision support collaborative system". In: *Proceedings of the 16th ACM SAC2001 symposium on Applied computing*, pp. 182-186, Las Vegas, NV USA, March.

PFALTZ, J. L., 1990, "Differences Between Commercial and Scientific Data". In: *NSF Workshop on Scientific Database Management*, pp. 80-82, Charlottesville, VA, USA, March.

PINHO, S. F. C. D., 2001, *Avaliação da Qualidade de Dados pela Não Conformidade*, Tese de M.Sc., COPPE / UFRJ, Rio de Janeiro, Brasil.

PINTO, G. R. B., SOUZA, J. M., PALMA, S. J. M., et al, 2001, "X-Arc Spatial Data Integration in the SPeCS Collaborative Design Framework". In: *The Sixth International Conference on CSCW in Design*, Ontario, Canada, July.

PINTO, G. R. B., STRAUCH, J. C. M., CARDOSO, L. F., et al, 2002, "A Framework to Support Scientific Knowledge Management: a Case Study in Agro-meteorology". In: *The Seventh International Conference on CSCW in Design*, pp. 320-324, Rio de Janeiro, Brasil, September.

PIROLA, V. S., 2002, *A Proposição de uma Ferramenta de Apoio ao Mapeamento do Conhecimento em uma Organização*, Tese de M.Sc., COPPE UFRJ, Rio de Janeiro, Brasil.

PLESUMS, C., 2002, "Introduction to Workflow". In Fischer, Liana, *The Workflow Handbook 2002*, Future Strategies.

POLANYI, M., 1967, *The tacit dimension*, Garden City, N. Y, Doubleday.

- PRODAN, R., FAHRINGER, T., 2002, "ZENTURIO: An Experiment Management System for Cluster and Grid Computing". In: *4th IEEE International Conference on Cluster Computing (Cluster 2002)*, pp. 9-18, Chicago, USA, September.
- RENAUX, D. P. B., STRAUHS, F. R., ABREU, A. F., et al, 2001, "Gestão do Conhecimento de um Laboratório de Pesquisa: uma Abordagem Prática". In: *IV Simposio Internacional de Gestão do Conhecimento/Gestão de Documentos*, pp. 195-208, Curitiba, Brasil, August.
- REPELLI, C., 2001, "O mercado de serviços de meteorologia no Brasil", *Boletim da Sociedade Brasileira de Meteorologia (BSBMET)*, v. 25, n. 2, pp. 45-50.
- ROBBINS, R. J., 1990, "Types of Scientific Databases". In: *NSF Workshop on Scientific Database Management*, pp. 83-86, Charlottesville, VA, USA, March.
- RUGGLES, R. L., 1997, *Knowledge management tools*, Boston, Butterworth-Heinemann.
- RUMBAUGH, J., 1991, *Object-oriented modeling and design*, Englewood Cliffs, N.J, Prentice Hall.
- SALTZ, J., 2002, "Data Provenance". In: *Workshop on Data Derivation and Provenance*, Chigaco, USA, October.
- SANTOS, E. G. D., CERANTE, L. L., 2000, *Gestão do Conhecimento: Um Estudo para Facilitar sua Implantação nas Empresas*, Tese de BSc, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.
- SARAIVA, L. S., 2003, "Citometria de Fluxo". In: <http://www.ufv.br/dbg/LabC/citome3.html>, Accessed in 10/01/2003.
- SAS INSTITUTE INC, 1999, "SAS OnlineDoc®, Version 8", Cary, NC.
- SOUZA, J. M., CARDOSO, L. F., SULAIMAN, A., et al, 2002, *Bill of Experiments: A Tool for Scientific Knowledge Management*. ES-580/02 COPPE / UFRJ
- SOUZA, J. M., CARDOSO, L. F., SULAIMAN, A., et al, 2001, "Gestão de Conhecimento para Experimentos Científicos: Bill of Experiments". In: *IV International Symposium on Knowledge Management/Document Management*, Curitiba, August.
- SOUZA, J. M., STRAUCH, J. C. M., MARTORANO, L. G., et al, 2002, *Agromet: Gestão do Conhecimento em Agrometeorologia*. ES-581/02 COPPE / UFRJ
- STONEBRAKER, M., ROWE, L. A., HIROAMA, M., 1990, "The Implementation of Postgres", *IEEE Transactions on Knowledge and Data Engineering*, v. 2, n. 1, pp. 125-142.
- SULAIMAN, A., 1999, "Using Datalog Data Model and Data Mining to Solve Bill of Materials Personalization Problems". In: *International Workshop on Advance Issues of E-Commerce and Web-Based Information Systems*, Santa Clara, California, April.
- SUN, 2003a, "APPLETS". In: <http://java.sun.com/applets/>, Accessed in 02/2003a.

SUN, 2003b, "JavaServer Pages". In: <http://java.sun.com/products/jsp/>, Accessed in 02/2003b.

SVEIBY, K.-E., 2001, "What is Knowledge Management?". In: <http://www.sveiby.com/articles/KnowledgeManagement.html>, Accessed in 26/08/2002.

SWANSON, D. R., 1991, "Analysis of Unintended Connections Between Disjoint Science Literatures". In: *Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 280-289, Chicago, Illinois, USA, October.

SWI, 2003, "What is SWI-Prolog?". In: <http://www.swi-prolog.org/>, Accessed in 19/02/2003.

TANCO, R. A., 2000, "CLIMLAB 2000". *International Research Institute/University of La plata*, La Plata, Argentina.

TAYI, G. K.,BALLOU, D. P., 1998, "Examining data quality", *Communications of the ACM*, v. 41, n. 2, pp. 54-57.

TERRA, J. C. C., 1999, *Gestão do Conhecimento: Aspectos Conceituais e Estudo Exploratório Sobre as Práticas de Empresas Brasileiras*, Tese de Doutorado, Universidade de São Paulo.

TIJIBOY, A. V.,MAÇADA, D. L., 1997, "Cooperação e Colaboração: O Nosso Conceito". In: <http://www.niee.ufrgs.br/cursos/topicos-ie/ana/conceito.htm>, Accessed in 21/12/2002.

TURK, Z., BJORK, B.-C., MARTNES, B., et al, 2002, "The new role of professional organisations in the management of scientific knowledge". In: *CIB W78 conference 2002*, pp. 1-8, Denmark, June.

WAINER, J., WESKE, M., VOSSSEN, G., et al, 1996, "Scientific workflow systems". In: *NSF Workshop on Workflow and Process Automation*, Athens, Georgia, USA, May.

WENSLEY, A., 2000, "Tools for Knowledge Management". In: <http://www.icasit.org/km/resources/toolsforkm.htm>, Accessed in 06/12/2002.

WESKE, M., GOTTFRIED, V., MEDEIROS, C. M. B., et al, 1998, "Workflow Management in Geoprocessing Applications". In: *ACM-GIS '98, Proceedings of the 6th international symposium on Advances in Geographic Information Systems*, pp. 88-93, Washington, DC, USA, November.

WESKE, M., VOSSSEN, G., MEDEIROS, C. M. B., 1996, *Scientific Workflow Management: WASA Architecture and Applications*. 03/96-I Fachbericht Angewandte Mathematik und Informatik

WFMC, 1999, *Workflow Management Coalition Terminology & Glossary* WFMC-TC-1011

WILLSHIRE, M. J.,MEYEN, D., 1997, "A Process for Improving Data Quality", *Data Quality Journal*, v. 3, n. 1.

