

Universidade Federal de Santa Catarina
Programa de Pós-Graduação em
Engenharia de Produção

METODOLOGIA PARA DEFINIÇÃO DE UNIDADES DE INFORMAÇÃO PARA
PLATAFORMAS DE GOVERNO ELETRÔNICO: UMA APLICAÇÃO À
PLATAFORMA LATTES

Paulo Henrique de Souza Bermejo

Dissertação apresentada ao
Programa de Pós-Graduação em
Engenharia de Produção da
Universidade Federal de Santa Catarina
como requisito parcial para obtenção
do título de Mestre em
Engenharia de Produção

Florianópolis
2004

Paulo Henrique de Souza Bermejo

METODOLOGIA PARA DEFINIÇÃO DE UNIDADES DE INFORMAÇÃO PARA
PLATAFORMAS DE GOVERNO ELETRÔNICO: UMA APLICAÇÃO À
PLATAFORMA LATTES

Esta dissertação foi julgada e aprovada para a obtenção do grau de Mestre em
Engenharia de Produção no Programa de Pós-Graduação em Engenharia de
Produção da Universidade Federal de Santa Catarina

Florianópolis, 3 de julho de 2004.

Prof. Edson Pacheco Paladini, Dr.
Coordenador do Programa

BANCA EXAMINADORA

Prof. Roberto C. dos S. Pacheco, Dr.
Orientador
Universidade Federal de Santa Catarina

Prof. Aran Bey Tcholakian Morales, Dr.
Universidade Federal de Santa Catarina

Prof. Vinícius Medina Kern, Dr.
Universidade Federal de Santa Catarina

*Dedico aos meus queridos pais, Cacá e Margareth,
e aos meus irmãos amigos, Felipe e Vanessa.*

AGRADECIMENTOS

A Deus pela vida, saúde, paz, proteção, família e pelos bons amigos.

Aos meus queridos pais, Antonio Carlos Bermejo e Margareth A. de Souza Bermejo, e aos meus irmãos amigos, Felipe de Souza Bermejo e Vanessa Cristina Bermejo, que mesmo morando distante de mim sempre estão presentes em minha vida.

Ao professor Roberto Carlos dos Santos Pacheco pela amizade e orientação. Aos professores e amigos José Francisco Salm Jr., Vinicius Medina Kern e Aran Bey Tcholakian Morales pelo apoio e consideração.

Ao professor Ulysses T. Guariba Neto pelo incentivo.

Aos amigos que fiz durante este mestrado Alessandro Bovo, Amim Moisés, André Castoldi, Fernando Montenegro, Jaime Miranda Jr., Marcelo Cícero, Márcio Nápoli, Renato Balancieri, Sandra Martins, Tiago Fascin e Wagner Watanabe.

Aos professores que acompanharam os meus primeiros passos na academia e me incentivaram para que eu pleiteasse um curso de mestrado, Luiz Ricardo Begosso, Alex S. R. S. Poletto, Marisa Atsuko Nitto, Marcelo Manfio e Almir Rogério Camolesi.

Ao Grupo Stela por ter possibilitado a realização e aplicação da minha pesquisa em seus projetos, e a todos os outros amigos que conquistei nessa grande família.

RESUMO

Unidades de informação são as dimensões fundamentais de um domínio. Este trabalho apresenta uma metodologia para definição de unidades de informação para sistemas de governo eletrônico articulada com uma arquitetura de desenvolvimento de sistemas de e-gov. A metodologia é organizada nas etapas de identificação, especificação, padronização e gestão de unidades de informação. Para a sua descrição, são contemplados exemplos práticos dessas etapas por meio da unidade de informação de Grupos de Pesquisa da Plataforma Lattes. A metodologia é ainda usada neste trabalho para a definição de uma unidade de informação para o Programa de Fomento em Iniciação Científica, que será submetida à homologação da Comunidade CONSCIENTIAS - LMPL e adotada pelo CNPq e por outros interessados. A aplicabilidade permitiu a definição de unidades de informação padronizadas que forneçam uma especificação explícita e completa do seu domínio.

Palavras-chave: E-gov, Plataforma Lattes, UML e XML.

ABSTRACT

Information units are the fundamental dimensions of a domain. This work presents a methodology for the definition of information units for e-gov systems in the context of an architecture for e-gov platforms. The methodology is organized in four stages: identification, specification, standardization, and management of information units. It is illustrated through the definition of an existing information unit – Research Group – from the Lattes Platform. The methodology is applied in the definition of a new unit for the Institutional Program of Research Initiation Scholarships to be submitted for homologation by the Conscientias-LMPL Community and adopted by the Brazilian National Research Council and other interested parties. The methodology allows for the definition of standardized information units with explicit and complete specification of a domain.

SUMÁRIO

RESUMO	V
ABSTRACT	VI
LISTA DE FIGURAS	IX
LISTA DE TABELAS	XI
LISTA DE REDUÇÕES	XII
1 INTRODUÇÃO.....	1
1.1 APRESENTAÇÃO	1
1.2 DEFINIÇÃO DO PROBLEMA.....	2
1.3 ARQUITETURA REFERÊNCIA.....	3
1.4 OBJETIVO GERAL.....	5
1.5 OBJETIVOS ESPECÍFICOS.....	5
1.6 JUSTIFICATIVA.....	6
1.7 METODOLOGIA	6
1.8 ESTRUTURA DO TRABALHO	9
2 SISTEMAS DE E-GOV E UNIDADES DE INFORMAÇÃO.....	10
2.1 APRESENTAÇÃO	10
2.2 GOVERNO ELETRÔNICO E SISTEMAS E-GOV	11
2.3 ARQUITETURA DE SISTEMAS E-GOV	16
2.4 ENGENHARIA DE SOFTWARE EM ARQUITETURA DE E-GOV E UNIDADES DE INFORMAÇÃO	26
2.4.1 <i>LEVANTAMENTO DE REQUISITOS</i>	27
2.4.2 <i>ANÁLISE DE FONTES DE INFORMAÇÃO</i>	28
2.4.3 <i>ESPECIFICAÇÃO DO DOMÍNIO DE PROBLEMA</i>	29
2.4.4 <i>ONTOLOGIAS</i>	30
2.4.5 <i>PADRONIZAÇÃO DE INFORMAÇÕES</i>	33
2.4.6 <i>EXTENSIBLE MARKUP LANGUAGE (XML)</i>	34
3 METODOLOGIA PARA DEFINIÇÃO DE UNIDADES DE INFORMAÇÃO EM PLATAFORMAS DE E-GOV	37
3.1 APRESENTAÇÃO	37
3.2 IDENTIFICAÇÃO DE UNIDADE DE INFORMAÇÃO	38
3.3 ESPECIFICAÇÃO DE UNIDADE DE INFORMAÇÃO	40
3.3.1 <i>DEFINIÇÃO DO CONTEÚDO DE UMA UNIDADE DE INFORMAÇÃO</i>	41
3.3.2 <i>REPRESENTAÇÃO DA DEFINIÇÃO DO CONTEÚDO EM UM MODELO DESCRITIVO (NA FORMA DE UM DIAGRAMA DE CLASSES)</i>	49
3.4 PADRONIZAÇÃO DE UNIDADE DE INFORMAÇÃO.....	62
3.5 GESTÃO DE UNIDADE DE INFORMAÇÃO	75

3.6	FASE DE APLICAÇÃO À PLATAFORMA DE E-GOV.....	78
4	APLICAÇÃO: GESTÃO DE UNIDADES DE INFORMAÇÃO EM C&T – PLATAFORMA LATTES	79
4.1	APRESENTAÇÃO	79
4.2	PLATAFORMA EUROPÉIA – EUROCRIS	80
4.3	PLATAFORMA LATTES.....	81
4.4	UNIDADES DE INFORMAÇÃO DA PLATAFORMA LATTES	85
4.4.1	<i>IDENTIFICAÇÃO: PROCESSO DE DEFINIÇÃO DA UNIDADE DE INFORMAÇÃO PARA O PROGRAMA DE FOMENTO EM INICIAÇÃO CIENTÍFICA.....</i>	85
4.4.2	<i>ETAPA DE ESPECIFICAÇÃO DA UNIDADE DE INFORMAÇÃO PARA O PROGRAMA DE FOMENTO EM INICIAÇÃO CIENTÍFICA</i>	86
4.4.3	<i>ETAPA DE PADRONIZAÇÃO DA UNIDADE DE INFORMAÇÃO PARA O PROGRAMA DE FOMENTO EM INICIAÇÃO CIENTÍFICA.....</i>	88
4.4.4	<i>GESTÃO: COMO SÃO GERIDAS AS UNIDADES DE INFORMAÇÃO PADRONIZADAS EM TERMOS DA PLATAFORMA LATTES.....</i>	89
4.5	RECURSOS DA PLATAFORMA LATTES VIABILIZADOS PELA GESTÃO DE SUAS UNIDADES DE INFORMAÇÃO.....	91
5	CONCLUSÕES E FUTUROS TRABALHOS.....	93
5.1	CONCLUSÕES	93
5.2	TRABALHOS FUTUROS	94
	REFERÊNCIAS BIBLIOGRÁFICAS.....	95
	ANEXO I – CODIFICAÇÃO EM XML SCHEMA PARA A UNIDADE PARA O PROGRAMA DE FOMENTO EM IC PROVENIENTE DA APLICAÇÃO DA ETAPA DE PADRONIZAÇÃO	103

LISTA DE FIGURAS

Figura 1 – Arquitetura conceitual proposta para projetos de e-gov (PACHECO, 2003) .	5
Figura 2 – Metodologia do trabalho	8
Figura 3 – Arquitetura proposta pela FEA (2002) segmentada em quatro níveis	18
Figura 4 – Arquitetura proposta pela FEA (2002) associada com os modelos de referência e agências de E-gov norte-americanas.....	19
Figura 5 – Modelo de Referência Técnica de governo eletrônico – TRM (FEA, 2002, p. 26)	20
Figura 6 - Arquitetura para projetos de e-gov proposta por Pacheco (2003) com destaque para os seus quatro componentes.....	23
Figura 7 – Enquadramento da metodologia deste trabalho na arquitetura de sistemas de e-gov proposta por Pacheco (2003).	37
Figura 8 - Metodologia proposta para a definição de unidades de informação em plataformas de sistemas de governo eletrônico	38
Figura 9 – Modelo do cartão para descrição dos candidatos a objetos na fase de validação.....	43
Figura 10 – Cartão de identificação para o objeto “Identificação do grupo”	44
Figura 11 – Cartão de identificação para o objeto “ENDEREÇO INSTITUCIONAL DO GRUPO”	44
Figura 12 – Cartão de identificação para o objeto “LÍDER”	44
Figura 13 – Cartão de identificação para o objeto “REPERCUSSÃO”	45
Figura 14 – Cartão de identificação para o objeto “PESQUISADOR”	45
Figura 15 – Cartão de identificação para o objeto “ESTUDANTE”.....	45
Figura 16 – Cartão de identificação para o objeto “TÉCNICO”.....	45
Figura 17 – Cartão de identificação para o objeto “LINHA DE PESQUISA”.....	46
Figura 18 – Cartão de identificação para o objeto “EMPRESA”.....	46
Figura 19 – Cartão de identificação para o objeto “PALAVRA-CHAVE”.....	47
Figura 20 – Cartão de identificação para o objeto “SETOR DE APLICAÇÃO”	47
Figura 21 – Cartão de identificação para o objeto “SETOR DE ATIVIDADE ECONÔMICA”	47
Figura 22 – Cartão de identificação para o objeto “TIPO DE RELAÇÃO”.....	47
Figura 23 – Cartão de identificação para o objeto “TIPO DE REMUNERAÇÃO”	48
Figura 24 – Exemplo de representação do conceito de herança em um diagrama de classes da UML.....	50
Figura 24 – Exemplo da organização de tipos e subtipos de dados em um diagrama de classes da UML.....	51
Figura 25 – Diagrama de classe que contém os objetos da unidade de informação de “Grupo de Pesquisa” organizados em uma hierarquia de classes.....	53
Figura 26 – Exemplo de relação do tipo composição e agregação.....	55
Figura 27 – Relacionamento de associação com cardinalidade de 1 para muitos, entre a classe “ClasseA” e a classe “ClasseB”	55
Figura 28 – Definição e apresentação dos relacionamentos de composição da classe principal da unidade de informação de Grupo de Pesquisa.....	56
Figura 29 – Especificação das classes relacionadas com a classe “IdentificacaoDoGrupo”.	57

Figura 30 – Apresentação do relacionamento de associação utilizado na especificação da unidade de informação de Grupo de Pesquisa	58
Figura 31 – Apresentação das classes para a unidade de informação de Grupo de Pesquisa com detalhamento de propriedades e tipos de dados.....	60
Figura 32 – Parte do diagrama de classes UML, criado para a unidade de informação de Grupo de Pesquisa, que contém classes com relacionamento do tipo associação..	71
Figura 33 – Relacionamento de agregação entre as unidades de informação Grupo de Pesquisa e Currículo.....	74
Figura 34 – Ciclo da Comunidade CONSCIENTIAS para desenvolvimento de ontologias (CONSCIENTIAS, 2004)	76
Figura 35 – Ferramenta da Comunidade CONSCIENTIAS/LMPL para acompanhamento das versões das unidades de informação.....	77
Figura 37 – Mecanismo de conversão para integração das informações entre os sistemas que mantenham compatibilidade com o padrão XML especificado em diferentes idiomas.	91
Figura 38 – Estrutura do XML Schema criado para a unidade de informação para o Programa de Fomento em Iniciação Científica	103

LISTA DE TABELAS

Tabela 1 – Projetos premiados no III Prêmio Excelência em Governo Eletrônico (PRÊMIO E-GOV, 2004).....	15
Tabela 2 – Lista inicial dos candidatos a objetos para a unidade de informação de grupo de pesquisa.....	42

LISTA DE REDUÇÕES

- ABEP - Associação Brasileira de Empresas Estaduais de Processamento de Dados
- API - Interface de programa de aplicação (*Application Program Interface*)
- ARM - Modelo de Referência de Capacidade de Aplicação (*Application Capability Reference Model*)
- BRM - Modelo de Referência Administrativo ou de Negócios (*Business Reference Model*)
- BTD - Banco de Teses e Dissertações
- C&T - Ciência e Tecnologia
- CERIF - Formato comum de informações de pesquisa usado pela Europa (*Common European Research Information Format*)
- CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico
- CONSCIENTIAS - Comunidade para Ontologias em Ciência, Tecnologia e Informações de Aperfeiçoamento de Nível Superior
- CRC - Classe de Responsabilidade e Colaboração (*Class-Responsibility-Collaboration*)
- CRIS - Sistemas de Informação de Pesquisas Correntes (*Current Research Information Systems*)
- DRIS - Diretório de Pesquisa de Sistemas de Informação (*Directory of Research Information Systems*)
- DRM - Modelo de Referência de Dados e Informação (*Data and Information Reference Model*)
- EBAPE/FGV - Escola Brasileira de Administração Pública e de Empresas da Fundação Getúlio Vargas
- EEI - Interface de Ambiente Externo (*External Environment Interface*)

e-gov	- governo eletrônico
ERA	- Área de Pesquisa Européia (<i>European Research Area</i>)
ERGO	- Portal Europeu On-line de Pesquisa (<i>European Research Gateways On-line</i>)
FIOCRUZ	- Fundação Oswaldo Cruz
G2B	- Governo para Empresas (<i>Government to Businesses</i>)
G2C	- Governo para Cidadãos (<i>Government to Citizens</i>)
G2G	- Governo para Governo (<i>Government to Government</i>)
HTML	- Linguagem de Hipertexto por Anotações (<i>HyperText Markup Language</i>)
IBICT	- Instituto Brasileiro de Informação em Ciência e Tecnologia
IC	- Iniciação Científica
IST	- Tecnologias da Sociedade da Informação (<i>Information Society Technologies</i>)
LMPL	- Linguagem de Marcação da Plataforma Lattes
MCT	- Ministério da Ciência e Tecnologia
MPOG	- Ministério do Planejamento, Orçamento e Gestão
PIBIC	- Programa Institucional de Bolsas de Iniciação Científica
PPGEP	- Programa de Pós-Graduação em Engenharia de Produção
PRM	- Modelo de Referência de Performance (<i>Performance Reference Model</i>)
SABESP	- Companhia de Saneamento Básico do Estado de São Paulo
SBC	- Sociedade Brasileira de Computação
SGML	- Linguagem de Marcação Padrão Generalizada (<i>Standard Generalized Markup Language</i>)
TI	- Tecnologia da Informação

TRM	-	Modelo de Referência Técnica (<i>Technical Reference Model</i>)
UFBA	-	Universidade Federal da Bahia
UFMG	-	Universidade Federal de Minas Gerais
UFRGS	-	Universidade Federal do Rio Grande do Sul
UFRJ	-	Universidade Federal do Rio de Janeiro
UFRN	-	Universidade Federal do Rio Grande do Norte
UFSC	-	Universidade Federal de Santa Catarina
Unicamp	-	Universidade Estadual de Campinas
URI	-	Indicador Uniforme de Recurso (<i>Uniform Resource Identifier</i>)
USP	-	Universidade de São Paulo
W3C	-	<i>World Wide Web Consortium</i>
WWW	-	<i>World Wide Web</i>
XMI	-	Modelo XML para intercâmbio de meta-dado (<i>XML Metadata Interchange</i>)

1 INTRODUÇÃO

Tecnologia é o principal meio de criar vantagem competitiva suficiente para tirar o país do subdesenvolvimento.

José Ellis Ripper Filho

1.1 APRESENTAÇÃO

O surgimento da *Web* e, conseqüentemente, a sua integração com ferramentas de tecnologia da informação provocaram o surgimento de serviços oferecidos pela Internet.

O uso das novas tecnologias da informação pelos governos na prestação de serviços e informações para cidadãos, fornecedores e servidores constitui o que se convencionou chamar de governo eletrônico (SF/BNDES, 2003). Holmes (2001) define o governo eletrônico, ou e-gov, como o uso da tecnologia da informação, em particular da Internet, para disponibilizar serviços públicos de forma orientada ao cliente, com muito mais comodidade e com um custo compensador. Segundo Pacheco e Kern (2003), o e-gov teve o seu surgimento impulsionado inicialmente pelo setor privado através do comércio eletrônico (*e-commerce*). De acordo com Hoeschl (2003), o governo eletrônico está passando por cima de referências culturais, econômicas e religiosas para se firmar como uma das principais instituições do terceiro milênio.

Com isso, diversos países têm investido nessa idéia, e prova disso são os seus investimentos nessa área, como, por exemplo, a criação de seus portais eletrônicos governamentais. Esse fenômeno é mundial. De um lado do mundo, temos os EUA, que estão apostando forte nessa iniciativa através de seu principal portal governamental, o "*First Gov*". Do outro lado, temos Cingapura, que não possui os mesmos referenciais econômicos, culturais e religiosos tal qual os EUA, mas que também está concentrando seus esforços no governo eletrônico, com o seu "*Government On-line Portal*".

O Brasil tem conquistado também o seu espaço no âmbito de governo eletrônico, o qual lhe proporcionou uma importante conquista: o título de país exemplar nas Américas para informações governamentais na Internet. Takahashi (2000, p. 74) afirma que "a partir de 1995, concomitante com a explosão da Internet no País, aumentou grandemente a quantidade de *sites* do Governo Federal na rede, e análise

independente de princípios de 1996 chegou a apontar o Brasil como o exemplo nas Américas para informações governamentais na Internet. Nos anos recentes, a variedade de informações aumentou, não somente no nível federal, como especialmente no nível estadual”.

Apesar do alcance nacional e internacional de projetos de e-gov, pesquisas têm mostrado o insucesso da maioria das iniciativas de governo eletrônico no alcance do público cidadão (BBC *apud* PACHECO, 2003). Uma das razões para esse cenário está na etapa de levantamento de requisitos e necessidades de informação junto aos usuários dos sistemas de e-gov, na qual há exigência de que os resultados dos trabalhos estejam em padrões generalizáveis e intercambiáveis de informação entre diferentes *sites* de governo. Somente essa interoperabilidade vai garantir ao cidadão a visão de governo para *sites* públicos. Com ela, serviços compatíveis passam a se referenciar e a dispensar a multiplicidade de instâncias do mesmo pedido de informação (e.g., solicitação de dados cadastrais em *sites* de governo cujo pedido já havia sido efetivo em outros *sites* de governo). Para tal, o processo de levantamento de requisitos e identificação das informações necessárias em projetos de e-gov necessita de uma metodologia que permita que essa tarefa crie a dimensão fundamental da informação no domínio de interesse, o que é denominada unidade de informação, padronizável, generalizável e referenciável.