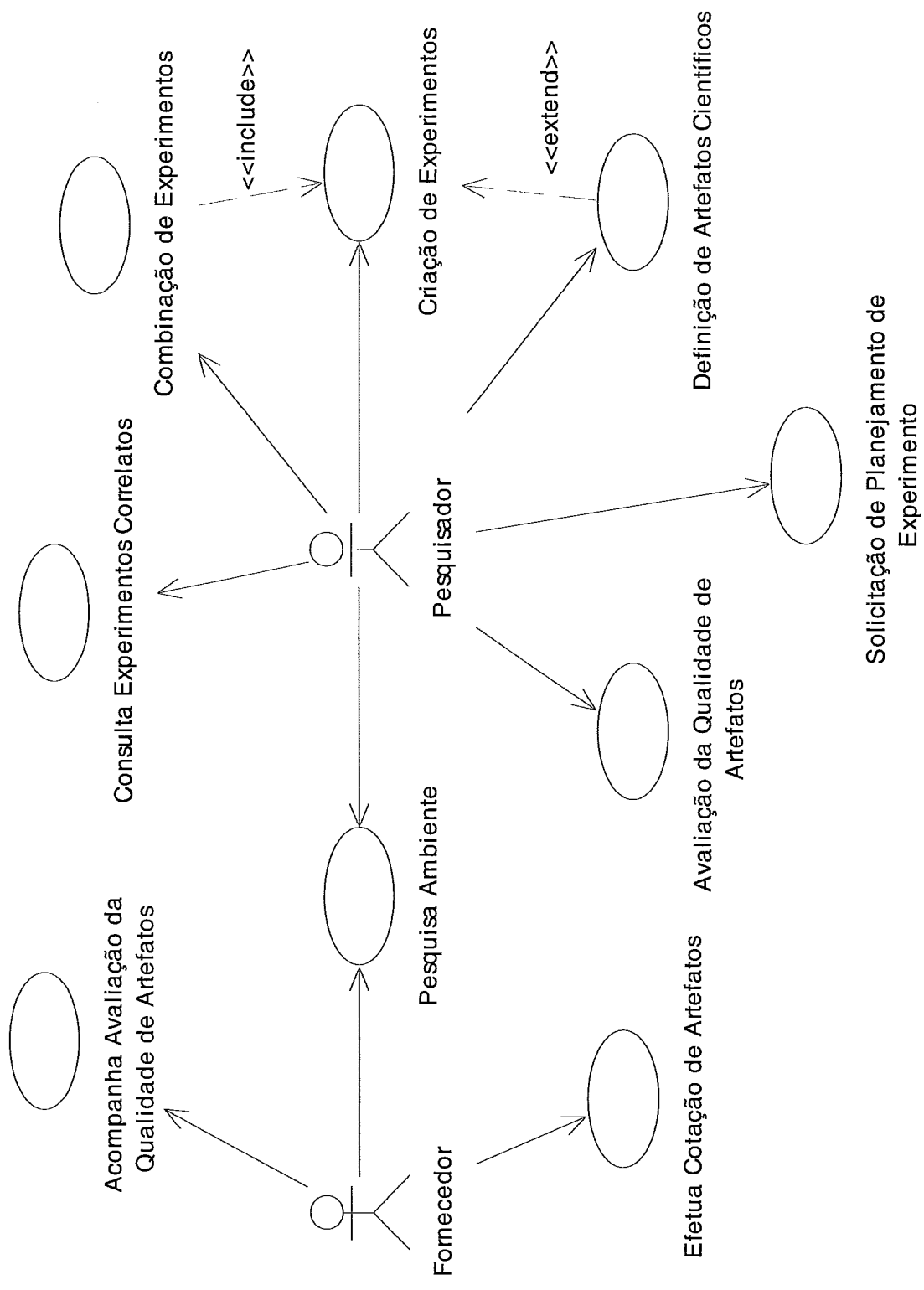
ZEMANKOVA, M., IOANNIDIS, Y. E., 1994, "Scientific Databases - State of the Art and Future Directions". In: *VLDB'94, Proceedings of 20th International Conference on Very Large Data Bases*, pp. 752-753, Santiago de Chile, Chile, September.

ZUBELLI, J., 1999, "An Introduction to Inverse Problems: Examples, Methods and Questions". In: *22º Colóquio Brasileiro de Matemática*, Rio de Janeiro, Brasil, July.

ZYMED LABORATORIES INC., 2000, "Zymed's Recommended Flow Cytometry Methods", November.

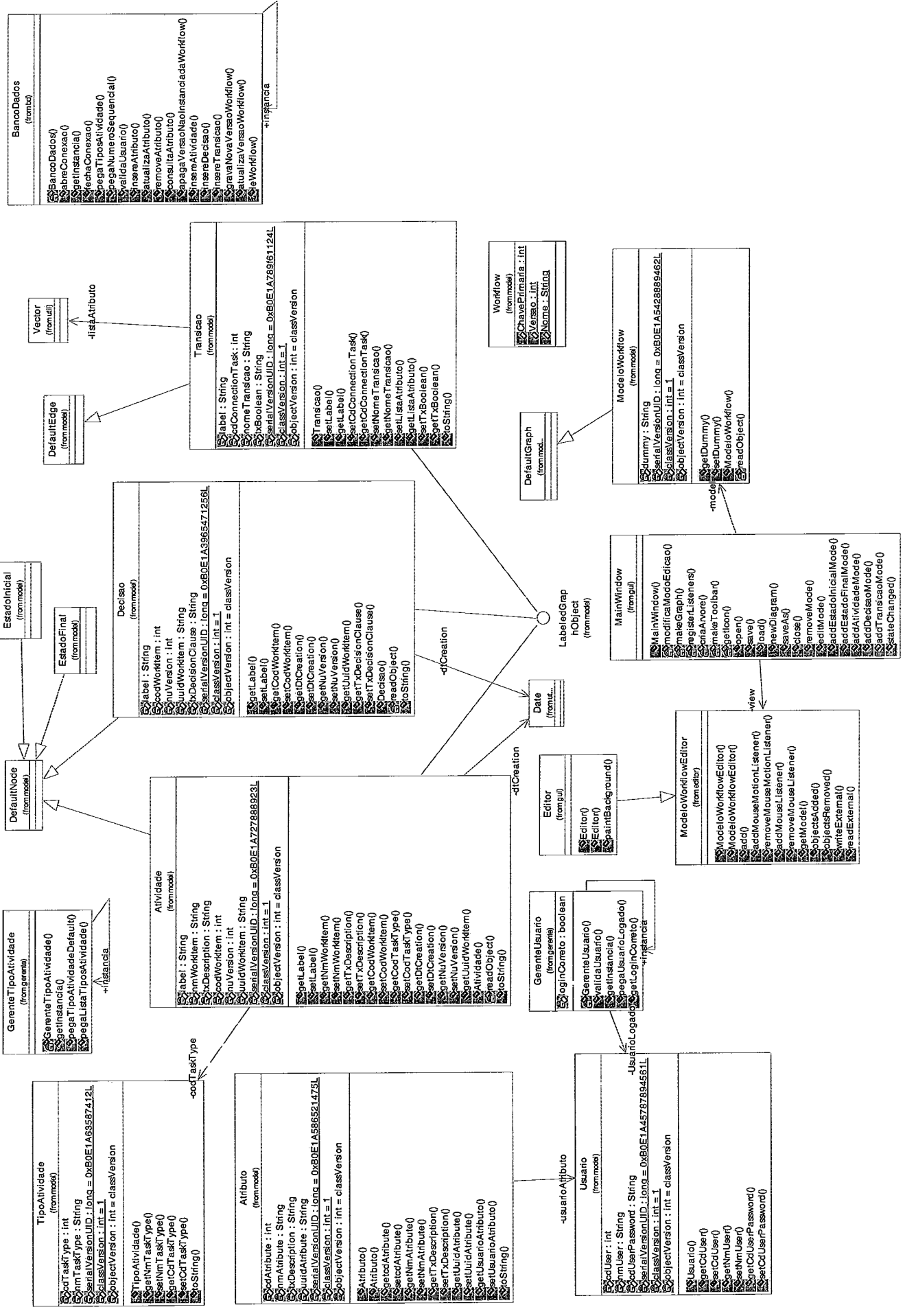
Apêndice A – Diagrama de Casos de Uso da Ferramenta BOE

Este apêndice descreve o modelo de casos de uso da ferramenta BOE.



Apêndice B – Modelo de Classes do Módulo de Projeto do *Workflow*

Este apêndice descreve sucintamente as classes criadas para a representação do *workflow* no módulo de projeto do *workflow*. Apenas as classes, com seus respectivos métodos e associações são listadas.



```

class TipoAtividade {
    -cdTaskType : int
    -nmTaskType : String
    -idTaskType : String
    -serialVersionUID : long = 0xB0E1A63587412L
    -classVersion : int = 1
    -objectVersion : int = classVersion
    TipoAtividade()
    TipoAtividade(int, String, String)
    getNmTaskType()
    getIdTaskType()
    getSerialVersionUID()
    getClassVersion()
    getObjectVersion()
}

```

```

class Atividade {
    -label : String
    -nmWorkItem : String
    -xDescription : String
    -codWorkItem : String
    -nuVersion : int
    -uidWorkItem : String
    -serialVersionUID : long = 0xB0E1A72788923L
    -classVersion : int = 1
    -objectVersion : int = classVersion
    Atividade()
    Atividade(String, String, String, String, int, String)
    getLabel()
    getNmWorkItem()
    getXDescription()
    getCodWorkItem()
    getNuVersion()
    getUidWorkItem()
    getSerialVersionUID()
    getClassVersion()
    getObjectVersion()
}

```

```

class Atividade {
    -label : String
    -nmWorkItem : String
    -xDescription : String
    -codWorkItem : String
    -nuVersion : int
    -uidWorkItem : String
    -serialVersionUID : long = 0xB0E1A72788923L
    -classVersion : int = 1
    -objectVersion : int = classVersion
    Atividade()
    Atividade(String, String, String, String, int, String)
    getLabel()
    getNmWorkItem()
    getXDescription()
    getCodWorkItem()
    getNuVersion()
    getUidWorkItem()
    getSerialVersionUID()
    getClassVersion()
    getObjectVersion()
}

```

```

class Decisao {
    -label : String
    -codWorkItem : int
    -nuVersion : int
    -xDecisionClause : String
    -serialVersionUID : long = 0xB0E1A3965471256L
    -classVersion : int = 1
    -objectVersion : int = classVersion
    Decisao()
    Decisao(String, int, int, String)
    getLabel()
    getCodWorkItem()
    getNuVersion()
    getXDecisionClause()
    getSerialVersionUID()
    getClassVersion()
    getObjectVersion()
}

```

```

class Transacao {
    -label : String
    -cdConnectionTask : int
    -nmConnectionTask : String
    -xBoolean : String
    -serialVersionUID : long = 0xB0E1A789161123L
    -classVersion : int = 1
    -objectVersion : int = classVersion
    Transacao()
    Transacao(String, int, String, String)
    getLabel()
    getCdConnectionTask()
    getNmConnectionTask()
    getXBoolean()
    getSerialVersionUID()
    getClassVersion()
    getObjectVersion()
}

```

```

class GerenteUsuario {
    -loginCorreto : boolean
    GerenteUsuario()
    loginCorreto()
    getLoginCorreto()
}

```

```

class Editor {
    Editor()
    Editor(String)
    paintBackground()
}

```

```

class ModeloWorkflowEditor {
    ModeloWorkflowEditor()
    ModeloWorkflowEditor(String)
    add()
    addMouseListener()
    addMouseMotionListener()
    addMouseListener()
    getModel()
    objectsAdded()
    objectsRemoved()
    writeExternal()
    readExternal()
}

```

```

class MainWindow {
    MainWindow()
    MainWindow(String)
    makeGraph()
    registerListeners()
    createAnore()
    createToolBar()
    getToolBar()
    open()
    save()
    load()
    newDiagram()
    saveAs()
    close()
    removeMode()
    removeMouseListener()
    removeMouseMotionListener()
    getModel()
    objectsAdded()
    objectsRemoved()
    writeExternal()
    readExternal()
    stateChanged()
}

```

```

class ModeloWorkflow {
    Dummy : String
    SerialVersionUID : long = 0xB0E1A542889462L
    ClassVersion : int = 1
    ObjectVersion : int = classVersion
    ModeloWorkflow()
    getDummy()
    getSerialVersionUID()
    getClassVersion()
    getObjectVersion()
}

```

```

class Workflow {
    ChavePrimaria : int
    Versao : int
    Nome : String
}

```

```

class EsaoInicial {
    EsaoInicial()
}

```

```

class EstadoFinal {
    EstadoFinal()
}

```

```

class DefaultEdge {
    DefaultEdge()
}

```

BancoDados (from:tbl)

abreConexao()
getInstancia()
fechaConexao()
pegaTipoAtividade()
pegaNumeroSequencial()
criaUsuario()
insereAtributo()
atualizaAtributo()
removeAtributo()
consultaAtributo()
pagaVersaoNoInstanciaWorkflow()
insereAtividade()
insereDecisao()
insereTransacao()
gravaNovaVersaoWorkflow()
atualizaVersaoWorkflow()
removeWorkflow()

Apêndice C – Programa Prolog Criado pelo Módulo de Inferência

Este apêndice traz a listagem do programa Prolog gerado pelo módulo de inferência para o planejamento do experimento descrito no estudo de caso do capítulo quatro. Esta implementação descreve a representação de um *workflow* através de regras proposta nesta dissertação.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Funcoes Auxiliares %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% minimum(Xs, X, Min) is true if Min is the smallest element in the
% list [X|Xs].
%EFICIENTE -> O(n)
minimum([Y|Ys], X, Z) :- Y <= X, !, minimum(Ys, Y, Z).
minimum([_]Ys, X, Z) :- minimum(Ys, X, Z).
minimum([], X, X).

maximum([Y|Ys], X, Z) :- Y >= X, !, maximum(Ys, Y, Z).
maximum([_]Ys, X, Z) :- maximum(Ys, X, Z).
maximum([], X, X).

%Implementação de mínimo para a lista de Custos
minimum_custo([Y|Ys], X, Z) :- Y=(_,A,_,_), X=(_,B,_,_), A <= B, !,
minimum_custo(Ys, Y, Z).
minimum_custo([_]Ys, X, Z) :- minimum_custo(Ys, X, Z).
minimum_custo([], X, X).

%Implementação de mínimo para a lista de Prazos
minimum_prazo([Y|Ys], X, Z) :- Y=(_,_,A,_), X=(_,_,B,_), A <= B, !,
minimum_prazo(Ys, Y, Z).
minimum_prazo([_]Ys, X, Z) :- minimum_prazo(Ys, X, Z).
minimum_prazo([], X, X).

%Implementação de máximo para a lista de Qualidades
maximum_qualidade([Y|Ys], X, Z) :- Y=(_,_,_,A), X=(_,_,_,B), A >= B, !,
maximum_qualidade(Ys, Y, Z).
maximum_qualidade([_]Ys, X, Z) :- maximum_qualidade(Ys, X, Z).
maximum_qualidade([], X, X).

%Chamada mais simplificada para a função mínimo de uma lista em O(N)
minimo_lista([H|T],X) :- minimum(T,H,X).

%Chamada mais simplificada para a função máximo de uma lista em O(N)
maximo_lista([H|T],X) :- maximum(T,H,X).

%Chamada mais simplificada para a função mínimo da lista de custos em
O(N)
minimo_lista_BOE_custo([H|T],X) :- minimum_custo(T,H,X).

%Chamada mais simplificada para a função mínimo da lista de prazos em
O(N)
minimo_lista_BOE_prazo([H|T],X) :- minimum_prazo(T,H,X).
```


%Chamada mais simplificada para a função máximo da lista de qualidades em O(N)

```
maximo_lista_BOE_qualidade([H|T],X) :- maximum_qualidade(T,H,X).
```

%Concatena Listas

```
conc([],L,L).
```

```
conc([X | L1], L2, [X | L3]) :-  
    conc(L1, L2, L3).
```

%Padrão para qualquer predicado

```
resolve(Pred, Lista_Instituicao, Custo, Prazo, Qualidade,  
Lista_Instituicao2, Custo2, Prazo2, Qualidade2, Lista_Instituicao3,  
Custo3, Prazo3, Qualidade3) :-
```

```
    K2=..[Pred,_X,_Y,_Z,_Q,_NVL], clause(K2, _L), %Verifica se existe o  
predicado
```

```
    K =..[Pred,X,Y,Z,Q,1],
```

```
    findall( (X,Y,Z,Q), K, L ), %Lista L instanciada com opções
```

```
    minimo_lista_BOE_custo(L, (Lista_Instituicao, Custo, Prazo,  
Qualidade)),
```

```
    minimo_lista_BOE_prazo(L, (Lista_Instituicao2, Custo2, Prazo2,  
Qualidade2)),
```

```
    maximo_lista_BOE_qualidade(L, (Lista_Instituicao3, Custo3, Prazo3,  
Qualidade3)).
```

%Escreve a lista composta de resultados como XML

%Caso seja composto

```
escreveXML([(X,Y,Z,W,Q)|L]) :-
```

```
    atom(X), Y=[_HEAD|_TAIL],
```

```
    write_ln('<CompositeArtifact>'),
```

```
    write('<Name>'), write(X), write_ln('</Name>'),
```

```
    write('<Composition>'), escreveXML(Y),
```

```
    write('<Cost>'), write(Z), write_ln('</Cost>'),
```

```
    write('<Delay>'), write(W), write_ln('</Delay>'),
```

```
    write('<Quality>'), write(Q), write_ln('</Quality>'),
```

```
    write_ln('</Composition>'),
```

```
    write_ln('</CompositeArtifact>'),
```

```
    escreveXML(L).
```

```
escreveXML([(X,Y,Z,W,Q)|L]) :-
```

```
    atom(X), atom(Y),
```

```
    write('<SimpleArtifact>'), write(X),
```

```
write_ln('<Institution>'),
```

```
    write('<Name>'), write(Y), write_ln('</Name>'),
```

```
    write('<Cost>'), write(Z), write_ln('</Cost>'),
```

```
    write('<Delay>'), write(W), write_ln('</Delay>'),
```

```
    write('<Quality>'), write(Q), write_ln('</Quality>'),
```

```
    write_ln('</Institution>'),
```

```
    write_ln('</SimpleArtifact>'),
```

```
    escreveXML(L).
```

```
escreveXML([]).
```

```
escreveXML(Y) :- write('<Institution>'), write(Y),
```

```
write_ln('</Institution>').
```

%Neste ponto já carregou-se o arquivo com funções auxiliares