Essas unidades de informação são entendidas como modelos de informação ou metadados para um determinado domínio de interesse. Pacheco (2003) relata que as unidades de informação descrevem subdomínios da área-fim para a qual uma plataforma de sistemas está sendo desenvolvida. Castoldi (2003) afirma que as unidades de informação são formadas por classes ou elementos do domínio da plataforma aos quais estão associados conteúdo, processos e serviços específicos.

1.2 DEFINIÇÃO DO PROBLEMA

Para desenvolvimento de uma metodologia que permita a geração de padrões técnicos ou de uma arquitetura que possa contribuir na definição de metadados para informações governamentais, foram considerados alguns questionamentos como:

- ⌘ *Metodologias tradicionais de levantamento de requisitos e especificação de sistemas servem para definição de unidades de informação para plataformas de sistemas de informação para governos?*
- ⌘ *Como identificar unidades de informação em uma área de governo, de forma a permitir a construção de uma plataforma integrada para gestão da informação pública?*

A primeira questão visa verificar a eficácia das metodologias tradicionais de levantamento de requisitos na identificação de unidades de informação em projetos de governo eletrônico. A fim de que as unidades de informação identificadas e especificadas em projetos de governo eletrônico possam ser padronizadas de forma articulada com os demais componentes de um projeto de e-gov, considera-se que é necessário atentar para uma possível variação nos procedimentos tradicionais de levantamento de requisitos, que pode ocorrer quando o sistema for de administração pública. A metodologia para efetivar a definição de unidades de informação nesses projetos é a resposta para a segunda questão de pesquisa, que deverá advir da metodologia proposta nesta dissertação para definição de unidades de informação padronizadas e de sua aplicação a um caso real de governo eletrônico.

1.3 ARQUITETURA REFERÊNCIA

Além de propor a metodologia para especificar e padronizar as unidades de informação, a presente dissertação visa contextualizar tal metodologia segundo a relação dessas unidades em arquiteturas de e-gov, apontando a necessidade da padronização na garantia da interoperabilidade em projetos de governo eletrônico. É necessário que essas unidades sejam compreendidas por sistemas de informação, *sites*, portais e por sistemas de extração de conhecimento desenvolvidos no projeto de e-gov.

A metodologia, foco desta dissertação, deve estar articulada com uma arquitetura integrada de definição de projetos de governo eletrônico, de forma a permitir a sua utilização em arquiteturas e metodologias para construção de governo eletrônico.

Embora o trabalho desta dissertação seja independente e possível de ser aplicado e contextualizado em demais metodologias e arquiteturas para construção de

governo eletrônico, a presente dissertação terá como base referencial a metodologia de desenvolvimento de arquitetura integrada de sistemas de governo eletrônico utilizada no estudo de caso da Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Esse projeto tem por base a *Metodologia de Desenvolvimento de Plataformas de Governo para Geração e Divulgação de Informações e de Conhecimento*, identificada por Metodologia Stela e-gov e relatada no trabalho de Pacheco (2003), cujo conteúdo apresenta componentes metodológicos e tecnológicos na forma de uma arquitetura que objetiva preservar a relação conceitual entre seus elementos e que é construída a partir da aplicação da metodologia proposta.

De acordo com Pacheco (2003), uma das principais demandas de governo é a geração contínua de informações atualizadas sobre as diversas áreas de ação do Estado bem como a produção de conhecimento a partir dessas áreas, de forma a permitir revisão e melhorias constantes dos processos de gestão dos serviços públicos. Para atender a essa necessidade, o autor propõe uma arquitetura com os seguintes componentes metodológicos e tecnológicos: a) unidades de informação, definição de ontologias e representação em uma linguagem de marcação (XML); b) fontes e sistemas de informação; c) portais e serviços Web; e d) sistemas de conhecimento e agentes. Essa proposta pode ser mais bem visualizada na Figura 1.

Considerando-se que projetos de governo eletrônico ultrapassam o escopo de análise de sistemas organizacionais, a proposta deste trabalho de construção de uma metodologia para a definição de unidades de informações em plataformas de sistemas de governo eletrônico visa contribuir e se enquadrar na arquitetura criada por Pacheco (2003).

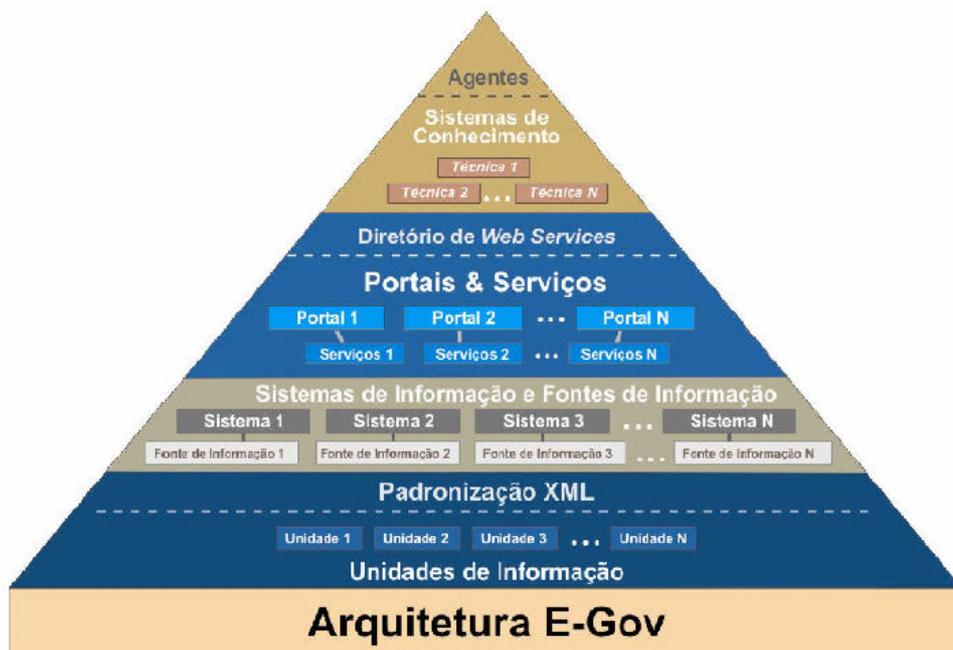


Figura 1 – Arquitetura conceitual proposta para projetos de e-gov (PACHECO, 2003)

1.4 OBJETIVO GERAL

Este trabalho tem como objetivo principal propor uma metodologia que possibilite a definição de unidades de informação em plataformas de sistemas para construção de governo eletrônico, de forma que possa ser seguida pelo especialista do negócio ou pelo engenheiro de software, com o intuito de definir a unidade de informação para o negócio ou para a aplicação.

1.5 OBJETIVOS ESPECÍFICOS

Para alcançar o objetivo principal deste trabalho, serão considerados os seguintes objetivos específicos:

- ✍ levantar informações teóricas sobre o desenvolvimento de plataformas de governo, incluindo governo eletrônico;
- ✍ abordar as técnicas da engenharia de software para o levantamento de requisitos;
- ✍ caracterizar as etapas para desenvolvimento de uma unidade de informação, tais como identificação, especificação, padronização e gestão;

- ✍ elaborar e descrever as diretrizes para cada etapa caracterizada para desenvolvimento de uma unidade de informação;
- ✍ avaliar a importância da especificação dessas unidades para a plataforma de sistemas de governo eletrônico; e
- ✍ aplicar a metodologia proposta no estudo de caso da Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), desenvolvida segundo a metodologia “Stela e-gov” (PACHECO, 2003).

1.6 JUSTIFICATIVA

A proposta para as unidades de informação consiste na apresentação de uma metodologia que permita a sua definição ou desenvolvimento. A necessidade para propor tais unidades pode ser explicada pela afirmação de Takahashi (2000, p. 78): “É preciso gerar padrões técnicos para aplicações governamentais”. O autor cita uma das ações necessárias a serem feitas para o desenvolvimento da sociedade da informação no Brasil: “Definir um modelo de metadados para informações governamentais”.

Através da metodologia proposta neste trabalho para a definição de unidades de informação, espera-se alcançar as metas do Programa Sociedade da Informação do Ministério da Ciência e Tecnologia (MCT) e do Portal de Serviços e Informações de Governos (e-gov), os quais promovem: (1) a informatização da administração pública; (2) o uso de padrões em seus sistemas aplicativos; e (3) a concepção, a prototipagem e o fomento a aplicações em serviços de governo, especialmente os que envolvem ampla disseminação de informações (como é o caso das plataformas de governo eletrônico) (MPOG, 2003).

1.7 METODOLOGIA

Existem várias formas de se classificar uma pesquisa. As mais tradicionais salientam os seguintes pontos: natureza da pesquisa, forma de abordagem do problema, seus objetivos e procedimentos técnicos (SILVA; MENEZES, 2000).

Com relação à natureza, a pesquisa pode ser classificada como Aplicada. Esse tipo de classificação é descrito por Silva e Menezes (2000) como uma pesquisa cujo objetivo é gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos, envolvendo verdades e interesses locais. Considerando-se o objetivo proposto neste trabalho, pressupõe-se o enquadramento nesse tipo de pesquisa.

A metodologia proposta neste trabalho, conforme citado anteriormente, visa abordar as etapas essenciais a fim de possibilitar uma seqüência de procedimentos que poderão ser seguidos para definição de uma unidade de informação. Com isso, no tocante à forma de abordagem do problema pode-se enquadrar a pesquisa como pesquisa descritiva, classificada também como pesquisa qualitativa. Silva e Menezes (2000) definem que nesse tipo de pesquisa o processo e seus significados são os focos principais da abordagem.

No que diz respeito aos objetivos, a pesquisa pode ser classificada como Exploratória. Segundo Richardson (1999), nesse tipo de pesquisa o objetivo é conhecer as características de um determinado fenômeno para que, posteriormente, seja possível obter explicações das causas e conseqüências de tal fenômeno. Esse tipo de pesquisa visa também proporcionar maior familiaridade com o problema, com vistas a torná-lo explícito ou a construir hipóteses.

A pesquisa exploratória assume, em geral, as formas de Pesquisa Bibliográfica e Estudo de Caso (GIL, 2000 *apud* SILVA E MENEZES, 2000, p. 21). Na presente dissertação, a pesquisa bibliográfica permitiu identificar as diferentes metodologias de levantamento de requisitos e sua aplicação ao caso de identificação e especificação de unidades de informação em projetos de governo eletrônico. A partir dessa análise, procede-se tanto à especificação da metodologia para definição de unidades de informação quanto à sua aplicação ao caso real da Plataforma Lattes do CNPq.

Do ponto de vista dos procedimentos técnicos, a pesquisa pode ser classificada como uma Pesquisa Bibliográfica e também como um Estudo de Caso (Gil *apud* Silva e Menezes, 2000).

Definidos os tipos de pesquisa nos quais o presente trabalho se enquadra, citam-se as três etapas que o fundamentam, a fim de que sejam efetivados os objetivos propostos:

(a) estudo de sistemas eletrônicos para governo e arquitetura para desenvolvimento desses sistemas, e, em particular, a representação das unidades de informação nessa arquitetura;

(b) descrição das etapas para desenvolvimento dos processos de identificação de uma unidade de informação;

(c) descrição das etapas para desenvolvimento dos processos de padronização de uma unidade de informação; e

(d) aplicação da metodologia proposta para definição de unidades de informação padronizadas em C&T (Plataforma Lattes - CNPq).

A Figura 2 ilustra uma visão esquemática da metodologia de construção do trabalho.

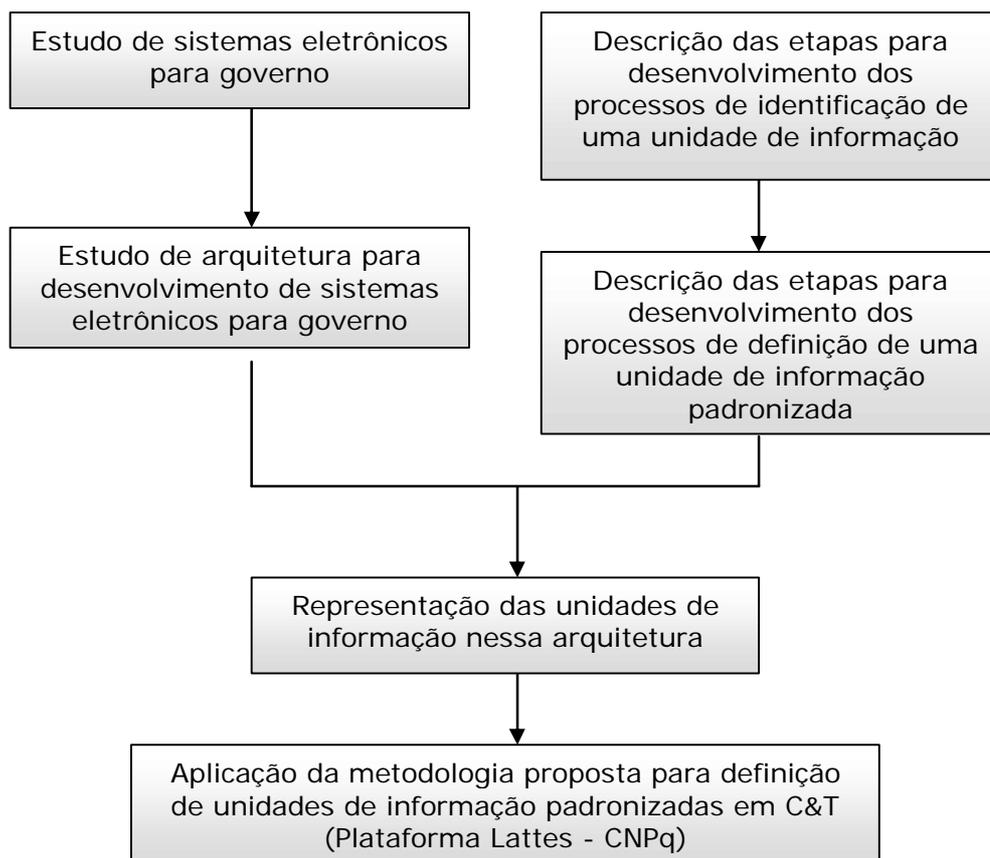


Figura 2 – Metodologia do trabalho

1.8 ESTRUTURA DO TRABALHO

O presente trabalho divide-se em cinco capítulos, os quais visam à abordagem de questões relacionadas à proposição de uma metodologia que possibilite a identificação, especificação, padronização e gestão de unidades de informação para plataformas de sistemas de governo. Assim sendo, os capítulos seguintes estão estruturados da seguinte forma:

Capítulo 2: Sistemas de e-gov e unidades de Informação – capítulo em que são apresentados conceitos e questões relacionadas à construção de plataformas de sistemas, incluindo-se as considerações da engenharia de software sobre o assunto. Nesse capítulo é contextualizada a arquitetura metodológica para desenvolvimento de plataformas de governo eletrônico, definida por Pacheco (2003);

Capítulo 3: Metodologia para definição de unidades de informação em plataformas de e-gov – nesse capítulo aborda-se uma descrição das etapas necessárias para o desenvolvimento de uma unidade de informação e as diretrizes, as metodologias, os cuidados, as indicações e as recomendações metodológicas referentes a esse desenvolvimento;

Capítulo 4: Aplicação: gestão de unidades de informação em C&T - Plataforma Lattes – nesse capítulo é aplicada a metodologia proposta para identificação, especificação, padronização das unidades de informação da Plataforma Lattes do CNPq bem como a gestão dessas unidades.

Capítulo 5: Conclusões e futuros trabalhos – Nessa parte final do trabalho são relatadas as conclusões sobre o trabalho e as recomendações para futuros trabalhos.

2 SISTEMAS DE E-GOV E UNIDADES DE INFORMAÇÃO

Governo eletrônico é o uso da tecnologia para reinventar relacionamentos e colocar o cidadão no centro.

Davide Vigano

2.1 APRESENTAÇÃO

Apresentando os principais conceitos e questões relacionados à construção de uma plataforma de sistemas para governo eletrônico e as considerações da engenharia de software sobre o assunto, é contextualizada nesse capítulo a relação das unidades de informação em arquiteturas de e-gov que visam à gestão de suas unidades, mencionando arquiteturas e metodologias para construção de governo eletrônico, como a *Metodologia de Desenvolvimento de Plataformas de Governo para Geração e Divulgação de Informações e de Conhecimento*, identificada como Metodologia Stela e-gov e descrita no trabalho de Pacheco (2003).

Tal metodologia Stela e-gov, que também é citada como arquitetura referência da metodologia proposta nesta dissertação para padronização, é utilizada para fins de contextualização da aplicação da metodologia proposta nesta dissertação. Contudo, o trabalho desta dissertação é independente e poderá ser aplicado em demais metodologias e arquiteturas para construção de governo eletrônico. A opção por utilização da Metodologia Stela e-gov também se dá pelo fato de ela possuir uma camada explícita e específica em sua arquitetura para tratar unidades de informação e sua padronização.

Ainda neste capítulo, com base na literatura das áreas da engenharia de software e outras, as quais apontam diretrizes para encontrar uma unidade de informação em um domínio de determinado problema, são abordados os consideráveis conceitos com o intuito de fundamentar a identificação, especificação e padronização de uma unidade de informação. Entre esses conceitos, além de outros, citam-se o levantamento de requisitos, a análise orientada a objetos, a reengenharia de processos, a análise de sistemas legados verificando entidades e relacionamentos previamente utilizados, a consulta a representantes da comunidade de usuários e outros.

2.2 GOVERNO ELETRÔNICO E SISTEMAS E-GOV

Administrações públicas não têm se tornado mais focadas em seus cidadãos devido aos poucos esforços realizados durante sua trajetória de existência no sentido de induzir uma mudança ou reinvenção de cultura dentro do governo, de modo que trabalhassem com o compartilhamento de informação em equipe e fossem trabalhadores do conhecimento (HOLMES, 2001).

Holmes (2001) cita que, para a reinvenção do governo, é extremamente necessário o surgimento de um habilitador. Algo que viabilize o trabalho em equipe e permita o compartilhamento da informação. Algo que excite os políticos e a imprensa. Que seja proveitoso para os negócios, relevante para o dia-a-dia dos cidadãos e se dirija às suas necessidades e circunstâncias pessoais. Algo como a Internet (*World Wide Web*).

É fácil ser cínico sobre novas tecnologias como um remédio para todas as doenças da sociedade, e algumas vezes é difícil separar a Internet de outros tipos de soluções tecnológicas. O Governo, particularmente, tem uma história de inserção de seus créditos em grandes projetos de tecnologia da informação (TI) somente para vê-los funcionar e mais tarde acabar o orçamento e agitar a confiança pública. A Internet está mostrando ser algo diferente e único no sentido da permissão da abertura de novos caminhos em comunicação, trabalho, compras, ensino e outras atividades. Ela pode agora habilitar uma nova forma de governar (HOLMES, 2001). Essa questão pode ser firmada pelas declarações feitas por Scacchi (2002). Segundo ele, o governo eletrônico, também designado pelo termo e-gov – definição estabelecida para governo eletrônico pelo relatório *National Performance Review and the Government Information Technology Services Board*, emitido em 3 de fevereiro de 1997 (AMÉRICA ACCESS, 1997, *apud* RELYEA, 2002, p. 1) – encoraja a adaptação das práticas modernas de tecnologia da informação que exploram a Internet e as capacidades do comércio eletrônico para aprimorar as operações do governo e os serviços públicos.

O governo eletrônico, ou e-gov, é o uso da tecnologia da informação, em particular da Internet, para disponibilizar serviços públicos em muito mais comodidade, de forma orientada ao cliente, com um custo compensador. O movimento de governo eletrônico está sendo dirigido com o intuito de cortar custos e fornecer eficiência, reunir

expectativas dos cidadãos, fornecer relacionamentos entre eles e facilitar o desenvolvimento econômico (HOLMES, 2001).

Costake e Jensen (2002) consideram o e-gov um componente importante da sociedade da informação no âmbito local, regional, setorial, nacional e internacional. Esses autores classificam o e-gov como um instrumento principal para o melhoramento das atividades dos poderes Executivo e Judiciário e de autoridades legislativas.

Tambouris, Gorilas e Boukis (2001) observam o e-gov como uma transformação fundamental de governo em escala, não vista desde o início da Era Industrial. Essa transformação é baseada na aplicação de tecnologias de informação e comunicação para transformar com eficiência, efetividade, transparência e responsabilidade as mudanças transacionais dentro do governo, entre cidadão e agências federais, municipais e locais do governo através do acesso e do uso da informação.