```
:- ['funcoes_aux.pl'].
```

```
a_10( u_10, 90, 2, 0.0, _Dummy ).
```

```
a_10( u_9, 100, 2, 0.0, _Dummy ).
```

```

a_11( u_9, 110, 3, 0.6, _Dummy ).
a_2( u_10, 130, 1, 0.0, _Dummy ).
a_2( u_9, 100, 2, 0.0, _Dummy ).
a_61( u_9, 200, 4, 0.978947368421053, _Dummy ).
a_61( u_10, 90, 7, 0.6, _Dummy ).
a_7( u_9, 100, 2, 0.0, _Dummy ).
a_7( u_10, 95, 3, 0.768421052631579, _Dummy ).
a_8( u_10, 120, 1, 0.0, _Dummy ).
a_8( u_9, 80, 2, 0.0, _Dummy ).
a_9( u_9, 100, 2, 0.0, _Dummy ).
a_9( u_11, 0, 0, 0.968421052631579, _Dummy ).
p_2( u_11, 35, 1, 1, _Dummy ).
p_3( u_11, 60, 2, 1, _Dummy ).
p_3( u_1, 200, 5, 1, _Dummy ).
p_4( u_11, 20, 1, 1, _Dummy ).
p_5( u_8, 300, 4, 1, _Dummy ).
p_5( u_11, 34, 3, 1, _Dummy ).
p_6( u_11, 40, 5, 1, _Dummy ).
p_41( u_11, 30, 2, 1, _Dummy ).
p_42( u_11, 30, 2, 1, _Dummy ).
p_43( u_11, 35, 1, 1, _Dummy ).
p_44( u_11, 31, 2, 1, _Dummy ).
p_45( u_11, 35, 1, 1, _Dummy ).
p_46( u_11, 60, 3, 1, _Dummy ).
p_47( u_11, 70, 2, 1, _Dummy ).
p_48( u_11, 30, 2, 1, _Dummy ).
p_49( u_11, 40, 2, 1, _Dummy ).
p_50( u_11, 28, 2, 1, _Dummy ).
p_51( u_11, 33, 1, 1, _Dummy ).
:- assert( mincusto_poda(a_2, 10000000000) ).
:- assert( minprazo_poda(a_2, 10000000000) ).
:- assert( maxqualidade_poda(a_2, 0) ).
:- assert( mincusto_poda(a_7, 10000000000) ).
:- assert( minprazo_poda(a_7, 10000000000) ).
:- assert( maxqualidade_poda(a_7, 0) ).
:- assert( mincusto_poda(a_8, 10000000000) ).
:- assert( minprazo_poda(a_8, 10000000000) ).
:- assert( maxqualidade_poda(a_8, 0) ).
:- assert( mincusto_poda(a_9, 10000000000) ).
:- assert( minprazo_poda(a_9, 10000000000) ).
:- assert( maxqualidade_poda(a_9, 0) ).
:- assert( mincusto_poda(a_10, 10000000000) ).
:- assert( minprazo_poda(a_10, 10000000000) ).
:- assert( maxqualidade_poda(a_10, 0) ).
:- assert( mincusto_poda(a_11, 10000000000) ).
:- assert( minprazo_poda(a_11, 10000000000) ).
:- assert( maxqualidade_poda(a_11, 0) ).
:- assert( mincusto_poda(a_13, 10000000000) ).
:- assert( minprazo_poda(a_13, 10000000000) ).
:- assert( maxqualidade_poda(a_13, 0) ).
:- assert( mincusto_poda(a_18, 10000000000) ).
:- assert( minprazo_poda(a_18, 10000000000) ).
:- assert( maxqualidade_poda(a_18, 0) ).
:- assert( mincusto_poda(a_19, 10000000000) ).
:- assert( minprazo_poda(a_19, 10000000000) ).
:- assert( maxqualidade_poda(a_19, 0) ).
:- assert( mincusto_poda(a_20, 10000000000) ).
:- assert( minprazo_poda(a_20, 10000000000) ).
:- assert( maxqualidade_poda(a_20, 0) ).
:- assert( mincusto_poda(a_21, 10000000000) ).
:- assert( minprazo_poda(a_21, 10000000000) ).

```

```

:- assert( maxqualidade_poda(a_21, 0) ).
:- assert( mincusto_poda(a_22, 10000000000) ).
:- assert( minprazo_poda(a_22, 10000000000) ).
:- assert( maxqualidade_poda(a_22, 0) ).
:- assert( mincusto_poda(a_50, 10000000000) ).
:- assert( minprazo_poda(a_50, 10000000000) ).
:- assert( maxqualidade_poda(a_50, 0) ).
:- assert( mincusto_poda(a_51, 10000000000) ).
:- assert( minprazo_poda(a_51, 10000000000) ).
:- assert( maxqualidade_poda(a_51, 0) ).
:- assert( mincusto_poda(a_52, 10000000000) ).
:- assert( minprazo_poda(a_52, 10000000000) ).
:- assert( maxqualidade_poda(a_52, 0) ).
:- assert( mincusto_poda(a_53, 10000000000) ).
:- assert( minprazo_poda(a_53, 10000000000) ).
:- assert( maxqualidade_poda(a_53, 0) ).
:- assert( mincusto_poda(a_59, 10000000000) ).
:- assert( minprazo_poda(a_59, 10000000000) ).
:- assert( maxqualidade_poda(a_59, 0) ).
:- assert( mincusto_poda(a_60, 10000000000) ).
:- assert( minprazo_poda(a_60, 10000000000) ).
:- assert( maxqualidade_poda(a_60, 0) ).
:- assert( mincusto_poda(a_61, 10000000000) ).
:- assert( minprazo_poda(a_61, 10000000000) ).
:- assert( maxqualidade_poda(a_61, 0) ).
:- assert( mincusto_poda(a_62, 10000000000) ).
:- assert( minprazo_poda(a_62, 10000000000) ).
:- assert( maxqualidade_poda(a_62, 0) ).
:- assert( mincusto_poda(a_63, 10000000000) ).
:- assert( minprazo_poda(a_63, 10000000000) ).
:- assert( maxqualidade_poda(a_63, 0) ).
:- assert( mincusto_poda(a_64, 10000000000) ).
:- assert( minprazo_poda(a_64, 10000000000) ).
:- assert( maxqualidade_poda(a_64, 0) ).
:- assert( mincusto_poda(a_65, 10000000000) ).
:- assert( minprazo_poda(a_65, 10000000000) ).
:- assert( maxqualidade_poda(a_65, 0) ).
:- assert( mincusto_poda(a_66, 10000000000) ).
:- assert( minprazo_poda(a_66, 10000000000) ).
:- assert( maxqualidade_poda(a_66, 0) ).
:- assert( mincusto_poda(a_67, 10000000000) ).
:- assert( minprazo_poda(a_67, 10000000000) ).
:- assert( maxqualidade_poda(a_67, 0) ).
:- assert( mincusto_poda(a_68, 10000000000) ).
:- assert( minprazo_poda(a_68, 10000000000) ).
:- assert( maxqualidade_poda(a_68, 0) ).
:- assert( mincusto_poda(a_69, 10000000000) ).
:- assert( minprazo_poda(a_69, 10000000000) ).
:- assert( maxqualidade_poda(a_69, 0) ).
:- assert( nivelmaximo(12) ).
%Início das Regras
a_50(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :-
(
    nivelmaximo(NVL_MAX), Nivel < NVL_MAX,
    a_21(Inst_a_21, Custo_a_21, Prazo_a_21, Qualidade_a_21,
Nivel+1),
    p_2(Inst_p_2, Custo_p_2, Prazo_p_2, Qualidade_p_2, Nivel+1),
    Lista_Instituicao = [ (a_21, Inst_a_21, Custo_a_21,
Prazo_a_21, Qualidade_a_21), (p_2, Inst_p_2, Custo_p_2, Prazo_p_2,
Qualidade_p_2)],
    Custo is Custo_a_21 + Custo_p_2,

```

```

Qualidade is (Qualidade_a_21 + Qualidade_p_2) / 2,
maximo_lista([Prazo_a_21], Prazo_Dados),
Prazo is Prazo_Dados + Prazo_p_2),
(
(mincusto_poda(a_50, X), Custo < X,
retract(mincusto_poda(a_50,X)), assert(mincusto_poda(a_50,Custo)) );
(minprazo_poda(a_50, Y), Prazo < Y,
retract(minprazo_poda(a_50,Y)), assert(minprazo_poda(a_50,Prazo)) );
(maxqualidade_poda(a_50, Z), Qualidade > Z,
retract(maxqualidade_poda(a_50,Z)),
assert(maxqualidade_poda(a_50,Qualidade))
)
).
a_63(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
nivelmaximo(NVL_MAX), Nivel < NVL_MAX, a_51(Inst_a_51, Custo_a_51,
Prazo_a_51, Qualidade_a_51, Nivel+1), a_52(Inst_a_52, Custo_a_52,
Prazo_a_52, Qualidade_a_52, Nivel+1), a_53(Inst_a_53, Custo_a_53,
Prazo_a_53, Qualidade_a_53, Nivel+1), a_61(Inst_a_61, Custo_a_61,
Prazo_a_61, Qualidade_a_61, Nivel+1), p_3(Inst_p_3, Custo_p_3,
Prazo_p_3, Qualidade_p_3, Nivel+1), Lista_Instituicao = [ (a_51,
Inst_a_51, Custo_a_51, Prazo_a_51, Qualidade_a_51), (a_52, Inst_a_52,
Custo_a_52, Prazo_a_52, Qualidade_a_52), (a_53, Inst_a_53, Custo_a_53,
Prazo_a_53, Qualidade_a_53), (a_61, Inst_a_61, Custo_a_61, Prazo_a_61,
Qualidade_a_61), (p_3, Inst_p_3, Custo_p_3, Prazo_p_3,
Qualidade_p_3)], Custo is Custo_a_51 + Custo_a_52 + Custo_a_53 +
Custo_a_61 + Custo_p_3, Qualidade is (Qualidade_a_51 + Qualidade_a_52
+ Qualidade_a_53 + Qualidade_a_61 + Qualidade_p_3) / 5,
maximo_lista([Prazo_a_51, Prazo_a_52, Prazo_a_53, Prazo_a_61],
Prazo_Dados), Prazo is Prazo_Dados + Prazo_p_3), (
(mincusto_poda(a_63, X), Custo < X, retract(mincusto_poda(a_63,X)),
assert(mincusto_poda(a_63,Custo)) ); (minprazo_poda(a_63, Y), Prazo
< Y, retract(minprazo_poda(a_63,Y)), assert(minprazo_poda(a_63,Prazo))
); (maxqualidade_poda(a_63, Z), Qualidade > Z,
retract(maxqualidade_poda(a_63,Z)),
assert(maxqualidade_poda(a_63,Qualidade)) ) ) .
a_64(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
nivelmaximo(NVL_MAX), Nivel < NVL_MAX, a_51(Inst_a_51, Custo_a_51,
Prazo_a_51, Qualidade_a_51, Nivel+1), a_52(Inst_a_52, Custo_a_52,
Prazo_a_52, Qualidade_a_52, Nivel+1), a_53(Inst_a_53, Custo_a_53,
Prazo_a_53, Qualidade_a_53, Nivel+1), a_61(Inst_a_61, Custo_a_61,
Prazo_a_61, Qualidade_a_61, Nivel+1), p_3(Inst_p_3, Custo_p_3,
Prazo_p_3, Qualidade_p_3, Nivel+1), Lista_Instituicao = [ (a_51,
Inst_a_51, Custo_a_51, Prazo_a_51, Qualidade_a_51), (a_52, Inst_a_52,
Custo_a_52, Prazo_a_52, Qualidade_a_52), (a_53, Inst_a_53, Custo_a_53,
Prazo_a_53, Qualidade_a_53), (a_61, Inst_a_61, Custo_a_61, Prazo_a_61,
Qualidade_a_61), (p_3, Inst_p_3, Custo_p_3, Prazo_p_3,
Qualidade_p_3)], Custo is Custo_a_51 + Custo_a_52 + Custo_a_53 +
Custo_a_61 + Custo_p_3, Qualidade is (Qualidade_a_51 + Qualidade_a_52
+ Qualidade_a_53 + Qualidade_a_61 + Qualidade_p_3) / 5,
maximo_lista([Prazo_a_51, Prazo_a_52, Prazo_a_53, Prazo_a_61],
Prazo_Dados), Prazo is Prazo_Dados + Prazo_p_3), (
(mincusto_poda(a_64, X), Custo < X, retract(mincusto_poda(a_64,X)),
assert(mincusto_poda(a_64,Custo)) ); (minprazo_poda(a_64, Y), Prazo
< Y, retract(minprazo_poda(a_64,Y)), assert(minprazo_poda(a_64,Prazo))
); (maxqualidade_poda(a_64, Z), Qualidade > Z,
retract(maxqualidade_poda(a_64,Z)),
assert(maxqualidade_poda(a_64,Qualidade)) ) ) .
a_65(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
nivelmaximo(NVL_MAX), Nivel < NVL_MAX, a_51(Inst_a_51, Custo_a_51,
Prazo_a_51, Qualidade_a_51, Nivel+1), a_52(Inst_a_52, Custo_a_52,
Prazo_a_52, Qualidade_a_52, Nivel+1), a_53(Inst_a_53, Custo_a_53,

```