Embora já iniciada essa transformação, é algo que ainda não se tem visto na maioria das ações dos líderes do governo. Holmes (2001) faz uma analogia da inserção do setor privado no século XX transformado pelo *e-commerce* e pelo *e-business* como algo já realizado, e o proveito da Internet, que, segundo ele, não tem permeado os pensamentos da maioria dos oficiais públicos e políticos. Em síntese, podemos inferir o quanto tem que ser trilhado para que se possa alcançar o setor privado na área de governo eletrônico.

Tambouris, Gorilas e Boukis (2001) citam algumas dimensões que representam um conjunto potencial de esforços que precisam ser feitos. Entre eles estão cooperação organizacional, financiamento e suporte político, infra-estrutura técnica e integração entre processos existentes.

De acordo com Devadoss, Pan e Huang (2002), é notado na literatura que muitos dos estudos existentes não são empíricos, ou seja, não são apenas baseados na experiência, são sem caráter científico e também retóricos ou superficiais no foco natural da estrutural transacional dos processos requeridos nas iniciativas. Podemos considerar essas iniciativas como o conjunto de esforços a serem feitos abordando as dimensões citadas por Tambouris, Gorilas e Boukis (2001).

Lenk e Traunmuller, citados por Devadoss, Pan e Huang (2002) consideram que essas iniciativas podem ser entendidas em (a) perspectivas de negócios; (b)

perspectivas de cidadãos; (c) perspectivas de conhecimento; (d) perspectivas de processo; e (e) perspectivas de telecooperação.

Schubert e Hausler (2001) definem que as perspectivas de negócios consideram o uso e o desenvolvimento de informação e comunicação para elevar a funcionalidade do governo. O impacto *e-businesses* no setor público é a fonte da transformação do governo em direção ao governo eletrônico (WIMMER; TRAUNMÜLLER; LENK, 2001).

O desenvolvimento de tecnologias de *e-commerce* pode servir para incrementar o sucesso do cidadão à informação e fornecer funcionalidade para o governo (CSETENYI *apud* DEVADOSS; PAN; HUANG, 2002).

A perspectiva de conhecimento identifica funcionários ou colaboradores como fontes em suas respectivas funções. Enquanto se replanejam as ações para o governo eletrônico, o conhecimento dos funcionários deve ser suficientemente utilizado, garantindo assim uma prevenção de perda de conhecimento. Isso fornece a continuação do saber acumulado sobre o tempo (LENK; TRAUNMULLER *apud* DEVADOSS, PAN E HUANG, 2002).

De acordo com Watson e Mundy (2001), a perspectiva de processo trata a utilização de tecnologia de informação – TI para aumentar a entrega de serviço eficiente.

Enquanto outras perspectivas estão completamente claras a várias partes das atividades do e-gov, Devadoss, Pan e Huang (2002) abordam que a perspectiva de telecooperação fornece o recurso mais efetivo para explorar os estágios iniciais do desenvolvimento do governo eletrônico. Ela trata da interação de várias agências e parceiros de negócio envolvidos no processo de trabalho. Geralmente, nos primeiros estágios de qualquer projeto de e-gov, a maioria das agências independentes precisa se inteirar no processo, exigindo, muitas vezes, uma mudança (no sentido de adaptação) de tecnologia, reengenharia do processo, novas posições de trabalho e treinamento dos funcionários para que eles atuem nessas novas posições.

Definidas as iniciativas, também classificadas como perspectivas, que representam um conjunto potencial de esforços a serem realizados para o desenvolvimento de governo eletrônico, a seguir é tratado o trabalho realizado nessa

esfera, focalizando sistemas de informações que visam disponibilizar serviços de governo eletrônico.

Gouscos, Mentzas e Georgiadis (2001) classificam os serviços eletrônicos oferecidos pelo governo, nos âmbitos nacional, regional e local para negócios como G2B¹ (*Government to Businesses*), para os cidadãos como G2C (*Government to Citizens*), ou para o próprio governo como G2G (*Government to Government*).

A seguir, abordam-se notícias relacionadas a fatos e medidas sobre governo eletrônico no âmbito mundial que nos permitem perceber a situação atual e nos fazem pensar no quanto pode ser feito ainda nesse segmento:

Holmes (2001) resume que:

- ✍ 75% dos australianos registram seus impostos de renda através da Internet;
- ✍ Singapura fornece 150 serviços públicos em um único portal;
- ✍ a Federação dos Estados Unidos faz anualmente quatro milhões de aquisições *on-line* de mercadorias e serviços, que giram em torno de US\$ 17 bilhões;
- ✍ custos administrativos no Departamento de Agricultura dos Estados Unidos da América foram reduzidos de US\$ 77 por transação para US\$ 17 depois da implantação de um sistema eletrônico de aquisição;
- ✍ o departamento de polícia da Escandinávia demora 25 segundos para transmitir a imagem da impressão digital de um indivíduo para outra autoridade;
- ✍ *Monterrey Tech* é o sistema de educação mais disperso no mundo, fornecendo cursos *on-line* para 43 mil estudantes por todo o México e América Latina;
- ✍ a Estônia não tinha computadores pessoais em 1991. Hoje ela está entre os 20 países mais conectados no mundo;

¹ Os acrônimos utilizados são oriundos do inglês. Sendo esses termos amplamente adotados, optou-se pela não tradução.

- ✍️ petições públicas para o parlamento da Escócia são criadas e submetidas através de sistemas eletrônicos; e
- ✍️ brasileiros votam eletronicamente em eleições nacionais e locais.

Outros casos de e-gov, agora mais focados no âmbito nacional, como o da Plataforma Lattes (<http://www.lattes.cnpq.br>), mantida pelo CNPq, podem ilustrar o porquê da considerável classificação brasileira conquistada no quesito serviços on-line de governo. Lançada em 16 de agosto de 1999, com a primeira versão do Sistema de Currículo Lattes, sistema pertencente à plataforma de sistemas de governo eletrônico em C&T, conhecida como Plataforma Lattes, proporcionou um aumento significativo no número de currículos enviados ao CNPq (CNPq, 2003). Além disso, essa plataforma de sistemas de e-gov, que será comentada mais adiante, foi premiada neste ano como primeira colocada na 3ª edição do Prêmio e-gov brasileiro, na categoria governo para cidadão – G2C (PRÊMIO E-GOV, 2004).

Entre outros projetos de destaque que confirmam a importância do Brasil nesse âmbito, cabe citar outros vencedores do Prêmio e-gov, que foi instituído pela Associação Brasileira de Empresas Estaduais de Processamento de Dados (ABEP), pela Escola Brasileira de Administração Pública e de Empresas da Fundação Getúlio Vargas (EBAPE/FGV) e pelo Ministério do Planejamento, Orçamento e Gestão do Governo Federal, e se encontra atualmente na sua terceira edição. A seguir, apresentam-se na Tabela 1, os projetos vencedores em suas categorias:

Projeto	Autor	Premiação - Categoria
Agência Goiana de Administração e Negócios Públicos - GO	Goiás Digital	1º lugar - Prêmio Excelência em Governo Eletrônico
Plataforma Lattes	Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - DF	1º lugar - Governo para Cidadão – G2C
Departamento de Inteligência da Polícia Civil do Estado de São Paulo - DIPOL	OMEGA: Sistema de Suporte à Investigação Policial	1º lugar – Governo para Governo – G2G
Secretaria da Indústria, Comércio e Mineração do Estado da Bahia	Portal Bahialvest / BahiaExport	1º lugar – Governo para Negócio – G2B

Tabela 1 – Projetos premiados no III Prêmio Excelência em Governo Eletrônico (PRÊMIO E-GOV, 2004)

Projetos como esses podem ter contribuído significativamente para a inserção do Brasil no Relatório Mundial de Tecnologia de Comunicação e Informação. Segundo a Sociedade Brasileira de Computação - SBC (2003), o Brasil se classificou em 29º lugar entre os 82 países analisados no Relatório Mundial de Tecnologia de Comunicação e Informação 2002-2003, do Fórum Econômico Mundial (*World Economic Forum*). O estudo verifica o quanto cada um dos países está preparado para a economia em rede. Com essa classificação, o País melhorou sua posição em relação ao relatório passado, quando ficou em 38.º lugar. A posição brasileira se deu devido à boa pontuação obtida pelo País no quesito governo eletrônico. No quesito serviços on-line de governo, o Brasil conquistou o 8º lugar.

Considerando-se a melhora do Brasil na classificação diante do relatório anterior e o número de países analisados pelo Relatório Mundial de Tecnologia de Comunicação e Informação, que chega a 82, conclui-se que se trata de uma boa posição no ranking.

Com base nessas considerações podemos ter uma visão parcial do que já foi alcançado e do quanto se pode trilhar nessa área. Accenture (2003) afirma que o e-gov está se movendo acima da curva da maturidade, mas que ainda tem um longo caminho a trilhar.

2.3 ARQUITETURA DE SISTEMAS E-GOV

Pacheco (2003) afirma que a ampla variedade de abordagens no conjunto de processos e, sobretudo, de recursos tecnológicos, dificulta a integração de esforços por parte do Estado e pode gerar a ineficácia de investimentos.

A disposição de modelos e arquiteturas que permitam facilitar o alcance do primor nos processos relacionados a governo eletrônico é evidente e indispensável. Tambouris *et al.* (2002) consideram que o desenvolvimento de plataformas de e-gov deveria ser baseado em modelos apropriados, que poderiam habilitar o aprimoramento dos processos e superar as barreiras organizacionais e culturais.

Com isso pode-se pressupor a importância do uso de arquiteturas para a definição de governo eletrônico. No entanto, o estudo conceitual de arquitetura de

sistemas transcende o escopo de trabalho. Serão abordados nesta dissertação exemplos práticos de arquiteturas ou propostas integradas de metodologias para agências de governo. Ainda são considerados os componentes arquiteturais, metodológicos, e conceituais da arquitetura referência deste trabalho proposta por Pacheco (2003).

A seguir, é apresentada a definição estabelecida pelo Grupo de Trabalho da FEA - *Federal Enterprise Architecture* (Arquitetura de Empreendimento Federal) e adotada pelo Comitê de Arquitetura e Infra-Estrutura do Conselho Federal de Informática dos Estados Unidos.

A definição proposta pelo Grupo de Trabalho da FEA foi descrita no documento “*E-gov Enterprise Architecture Guidance - Common Reference Model*” (Guia de orientação de arquitetura de empreendimento de governo eletrônico – modelo de referência comum), publicado e assinado em 25 de julho de 2002. O documento proposto foi considerado pelos presidentes da FEA como um guia para o Governo Federal Norte-Americano alcançar soluções tecnológicas de e-gov mais efetivas e interoperacionais para satisfazer as necessidades da missão administrativa (FEA, 2002).