```

Prazo_a_53, Qualidade_a_53, Nivel+1), a_61(Inst_a_61, Custo_a_61,
Prazo_a_61, Qualidade_a_61, Nivel+1), p_3(Inst_p_3, Custo_p_3,
Prazo_p_3, Qualidade_p_3, Nivel+1), Lista_Instituicao = [ (a_51,
Inst_a_51, Custo_a_51, Prazo_a_51, Qualidade_a_51), (a_52, Inst_a_52,
Custo_a_52, Prazo_a_52, Qualidade_a_52), (a_53, Inst_a_53, Custo_a_53,
Prazo_a_53, Qualidade_a_53), (a_61, Inst_a_61, Custo_a_61, Prazo_a_61,
Qualidade_a_61), (p_3, Inst_p_3, Custo_p_3, Prazo_p_3,
Qualidade_p_3)], Custo is Custo_a_51 + Custo_a_52 + Custo_a_53 +
Custo_a_61 + Custo_p_3, Qualidade is (Qualidade_a_51 + Qualidade_a_52
+ Qualidade_a_53 + Qualidade_a_61 + Qualidade_p_3) / 5,
maximo_lista([Prazo_a_51, Prazo_a_52, Prazo_a_53, Prazo_a_61],
Prazo_Dados), Prazo is Prazo_Dados + Prazo_p_3), (
(mincusto_poda(a_65, X), Custo < X, retract(mincusto_poda(a_65,X)),
assert(mincusto_poda(a_65,Custo)) ); (minprazo_poda(a_65, Y), Prazo
< Y, retract(minprazo_poda(a_65,Y)), assert(minprazo_poda(a_65,Prazo))
); (maxqualidade_poda(a_65, Z), Qualidade > Z,
retract(maxqualidade_poda(a_65,Z)),
assert(maxqualidade_poda(a_65,Qualidade)) ) ) .
a_66(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
nivelmaximo(NVL_MAX), Nivel < NVL_MAX, a_51(Inst_a_51, Custo_a_51,
Prazo_a_51, Qualidade_a_51, Nivel+1), a_52(Inst_a_52, Custo_a_52,
Prazo_a_52, Qualidade_a_52, Nivel+1), a_53(Inst_a_53, Custo_a_53,
Prazo_a_53, Qualidade_a_53, Nivel+1), a_61(Inst_a_61, Custo_a_61,
Prazo_a_61, Qualidade_a_61, Nivel+1), p_3(Inst_p_3, Custo_p_3,
Prazo_p_3, Qualidade_p_3, Nivel+1), Lista_Instituicao = [ (a_51,
Inst_a_51, Custo_a_51, Prazo_a_51, Qualidade_a_51), (a_52, Inst_a_52,
Custo_a_52, Prazo_a_52, Qualidade_a_52), (a_53, Inst_a_53, Custo_a_53,
Prazo_a_53, Qualidade_a_53), (a_61, Inst_a_61, Custo_a_61, Prazo_a_61,
Qualidade_a_61), (p_3, Inst_p_3, Custo_p_3, Prazo_p_3,
Qualidade_p_3)], Custo is Custo_a_51 + Custo_a_52 + Custo_a_53 +
Custo_a_61 + Custo_p_3, Qualidade is (Qualidade_a_51 + Qualidade_a_52
+ Qualidade_a_53 + Qualidade_a_61 + Qualidade_p_3) / 5,
maximo_lista([Prazo_a_51, Prazo_a_52, Prazo_a_53, Prazo_a_61],
Prazo_Dados), Prazo is Prazo_Dados + Prazo_p_3), (
(mincusto_poda(a_66, X), Custo < X, retract(mincusto_poda(a_66,X)),
assert(mincusto_poda(a_66,Custo)) ); (minprazo_poda(a_66, Y), Prazo
< Y, retract(minprazo_poda(a_66,Y)), assert(minprazo_poda(a_66,Prazo))
); (maxqualidade_poda(a_66, Z), Qualidade > Z,
retract(maxqualidade_poda(a_66,Z)),
assert(maxqualidade_poda(a_66,Qualidade)) ) ) .
a_69(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
nivelmaximo(NVL_MAX), Nivel < NVL_MAX, a_51(Inst_a_51, Custo_a_51,
Prazo_a_51, Qualidade_a_51, Nivel+1), a_52(Inst_a_52, Custo_a_52,
Prazo_a_52, Qualidade_a_52, Nivel+1), a_53(Inst_a_53, Custo_a_53,
Prazo_a_53, Qualidade_a_53, Nivel+1), a_61(Inst_a_61, Custo_a_61,
Prazo_a_61, Qualidade_a_61, Nivel+1), p_3(Inst_p_3, Custo_p_3,
Prazo_p_3, Qualidade_p_3, Nivel+1), Lista_Instituicao = [ (a_51,
Inst_a_51, Custo_a_51, Prazo_a_51, Qualidade_a_51), (a_52, Inst_a_52,
Custo_a_52, Prazo_a_52, Qualidade_a_52), (a_53, Inst_a_53, Custo_a_53,
Prazo_a_53, Qualidade_a_53), (a_61, Inst_a_61, Custo_a_61, Prazo_a_61,
Qualidade_a_61), (p_3, Inst_p_3, Custo_p_3, Prazo_p_3,
Qualidade_p_3)], Custo is Custo_a_51 + Custo_a_52 + Custo_a_53 +
Custo_a_61 + Custo_p_3, Qualidade is (Qualidade_a_51 + Qualidade_a_52
+ Qualidade_a_53 + Qualidade_a_61 + Qualidade_p_3) / 5,
maximo_lista([Prazo_a_51, Prazo_a_52, Prazo_a_53, Prazo_a_61],
Prazo_Dados), Prazo is Prazo_Dados + Prazo_p_3), (
(mincusto_poda(a_69, X), Custo < X, retract(mincusto_poda(a_69,X)),
assert(mincusto_poda(a_69,Custo)) ); (minprazo_poda(a_69, Y), Prazo
< Y, retract(minprazo_poda(a_69,Y)), assert(minprazo_poda(a_69,Prazo))
); (maxqualidade_poda(a_69, Z), Qualidade > Z,

```

```

retract(maxqualidade_poda(a_69,Z)),
assert(maxqualidade_poda(a_69,Qualidade)) ) ) .
%Mais uma regra formatada para clareza
a_62(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :-
(
    nivelmaximo(NVL_MAX), Nivel < NVL_MAX,
    a_50(Inst_a_50, Custo_a_50, Prazo_a_50, Qualidade_a_50,
Nivel+1),
    a_51(Inst_a_51, Custo_a_51, Prazo_a_51, Qualidade_a_51,
Nivel+1),
    a_52(Inst_a_52, Custo_a_52, Prazo_a_52, Qualidade_a_52,
Nivel+1),
    a_53(Inst_a_53, Custo_a_53, Prazo_a_53, Qualidade_a_53,
Nivel+1),
    a_59(Inst_a_59, Custo_a_59, Prazo_a_59, Qualidade_a_59,
Nivel+1),
    a_60(Inst_a_60, Custo_a_60, Prazo_a_60, Qualidade_a_60,
Nivel+1),
    p_4(Inst_p_4, Custo_p_4, Prazo_p_4, Qualidade_p_4, Nivel+1),
    Lista_Instituicao = [ (a_50, Inst_a_50, Custo_a_50,
Prazo_a_50, Qualidade_a_50), (a_51, Inst_a_51, Custo_a_51, Prazo_a_51,
Qualidade_a_51), (a_52, Inst_a_52, Custo_a_52, Prazo_a_52,
Qualidade_a_52), (a_53, Inst_a_53, Custo_a_53, Prazo_a_53,
Qualidade_a_53), (a_59, Inst_a_59, Custo_a_59, Prazo_a_59,
Qualidade_a_59), (a_60, Inst_a_60, Custo_a_60, Prazo_a_60,
Qualidade_a_60), (p_4, Inst_p_4, Custo_p_4, Prazo_p_4,
Qualidade_p_4)],
    Custo is Custo_a_50 + Custo_a_51 + Custo_a_52 + Custo_a_53 +
Custo_a_59 + Custo_a_60 + Custo_p_4,
    Qualidade is (Qualidade_a_50 + Qualidade_a_51 + Qualidade_a_52
+ Qualidade_a_53 + Qualidade_a_59 + Qualidade_a_60 + Qualidade_p_4) /
7,
    maximo_lista([Prazo_a_50, Prazo_a_51, Prazo_a_52, Prazo_a_53,
Prazo_a_59, Prazo_a_60], Prazo_Dados), Prazo is Prazo_Dados +
Prazo_p_4),
    (
        (mincusto_poda(a_62, X), Custo < X,
retract(mincusto_poda(a_62,X)), assert(mincusto_poda(a_62,Custo)) );
        (minprazo_poda(a_62, Y), Prazo < Y,
retract(minprazo_poda(a_62,Y)), assert(minprazo_poda(a_62,Prazo)) );
        (maxqualidade_poda(a_62, Z), Qualidade > Z,
retract(maxqualidade_poda(a_62,Z)),
assert(maxqualidade_poda(a_62,Qualidade))
    )
).
a_67(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
nivelmaximo(NVL_MAX), Nivel < NVL_MAX, a_62(Inst_a_62, Custo_a_62,
Prazo_a_62, Qualidade_a_62, Nivel+1), a_69(Inst_a_69, Custo_a_69,
Prazo_a_69, Qualidade_a_69, Nivel+1), p_5(Inst_p_5, Custo_p_5,
Prazo_p_5, Qualidade_p_5, Nivel+1), Lista_Instituicao = [ (a_62,
Inst_a_62, Custo_a_62, Prazo_a_62, Qualidade_a_62), (a_69, Inst_a_69,
Custo_a_69, Prazo_a_69, Qualidade_a_69), (p_5, Inst_p_5, Custo_p_5,
Prazo_p_5, Qualidade_p_5)], Custo is Custo_a_62 + Custo_a_69 +
Custo_p_5, Qualidade is (Qualidade_a_62 + Qualidade_a_69 +
Qualidade_p_5) / 3, maximo_lista([Prazo_a_62, Prazo_a_69],
Prazo_Dados), Prazo is Prazo_Dados + Prazo_p_5), (
(mincusto_poda(a_67, X), Custo < X, retract(mincusto_poda(a_67,X)),
assert(mincusto_poda(a_67,Custo)) ); (minprazo_poda(a_67, Y), Prazo
< Y, retract(minprazo_poda(a_67,Y)), assert(minprazo_poda(a_67,Prazo))
); (maxqualidade_poda(a_67, Z), Qualidade > Z,

```

```

retract(maxqualidade_poda(a_67,Z)),
assert(maxqualidade_poda(a_67,Qualidade)) ) ) .
a_68(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
niveleximo(NVL_MAX), Nivel < NVL_MAX, a_67(Inst_a_67, Custo_a_67,
Prazo_a_67, Qualidade_a_67, Nivel+1), p_6(Inst_p_6, Custo_p_6,
Prazo_p_6, Qualidade_p_6, Nivel+1), Lista_Instituicao = [ (a_67,
Inst_a_67, Custo_a_67, Prazo_a_67, Qualidade_a_67), (p_6, Inst_p_6,
Custo_p_6, Prazo_p_6, Qualidade_p_6)], Custo is Custo_a_67 +
Custo_p_6, Qualidade is (Qualidade_a_67 + Qualidade_p_6) / 2,
maximo_lista([Prazo_a_67], Prazo_Dados), Prazo is Prazo_Dados +
Prazo_p_6), ( (mincusto_poda(a_68, X), Custo < X,
retract(mincusto_poda(a_68,X)), assert(mincusto_poda(a_68,Custo)) );
(minprazo_poda(a_68, Y), Prazo < Y, retract(minprazo_poda(a_68,Y)),
assert(minprazo_poda(a_68,Prazo)) ); (maxqualidade_poda(a_68, Z),
Qualidade > Z, retract(maxqualidade_poda(a_68,Z)),
assert(maxqualidade_poda(a_68,Qualidade)) ) ) ) .
a_21(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
niveleximo(NVL_MAX), Nivel < NVL_MAX, a_10(Inst_a_10, Custo_a_10,
Prazo_a_10, Qualidade_a_10, Nivel+1), p_41(Inst_p_41, Custo_p_41,
Prazo_p_41, Qualidade_p_41, Nivel+1), Lista_Instituicao = [ (a_10,
Inst_a_10, Custo_a_10, Prazo_a_10, Qualidade_a_10), (p_41, Inst_p_41,
Custo_p_41, Prazo_p_41, Qualidade_p_41)], Custo is Custo_a_10 +
Custo_p_41, Qualidade is (Qualidade_a_10 + Qualidade_p_41) / 2,
maximo_lista([Prazo_a_10], Prazo_Dados), Prazo is Prazo_Dados +
Prazo_p_41), ( (mincusto_poda(a_21, X), Custo < X,
retract(mincusto_poda(a_21,X)), assert(mincusto_poda(a_21,Custo)) );
(minprazo_poda(a_21, Y), Prazo < Y, retract(minprazo_poda(a_21,Y)),
assert(minprazo_poda(a_21,Prazo)) ); (maxqualidade_poda(a_21, Z),
Qualidade > Z, retract(maxqualidade_poda(a_21,Z)),
assert(maxqualidade_poda(a_21,Qualidade)) ) ) ) .
a_18(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
niveleximo(NVL_MAX), Nivel < NVL_MAX, a_7(Inst_a_7, Custo_a_7,
Prazo_a_7, Qualidade_a_7, Nivel+1), p_42(Inst_p_42, Custo_p_42,
Prazo_p_42, Qualidade_p_42, Nivel+1), Lista_Instituicao = [ (a_7,
Inst_a_7, Custo_a_7, Prazo_a_7, Qualidade_a_7), (p_42, Inst_p_42,
Custo_p_42, Prazo_p_42, Qualidade_p_42)], Custo is Custo_a_7 +
Custo_p_42, Qualidade is (Qualidade_a_7 + Qualidade_p_42) / 2,
maximo_lista([Prazo_a_7], Prazo_Dados), Prazo is Prazo_Dados +
Prazo_p_42), ( (mincusto_poda(a_18, X), Custo < X,
retract(mincusto_poda(a_18,X)), assert(mincusto_poda(a_18,Custo)) );
(minprazo_poda(a_18, Y), Prazo < Y, retract(minprazo_poda(a_18,Y)),
assert(minprazo_poda(a_18,Prazo)) ); (maxqualidade_poda(a_18, Z),
Qualidade > Z, retract(maxqualidade_poda(a_18,Z)),
assert(maxqualidade_poda(a_18,Qualidade)) ) ) ) .
a_51(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
niveleximo(NVL_MAX), Nivel < NVL_MAX, a_18(Inst_a_18, Custo_a_18,
Prazo_a_18, Qualidade_a_18, Nivel+1), p_43(Inst_p_43, Custo_p_43,
Prazo_p_43, Qualidade_p_43, Nivel+1), Lista_Instituicao = [ (a_18,
Inst_a_18, Custo_a_18, Prazo_a_18, Qualidade_a_18), (p_43, Inst_p_43,
Custo_p_43, Prazo_p_43, Qualidade_p_43)], Custo is Custo_a_18 +
Custo_p_43, Qualidade is (Qualidade_a_18 + Qualidade_p_43) / 2,
maximo_lista([Prazo_a_18], Prazo_Dados), Prazo is Prazo_Dados +
Prazo_p_43), ( (mincusto_poda(a_51, X), Custo < X,
retract(mincusto_poda(a_51,X)), assert(mincusto_poda(a_51,Custo)) );
(minprazo_poda(a_51, Y), Prazo < Y, retract(minprazo_poda(a_51,Y)),
assert(minprazo_poda(a_51,Prazo)) ); (maxqualidade_poda(a_51, Z),
Qualidade > Z, retract(maxqualidade_poda(a_51,Z)),
assert(maxqualidade_poda(a_51,Qualidade)) ) ) ) .
a_22(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
niveleximo(NVL_MAX), Nivel < NVL_MAX, a_11(Inst_a_11, Custo_a_11,
Prazo_a_11, Qualidade_a_11, Nivel+1), p_44(Inst_p_44, Custo_p_44,

```

```

Prazo_p_44, Qualidade_p_44, Nivel+1), Lista_Instituicao = [ (a_11,
Inst_a_11, Custo_a_11, Prazo_a_11, Qualidade_a_11), (p_44, Inst_p_44,
Custo_p_44, Prazo_p_44, Qualidade_p_44)], Custo is Custo_a_11 +
Custo_p_44, Qualidade is (Qualidade_a_11 + Qualidade_p_44) / 2,
maximo_lista([Prazo_a_11], Prazo_Dados), Prazo is Prazo_Dados +
Prazo_p_44), ( (mincusto_poda(a_22, X), Custo < X,
retract(mincusto_poda(a_22,X)), assert(mincusto_poda(a_22,Custo)) );
(minprazo_poda(a_22, Y), Prazo < Y, retract(minprazo_poda(a_22,Y)),
assert(minprazo_poda(a_22,Prazo)) ); (maxqualidade_poda(a_22, Z),
Qualidade > Z, retract(maxqualidade_poda(a_22,Z)),
assert(maxqualidade_poda(a_22,Qualidade)) ) ) .
a_52(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
nivelmaximo(NVL_MAX), Nivel < NVL_MAX, a_22(Inst_a_22, Custo_a_22,
Prazo_a_22, Qualidade_a_22, Nivel+1), p_45(Inst_p_45, Custo_p_45,
Prazo_p_45, Qualidade_p_45, Nivel+1), Lista_Instituicao = [ (a_22,
Inst_a_22, Custo_a_22, Prazo_a_22, Qualidade_a_22), (p_45, Inst_p_45,
Custo_p_45, Prazo_p_45, Qualidade_p_45)], Custo is Custo_a_22 +
Custo_p_45, Qualidade is (Qualidade_a_22 + Qualidade_p_45) / 2,
maximo_lista([Prazo_a_22], Prazo_Dados), Prazo is Prazo_Dados +
Prazo_p_45), ( (mincusto_poda(a_52, X), Custo < X,
retract(mincusto_poda(a_52,X)), assert(mincusto_poda(a_52,Custo)) );
(minprazo_poda(a_52, Y), Prazo < Y, retract(minprazo_poda(a_52,Y)),
assert(minprazo_poda(a_52,Prazo)) ); (maxqualidade_poda(a_52, Z),
Qualidade > Z, retract(maxqualidade_poda(a_52,Z)),
assert(maxqualidade_poda(a_52,Qualidade)) ) ) .
a_13(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
nivelmaximo(NVL_MAX), Nivel < NVL_MAX, a_2(Inst_a_2, Custo_a_2,
Prazo_a_2, Qualidade_a_2, Nivel+1), p_46(Inst_p_46, Custo_p_46,
Prazo_p_46, Qualidade_p_46, Nivel+1), Lista_Instituicao = [ (a_2,
Inst_a_2, Custo_a_2, Prazo_a_2, Qualidade_a_2), (p_46, Inst_p_46,
Custo_p_46, Prazo_p_46, Qualidade_p_46)], Custo is Custo_a_2 +
Custo_p_46, Qualidade is (Qualidade_a_2 + Qualidade_p_46) / 2,
maximo_lista([Prazo_a_2], Prazo_Dados), Prazo is Prazo_Dados +
Prazo_p_46), ( (mincusto_poda(a_13, X), Custo < X,
retract(mincusto_poda(a_13,X)), assert(mincusto_poda(a_13,Custo)) );
(minprazo_poda(a_13, Y), Prazo < Y, retract(minprazo_poda(a_13,Y)),
assert(minprazo_poda(a_13,Prazo)) ); (maxqualidade_poda(a_13, Z),
Qualidade > Z, retract(maxqualidade_poda(a_13,Z)),
assert(maxqualidade_poda(a_13,Qualidade)) ) ) .
a_53(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
nivelmaximo(NVL_MAX), Nivel < NVL_MAX, a_13(Inst_a_13, Custo_a_13,
Prazo_a_13, Qualidade_a_13, Nivel+1), p_47(Inst_p_47, Custo_p_47,
Prazo_p_47, Qualidade_p_47, Nivel+1), Lista_Instituicao = [ (a_13,
Inst_a_13, Custo_a_13, Prazo_a_13, Qualidade_a_13), (p_47, Inst_p_47,
Custo_p_47, Prazo_p_47, Qualidade_p_47)], Custo is Custo_a_13 +
Custo_p_47, Qualidade is (Qualidade_a_13 + Qualidade_p_47) / 2,
maximo_lista([Prazo_a_13], Prazo_Dados), Prazo is Prazo_Dados +
Prazo_p_47), ( (mincusto_poda(a_53, X), Custo < X,
retract(mincusto_poda(a_53,X)), assert(mincusto_poda(a_53,Custo)) );
(minprazo_poda(a_53, Y), Prazo < Y, retract(minprazo_poda(a_53,Y)),
assert(minprazo_poda(a_53,Prazo)) ); (maxqualidade_poda(a_53, Z),
Qualidade > Z, retract(maxqualidade_poda(a_53,Z)),
assert(maxqualidade_poda(a_53,Qualidade)) ) ) .
a_19(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
nivelmaximo(NVL_MAX), Nivel < NVL_MAX, a_8(Inst_a_8, Custo_a_8,
Prazo_a_8, Qualidade_a_8, Nivel+1), p_48(Inst_p_48, Custo_p_48,
Prazo_p_48, Qualidade_p_48, Nivel+1), Lista_Instituicao = [ (a_8,
Inst_a_8, Custo_a_8, Prazo_a_8, Qualidade_a_8), (p_48, Inst_p_48,
Custo_p_48, Prazo_p_48, Qualidade_p_48)], Custo is Custo_a_8 +
Custo_p_48, Qualidade is (Qualidade_a_8 + Qualidade_p_48) / 2,
maximo_lista([Prazo_a_8], Prazo_Dados), Prazo is Prazo_Dados +

```