A arquitetura proposta pela FEA, relatada no documento citado e adotada pelo Conselho Federal de Informática dos Estados Unidos (FEA, 2002), foi estruturada em quatro níveis segmentados com seus respectivos modelos, conforme é apresentado na Figura 3.

Esses quatros níveis consistem nas seguintes arquiteturas: Arquitetura de Negócios (*Business Architecture*), Arquitetura de Dados (*Data Architecture*), Arquitetura de Aplicação (*Applications Architecture*) e Arquitetura Tecnológica (*Technology Architecture*).

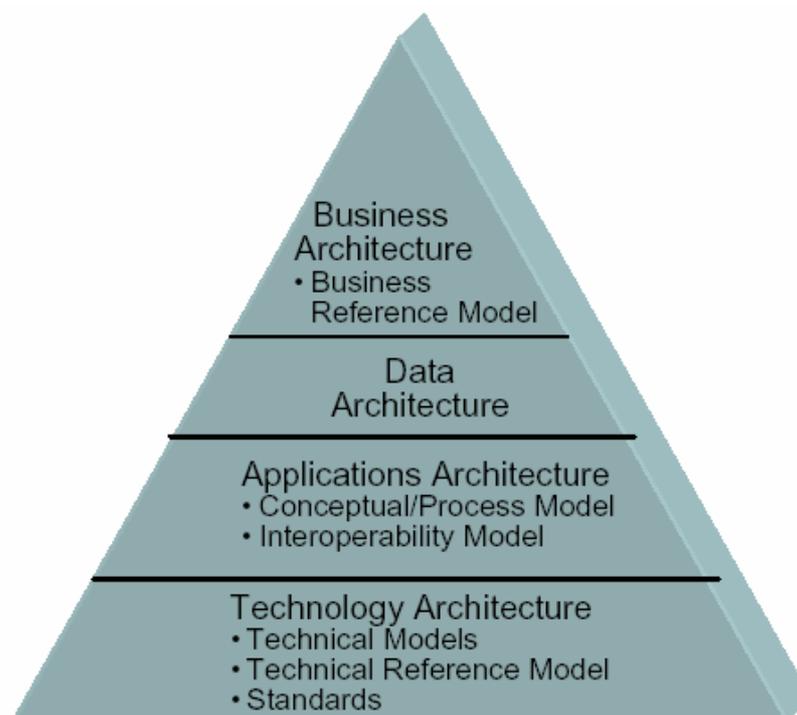


Figura 3 – Arquitetura proposta pela FEA (2002) segmentada em quatro níveis

Os modelos de referência encontrados nas arquiteturas são essencialmente descrições conceituais para estabelecer uma linha de base dos conceitos arquitetônicos do governo eletrônico efetivo e de um vocabulário comum. Os modelos e padrões associados com Arquitetura Tecnológica apresentam orientação e exemplos mais pragmáticos para as iniciativas de e-gov (FEA, 2002).

A Figura 4, a seguir, mostra como a FEA associa as arquiteturas definidas no documento “*E-gov Enterprise Architecture Guidance - Common Reference Model*” (Guia de Orientação de Arquitetura de Empreendimento de Governo Eletrônico – Modelo de Referência Comum) com os referidos Modelos de Referência: Administrativo ou de Negócios (*Business Reference Model - BRM*), de Performance (*Performance Reference Model - PRM*), de Capacidade de Aplicação (*Application Capability Reference Model - ARM*), de Dados e Informação (*Data and Information Reference Model - DRM*), e de Referência Técnica (*Technology Reference Model - TRM*).

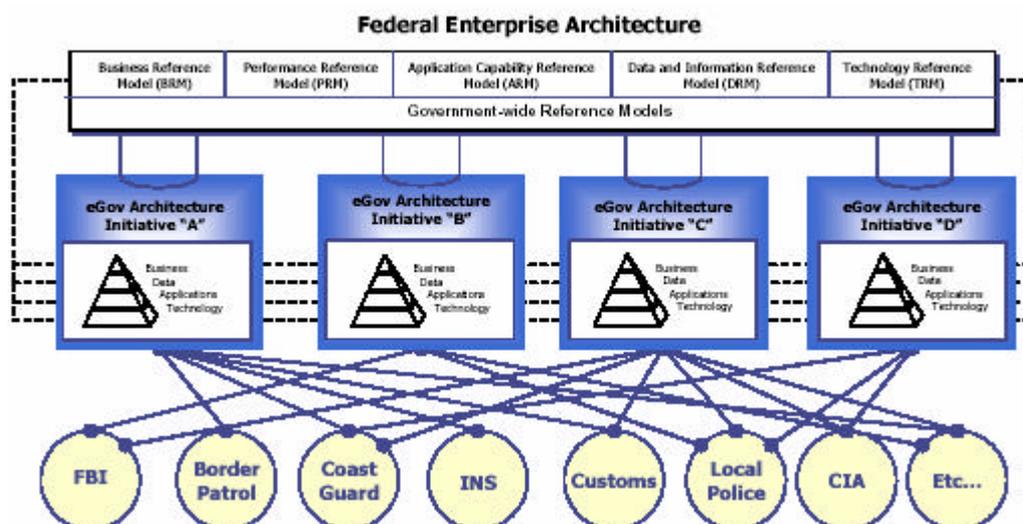


Figura 4 – Arquitetura proposta pela FEA (2002) associada com os modelos de referência e agências de E-gov norte-americanas

É apresentada a seguir uma descrição sucinta das arquiteturas propostas pela FEA baseada em seu guia de orientação (FEA, 2002).

A Arquitetura de Negócios ou Administrativa (*Business Architecture*) identifica as funções, os processos, a organização e o fluxo de informação para realizar a missão de uma instituição. As soluções de e-gov frequentemente envolvem as de negócio, que vão além dos limites funcionais ou organizacionais tradicionais – ambos dentro e através das agências e com constituintes externos, tais como os cidadãos, os governos locais e estaduais, e a indústria.

A Arquitetura de Dados (*Data Architecture*) define os principais tipos de dados necessários para auxiliar a administração, seu significado e sua forma. Vocabulário e definições de dados comuns são especialmente críticos para as soluções de governo eletrônico, as quais frequentemente atravessam os limites tradicionais, organizacionais, funcionais e do sistema. Isso inclui não apenas dados operacionais, mas também dados analíticos e conteúdo da *Web*. São relatadas na Arquitetura de Dados considerações importantes para as iniciativas de governo eletrônico, como princípios de operabilidade de dados, utilização da linguagem de marcação XML para o fornecimento de uma fundação crítica para as arquiteturas de dados de governo eletrônico, privacidade e integração física de dados.

A Arquitetura de Aplicação (*Application Architecture*) define as aplicações e habilidades de suporte para efetivamente gerenciar os dados e as informações necessários para auxiliar os objetivos administrativos.

A Arquitetura Tecnológica (*Technology Architecture*) define questões sobre *hardware* e *software* para auxiliar as aplicações e funções administrativas. Essa arquitetura associa o Modelo de Referência Técnica de governo eletrônico (*e-gov Technical Reference Model – TRM*), apresentado na Figura 4. Diferentemente dos outros modelos de referência apenas citados nas demais arquiteturas, foi realizada nesse trabalho uma descrição do TRM e de seus elementos, devido ao fato de este modelo fornecer dentre os demais modelos de referência uma visão mais aproximada de arquitetura de sistemas com ênfase para governo eletrônico.

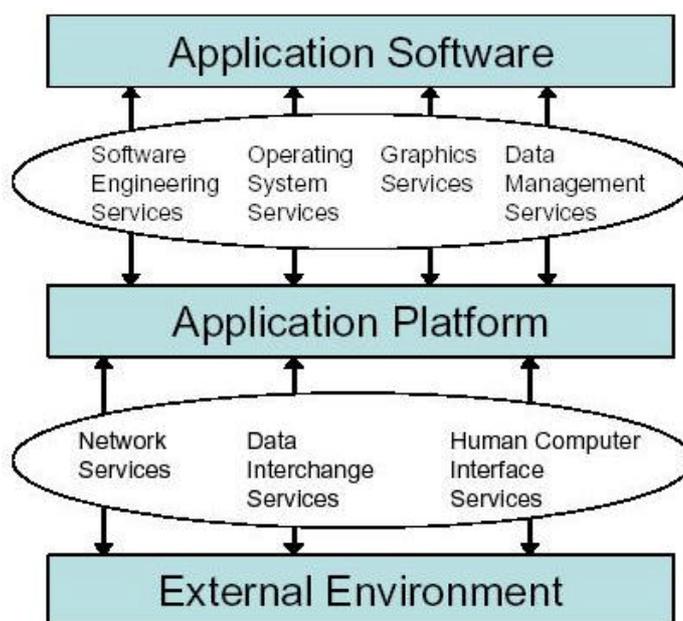


Figura 5 – Modelo de Referência Técnica de governo eletrônico – TRM (FEA, 2002, p. 26)

O propósito desse Modelo de Referência Técnica – TRM – é fornecer uma estrutura conceitual que defina um vocabulário técnico comum, de forma que as Agências Federais possam coordenar melhor aquisição, desenvolvimento e apoio organizados para as iniciativas de governo eletrônico. O TRM de e-gov engloba os elementos descritos a seguir.

- ? **Software de Aplicação**: inclui dados, documentação, treinamento e também programas.
- ? **Plataforma de Aplicação**: composta da coleção de componentes de hardware e software que fornecem serviços ou recursos de software para a sua aplicação.
- ? **Ambiente Externo**: elementos de sistemas que são externos ao software e à plataforma de aplicação. Por exemplo, serviços fornecidos por outras plataformas ou dispositivos periféricos.
- ? **Interface de Programa de Aplicação** (*Application Program Interface - API*): é a interface entre o software e a plataforma de aplicação. A sua funcionalidade primária é prover a portabilidade de software de aplicação.
- ? **Interface de Ambiente Externo** (*External Environment Interface - EEI*): é a interface que auxilia a transferência de informação entre a plataforma de aplicação e o ambiente externo, e entre as aplicações executadas na mesma plataforma. Consistindo principalmente de formatos de protocolos e de dados de suporte, o EEI auxilia a interoperacionalidade para uma grande extensão.

Outra arquitetura para governo eletrônico abordada nesta seção é a proposta metodológica tomada como arquitetura referencial para este trabalho, desenvolvida por Pacheco (2003).