```

Prazo_p_48), ( (mincusto_poda(a_19, X), Custo < X,
retract(mincusto_poda(a_19,X)), assert(mincusto_poda(a_19,Custo)) );
(minprazo_poda(a_19, Y), Prazo < Y, retract(minprazo_poda(a_19,Y)),
assert(minprazo_poda(a_19,Prazo)) ); (maxqualidade_poda(a_19, Z),
Qualidade > Z, retract(maxqualidade_poda(a_19,Z)),
assert(maxqualidade_poda(a_19,Qualidade)) ) ) .
a_59(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
nivelmaximo(NVL_MAX), Nivel < NVL_MAX, a_19(Inst_a_19, Custo_a_19,
Prazo_a_19, Qualidade_a_19, Nivel+1), p_49(Inst_p_49, Custo_p_49,
Prazo_p_49, Qualidade_p_49, Nivel+1), Lista_Instituicao = [ (a_19,
Inst_a_19, Custo_a_19, Prazo_a_19, Qualidade_a_19), (p_49, Inst_p_49,
Custo_p_49, Prazo_p_49, Qualidade_p_49)], Custo is Custo_a_19 +
Custo_p_49, Qualidade is (Qualidade_a_19 + Qualidade_p_49) / 2,
maximo_lista([Prazo_a_19], Prazo_Dados), Prazo is Prazo_Dados +
Prazo_p_49), ( (mincusto_poda(a_59, X), Custo < X,
retract(mincusto_poda(a_59,X)), assert(mincusto_poda(a_59,Custo)) );
(minprazo_poda(a_59, Y), Prazo < Y, retract(minprazo_poda(a_59,Y)),
assert(minprazo_poda(a_59,Prazo)) ); (maxqualidade_poda(a_59, Z),
Qualidade > Z, retract(maxqualidade_poda(a_59,Z)),
assert(maxqualidade_poda(a_59,Qualidade)) ) ) .
a_20(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
nivelmaximo(NVL_MAX), Nivel < NVL_MAX, a_9(Inst_a_9, Custo_a_9,
Prazo_a_9, Qualidade_a_9, Nivel+1), p_50(Inst_p_50, Custo_p_50,
Prazo_p_50, Qualidade_p_50, Nivel+1), Lista_Instituicao = [ (a_9,
Inst_a_9, Custo_a_9, Prazo_a_9, Qualidade_a_9), (p_50, Inst_p_50,
Custo_p_50, Prazo_p_50, Qualidade_p_50)], Custo is Custo_a_9 +
Custo_p_50, Qualidade is (Qualidade_a_9 + Qualidade_p_50) / 2,
maximo_lista([Prazo_a_9], Prazo_Dados), Prazo is Prazo_Dados +
Prazo_p_50), ( (mincusto_poda(a_20, X), Custo < X,
retract(mincusto_poda(a_20,X)), assert(mincusto_poda(a_20,Custo)) );
(minprazo_poda(a_20, Y), Prazo < Y, retract(minprazo_poda(a_20,Y)),
assert(minprazo_poda(a_20,Prazo)) ); (maxqualidade_poda(a_20, Z),
Qualidade > Z, retract(maxqualidade_poda(a_20,Z)),
assert(maxqualidade_poda(a_20,Qualidade)) ) ) .
a_60(Lista_Instituicao, Custo, Prazo, Qualidade, Nivel) :- (
nivelmaximo(NVL_MAX), Nivel < NVL_MAX, a_20(Inst_a_20, Custo_a_20,
Prazo_a_20, Qualidade_a_20, Nivel+1), p_51(Inst_p_51, Custo_p_51,
Prazo_p_51, Qualidade_p_51, Nivel+1), Lista_Instituicao = [ (a_20,
Inst_a_20, Custo_a_20, Prazo_a_20, Qualidade_a_20), (p_51, Inst_p_51,
Custo_p_51, Prazo_p_51, Qualidade_p_51)], Custo is Custo_a_20 +
Custo_p_51, Qualidade is (Qualidade_a_20 + Qualidade_p_51) / 2,
maximo_lista([Prazo_a_20], Prazo_Dados), Prazo is Prazo_Dados +
Prazo_p_51), ( (mincusto_poda(a_60, X), Custo < X,
retract(mincusto_poda(a_60,X)), assert(mincusto_poda(a_60,Custo)) );
(minprazo_poda(a_60, Y), Prazo < Y, retract(minprazo_poda(a_60,Y)),
assert(minprazo_poda(a_60,Prazo)) ); (maxqualidade_poda(a_60, Z),
Qualidade > Z, retract(maxqualidade_poda(a_60,Z)),
assert(maxqualidade_poda(a_60,Qualidade)) ) ) .
info(Pred) :- resolve(Pred, Lista_Instituicao1, Custos1, Prazos1,
Qualidade1, Lista_Instituicao2, Custos2, Prazos2, Qualidade2,
Lista_Instituicao3, Custos3, Prazos3, Qualidade3),
tell('C:\\TEMP\\xwdxectx.boe.txt'), write_ln('<MinCost>'),
escreveXML(Lista_Instituicao1), write('<Cost>'), write(Custos1),
write_ln('</Cost>'), write('<Delay>'), write(Prazos1),
write_ln('</Delay>'), write('<Quality>'), write(Qualidade1),
write_ln('</Quality>'), write_ln('</MinCost>'),
write_ln('<MinDelay>'), escreveXML(Lista_Instituicao2),
write('<Cost>'), write(Custos2), write_ln('</Cost>'), write('<Delay>'),
write(Prazos2), write_ln('</Delay>'), write('<Quality>'),
write(Qualidade2), write_ln('</Quality>'), write_ln('</MinDelay>'),
write_ln('<MaxQuality>'), escreveXML(Lista_Instituicao3),

```

```
write('<Cost>'), write(Custo3), write_ln('</Cost>'), write('<Delay>'),  
write(Prazo3), write_ln('</Delay>'), write('<Quality>'),  
write(Qualidade3), write_ln('</Quality>'), write_ln('</MaxQuality>'),  
told.  
:- info(a_68).  
:- halt.
```

Apêndice D – Esquema Lógico da Camada de Banco de Dados

Neste anexo, o esquema lógico de toda a camada de banco de dados é descrito sucintamente através de representação gráfica.