Conforme já mencionado neste trabalho, Pacheco (2003) apresenta uma proposta integrada de metodologia e arquitetura para a produção de plataformas tecnológicas para agências de governo produzirem e publicarem informações na Internet. Essa arquitetura ou metodologia de e-gov apresenta ainda os componentes metodológicos, conceituais e arquitetônicos de uma plataforma de informações de governo, desde a formação das bases de dados aos portais de conhecimento na área a que se destina a plataforma.

Essa arquitetura considerou como estudo de caso a Plataforma Lattes, projeto do Conselho Nacional de Desenvolvimento Científico e Tecnológico do Brasil - CNPq

para promover Ciência, Tecnologia e Inovação. Esse projeto, que é estudo de caso da metodologia abordada, serviu recentemente como base para a formação de rede internacional de intercâmbio de informações e de conhecimento entre as agências de governo, conhecida como Rede ScienTI (PACHECO, 2003).

Tal metodologia apresenta componentes metodológicos e tecnológicos na forma de uma arquitetura que tem por objetivo preservar a relação conceitual entre seus elementos, construída a partir da aplicação da metodologia proposta.

Conforme citado anteriormente, Pacheco (2003) relata que uma das principais demandas de governo é a geração contínua de informações atualizadas sobre as diversas áreas de ação do Estado bem como a produção de conhecimento a partir dessas áreas, de forma a permitir revisão e melhorias constantes dos processos de gestão dos serviços públicos. Com o intuito de atender a essa necessidade, ele propõe uma arquitetura com quatro componentes metodológicos e tecnológicos: a) unidades de informação, definição de ontologias e representação em uma linguagem de marcação (XML); b) fontes e sistemas de informação; c) portais e serviços Web; e d) sistemas de conhecimento e agentes.

A Figura 6 apresenta a arquitetura proposta por Pacheco (2003) com destaque para os quatros componentes a serem abordados a seguir.

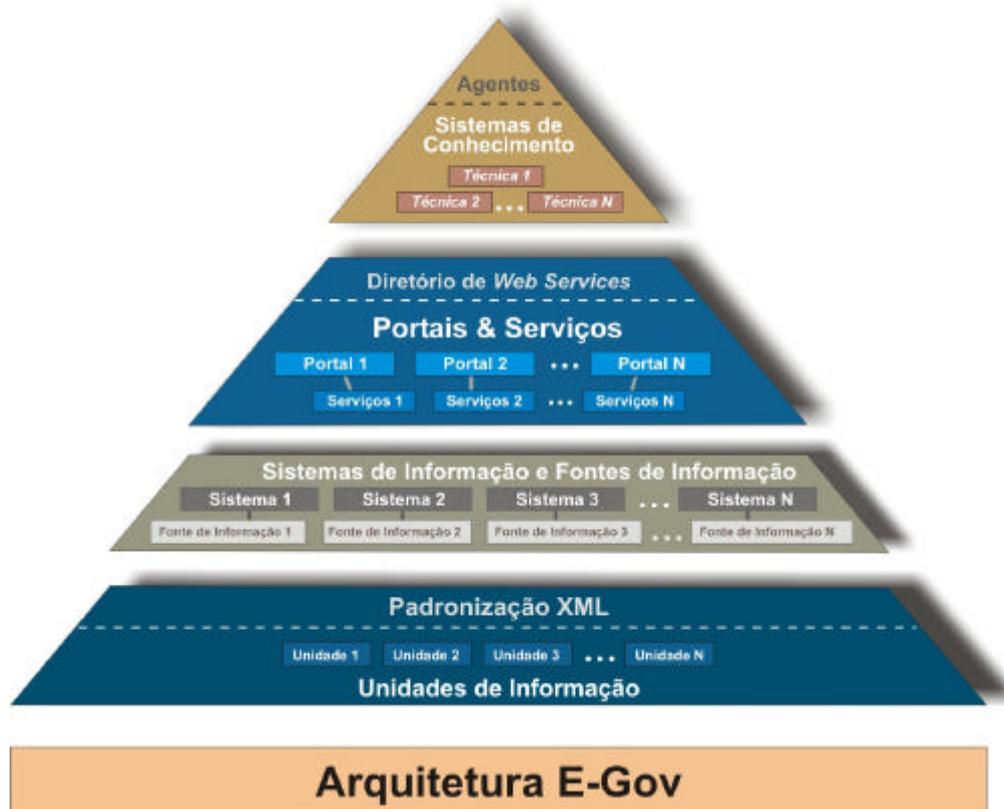


Figura 6 - Arquitetura para projetos de e-gov proposta por Pacheco (2003) com destaque para os seus quatro componentes

A seguir é apresentada uma descrição desses componentes metodológicos e tecnológicos que formam a arquitetura proposta por Pacheco (2003).

a) Camada de Unidades de Informação, definição de Ontologias e representação em uma Linguagem de Marcação (XML)

i) Camada de Unidades de Informação

Pacheco (2003) relata que cabe aos responsáveis pelo projeto a identificação das unidades, a especificação do conteúdo inicial, a proposição de um padrão inicial para essas unidades e a promoção contínua de uma revisão por parte da comunidade interessada. Ele afirma ainda que, para se proceder ao estabelecimento desses padrões, é necessário definir a ontologia das unidades de informação e sua padronização XML. As unidades de informação descreverão subdomínios da área-fim para a qual a plataforma está sendo desenvolvida. Os principais elementos metodológicos e tecnológicos dessa camada, a de unidades de informação, estão relacionados à padronização do conteúdo de cada unidade.

Comparando-se a camada de unidade de informação com as demais, as decisões no nível dessa camada terão impacto nas camadas superiores da arquitetura e nas características de generalidade, flexibilidade e interoperabilidade da plataforma.

ii) Definição de ontologias e representação em uma Linguagem de Marcação (XML)

Na arquitetura proposta, as ontologias deverão ser estabelecidas ao nível da primeira camada por se refletirem ao longo de todos os demais instrumentos da plataforma de governo. Pacheco (2003) considera que o desafio para os responsáveis será a construção de modelos ontológicos compreensíveis e dedicados aos diversos atores partícipes de uma plataforma de governo.

Essas e outras experiências deverão nortear os procedimentos de construção de comunidades de definição e atualização de ontologias na plataforma de governo. Para os instrumentos de informação e bases de dados, componentes da segunda camada da arquitetura, o resultado deverá ser a padronização XML de cada unidade de informação da plataforma.

Com base nisso, Pacheco (2003) estabelece em sua arquitetura que deverão ser definidos para cada unidade de informação metadados específicos, que deverão ser descritos em uma linguagem de marcação para promover a interoperabilidade dos sistemas e informações da plataforma. Khare e Rifkin (*apud* Pacheco, 2003) relatam que os últimos anos foram caracterizados por uma evolução crescente de padronização para interoperabilidade na Web, com a *eXtensible Markup Language* (Linguagem de Marcação Extensível - XML), sendo o padrão evolutivo de maior aceitação. Nesses termos, Pacheco (2003) propõe a utilização desse padrão para a especificação das unidades de informação da plataforma de governo. Assim como ocorre com os demais aspectos metodológicos da construção dos componentes tecnológicos, os responsáveis deverão estar atentos à evolução das formas de padronização e à interoperabilidade de seus sistemas, podendo optar por padrões que melhor atendam aos objetivos da plataforma.

b) Fontes e Sistemas de Informação

A camada de fontes e sistemas de informação, segundo Pacheco (2003), contempla os repositórios de cada unidade de informação da plataforma e os respectivos

sistemas de informação que captam, tratam e armazenam os dados da unidade junto à comunidade usuária. O desenvolvimento desses componentes tecnológicos tem por base as padronizações da camada inferior e seguem a metodologia e as tecnologias estabelecidas na fase de projeto da plataforma.

O propósito das fontes de informação, que constituem parte desta camada, são o armazenamento e a gestão das informações públicas, produzidas pela interação com a comunidade usuária ou por enlaces dinâmicos com outras fontes de informação. Já os Sistemas de Informação têm como um de seus principais objetivos garantir que as informações incluídas nas bases de dados sigam as regras e especificações estabelecidas para a unidade correspondente. Pacheco (2003) sugere em sua arquitetura que, para tal, devem incluir entre seus procedimentos de garantia de consistência e controle de obrigatoriedade de dados toda regra especificada na padronização.

c) Portais e Serviços Web

A terceira camada da plataforma proposta por Pacheco (2003) é composta (1) dos instrumentos desenvolvidos para apresentação de informações na Web, para publicação de informações com atualização dinâmica e com interação com a comunidade usuária, como, por exemplo, portais Web, e dos recursos de disseminação de serviços de informação de governo na Web - *Web Services*.

No domínio da arquitetura proposta, os portais têm por objetivo promover as fontes de informação da plataforma e torná-las disponíveis, segundo a classificação estabelecida na padronização das unidades. Pacheco (2003) considera ainda que serão excelentes subsídios para a hierarquização temática dos portais as ontologias elaboradas na padronização. Além do objetivo destacado dos portais, estes servem para divulgar os serviços e demais recursos elaborados para os usuários, sendo ainda um dos instrumentos de indução e gestão de comunidades virtuais.

Os Serviços Web (*Web Services*) surgem como proposição metodológica e tecnológica para a publicação de serviços do governo na Internet. Sendo assim, deverão valer-se dos padrões produzidos na camada de unidades de informação e serão recursos públicos dos portais da plataforma de governo. Combinados com sistemas de conhecimento, poderão, ainda, publicar serviços inteligentes.

d) Sistemas de Conhecimento e Agentes

No topo da arquitetura da plataforma de governo estão os sistemas de conhecimento e os agentes. Trata-se dos instrumentos projetados para gerar novos conhecimentos a partir das fontes de informação da plataforma e de sua operação por parte da comunidade usuária. Para elaborar os instrumentos dessa camada, os responsáveis pelo projeto deverão valer-se da área de descoberta de conhecimento, e, a partir de técnicas relacionadas como mineração de dados, estatística e reconhecimento de padrões, elaborar projetos específicos destinados à produção de conhecimento focado na área de governo atendida pela plataforma.

Os agentes surgem na arquitetura proposta por Pacheco (2003) como alternativa tecnológica e metodológica. Pacheco (2003) descreve a seguir essas duas alternativas. Em termos tecnológicos, os agentes inteligentes podem tornar os recursos dos sistemas de conhecimento disponíveis à comunidade usuária, da qual podem, também, obter informações e, por meio de procedimentos de inteligência, produzir novos conhecimentos no domínio da plataforma. Em termos metodológicos, a forma de conceber, especificar e elaborar esses agentes poderá ter reflexos nos procedimentos de projeto da plataforma.

Concluindo-se esta seção, pode-se inferir que a metodologia proposta nesta dissertação pode contribuir para projetos de e-gov que abordem a arquitetura proposta por Pacheco (2003) e para a proposta da FEA (2002), mais especificamente na Arquitetura de Dados.

Além das duas propostas apresentadas neste trabalho acredita-se que outras arquiteturas não relatadas que necessitem abordar o estabelecimento de unidades de informação padronizadas possam também tomar proveito da metodologia proposta nesta dissertação.

2.4 ENGENHARIA DE SOFTWARE EM ARQUITETURA DE E-GOV E UNIDADES DE INFORMAÇÃO

Abordaremos nesta seção uma visão da engenharia de software mais direcionada para a definição de unidades de informação em arquiteturas de governo

eletrônico. O estudo mais conceitual da engenharia de software transcende o escopo deste trabalho, contudo, será apresentada uma breve conceitualização dessa área.

O termo “Engenharia de Software” foi evidenciado em 1968 como resposta para o estado desolado da arte do desenvolvimento de software de qualidade com prazo estimado e dentro do orçamento previsto (BRUEGGE; DUTOIT, 2004).

Na coleção de padrões da IEEE (*apud* BOURQUE, 2002), engenharia de software é definida como a aplicação e o estudo da aproximação de uma sistemática, disciplinada e de quantificável acesso para o desenvolvimento, a operação e a manutenção de software, isto é, a aplicação da engenharia de software.

Farley (*apud* MELLON, 2004) concorda com as afirmações e cita que a engenharia de software é a disciplina tecnológica e gerencial preocupada com a produção planejada e com a manutenção de produtos de software, os quais são desenvolvidos e modificados com prazo estimado e dentro das estimativas de custo.

Bruegge e Dutoit (2004) fazem uma relação das atividades técnicas com a engenharia de software orientada a objetos. As atividades de desenvolvimento dividem a complexidade através da construção e validação de modelos de um domínio de aplicação ou sistema. Segundo eles, essas atividades de desenvolvimento incluem: a) Levantamento de requisitos; b) Análise; c) Projeto de sistema; d) Projeto de objeto; e) Implementação e f) Testes.

Visando às unidades de informação, dirige-se entre essas atividades uma atenção especial à atividade de requisitos.

2.4.1 LEVANTAMENTO DE REQUISITOS

Bruegge e Dutoit (2004) afirmam que os requisitos especificam um conjunto de características que o sistema deve ter.

Kaindl (1997) salienta a relevância dos requisitos mais especificamente na análise ao considerá-la como uma das mais importantes partes do ciclo de vida de qualquer projeto de software.

Entretanto, pode-se inferir que a atividade de levantamento de requisitos é uma questão que demonstra muita relevância em engenharia de software e é algo que, caso

não seja abordado corretamente, pode comprometer o sucesso do produto, considerando ainda que as formas tradicionais de levantamento de requisitos podem variar em se tratando da sua utilização para definir unidades de informação embasadas em arquiteturas voltadas para governo eletrônico.

Jackson (1997) associa alguns casos de insucesso em software à definição dos seus requisitos. Segundo ele, uma das causas pode ser atribuída a erros na engenharia de requisitos de sistemas como requisitos maldefinidos.

A engenharia de requisitos tem a finalidade de estabelecer os requisitos do sistema sujeito à construção (BRUEGGE; DUTOIT, 2004).

Escrito por Bubenko (*apud* NATURE, 1996), o principal desafio da engenharia de requisitos tem sido como proceder da informal, individual e confusa declaração dos requisitos para uma especificação formal entendida e aceita por todos os interessados.

Bruegge e Dutoit (2004) propõem uma divisão da engenharia de requisitos em duas principais atividades; levantamento de requisitos, na qual resulta na especificação de sistemas que o cliente compreende, e a análise, na qual resulta em um modelo de análise que os desenvolvedores podem interpretá-lo de forma precisa.

2.4.2 ANÁLISE DE FONTES DE INFORMAÇÃO

As fontes de informação são repositórios de dados referentes a cada unidade da plataforma de governo (PACHECO, 2003). Esses repositórios, considerados por Pacheco (2003) como fontes de informação, não necessariamente devem estar dispostos de uma única forma ou estrutura física.

As bases de dados, sejam elas do tipo, relacional, orientada a objetos, estruturada ou até mesmo textual, representam informações de domínio do problema e também podem ser ricas fontes de informação para análise.

No caso de sistemas implementados ou ainda em desenvolvimento, porém com especificação, pode-se realizar a análise das interfaces desses sistemas objetivando a busca de informações relevantes para definição e especificação das unidades de informação.

Outra considerável fonte que representa informações de domínio e regras de negócios, utilizada no levantamento de requisitos e que pode ser considerada relevante para a análise, é a modelagem do sistema. Bruegge e Dutoit (2004) citam algumas atividades que se originam da modelagem do sistema como identificação dos atores, identificação dos cenários, identificação e refinamento dos casos de uso, identificação dos relacionamentos entre as classes e a identificação dos requisitos não funcionais, todos eles pertencentes à modelagem do sistema.

Além disso, até mesmo conversas entre usuários e desenvolvedores, consideradas extremamente relevantes no processo de levantamento de requisitos (ALVARES, 2002), podem fornecer informações úteis para análise junto ao processo de definição e especificação de unidades de informação.

Unhelkar (2002) e Papaioannou e Theodoulidis (1996) afirmam que essa interação pode ser estabelecida em forma de entrevistas e *workshops*, *walkthrough*, e outros tipos de reunião entre usuários, técnicos e gerentes de projeto, procurando abordar ações, problemas, expectativas e necessidades relacionadas ao meio de trabalho.

2.4.3 ESPECIFICAÇÃO DO DOMÍNIO DE PROBLEMA

Um fator primário que pode ser considerado fundamental para a definição de unidades de informação em arquiteturas de sistemas de governo eletrônico é a delimitação do domínio do problema. Esse fator consiste obviamente no entendimento do problema. Unhelkar (2002) afirma que para o desenvolvimento de aplicações de software é essencial a obtenção desse entendimento.

Com essa finalidade, podem ser consideradas atividades propostas por Unhelkar (2002), tais como documentação e entendimento dos requisitos, análise dos requisitos, investigação dos problemas em detalhes e, opcionalmente, criação de um protótipo conceitual que possibilite o entendimento do fluxo dos processos dentro do negócio. Na criação do protótipo, o domínio do problema é inteiramente focado no que está acontecendo com o negócio ou com o usuário. Tracz, Coglianesi e Young (1993) confirmam a importância e a relação da definição dos requisitos com o domínio do problema.

Outra relação semelhante a essa é a apresentada por Jones *et al.* (1998), que, entre outros métodos para desenvolvimento de ontologia, abordam no seu trabalho o projeto IDEF5. Este projeto mostra como etapa pertencente a ele um conjunto de critérios incluindo objetivos e requisitos, com a finalidade de definir os limites da ontologia e partes específicas dos sistemas que devem ser incluídas ou excluídas.

As atividades consideradas por Unhelkar (2002) para entendimento do domínio do problema e a afinidade destas com a atividade de requisitos ilustraram a importância da consideração dos requisitos, a fim de proporcionar a especificação do domínio do problema, seja para o desenvolvimento de aplicações de software e ontologias, seja para unidades de informação.

2.4.4 ONTOLOGIAS

O termo “ontologia” possui sua origem na filosofia e se refere à disciplina que trata do assunto da existência. No nosso contexto (o tecnológico), o significado desse termo é sutilmente diferenciado: **é uma descrição formal dos conceitos e relacionamentos que existem dentro de um domínio** (sendo assim, não é uma disciplina, e sim um artefato). Isso significa que uma ontologia se relaciona com um vocabulário específico e uma linguagem específica, diferentemente da disciplina filosófica que trata da existência, mas não da linguagem (DAUM; MERTEN, 2002).

Segundo Gruber (1993), uma ontologia é uma especificação explícita de uma conceitualização. Além disso, ela é essencialmente um acordo, e esse acordo não necessariamente precisa abranger toda a conceitualização de determinado domínio, mas pode abranger apenas parte dele, ou seja, ele pode oferecer uma visão para o domínio. Assim, uma ontologia atua como um contrato entre parceiros, permitindo que se comuniquem com segurança dentro do domínio de informação (DAUM; MERTEN, 2002).

A finalidade principal de uma ontologia é permitir a comunicação entre os sistemas computadorizados de maneira independente de tecnologias de sistema, de arquiteturas de informação e de domínios individuais de aplicação (ONTOLOGY, 2003). Além disso, uma ontologia visa possibilitar a descrição de domínios de interesse

agregando as relações, propriedades, funções, processos e ainda regras e restrições dos objetos pertencentes a esses domínios (DACONTA *et al.*, 2003).

Um dos resultados do trabalho de definição de ontologias é a produção de gramáticas e vocabulários comuns à comunidade usuária interessada no domínio correspondente (PACHECO; KERN, 2001), uniformizando referências e conseqüentemente possibilitando e facilitando o processo de descoberta e geração de conhecimento (PACHECO, 2003).

No âmbito de governo eletrônico, o desafio para os responsáveis será a construção de modelos ontológicos compreensíveis e dedicados aos diversos atores partícipes de uma plataforma de governo (PACHECO, 2003). Segundo Aldrich *et al.* (2002), para que se possa cumprir esses objetivos, é necessário que seus recursos atendam simultaneamente ao que os cidadãos esperam do governo e ao que ele necessita da sociedade à qual se destina.

Pacheco (2003) relata a relação de ontologias com a sua arquitetura de e-gov. Segundo ele, as ontologias deverão ser estabelecidas ao nível da primeira camada em sua arquitetura proposta por se refletirem ao longo de todos os demais instrumentos da plataforma de governo. Para os instrumentos de informação e bases de dados, componentes da segunda camada, o resultado deverá ser a padronização XML de cada unidade de informação da plataforma. As ontologias também são relevantes na camada de construção de portais (e.g., podem ser a base para a construção de portais colaborativos da plataforma – STAAB *apud* PACHECO, 2003). Na camada de sistemas de conhecimento, as ontologias são referências para a descoberta e geração de conhecimento (e.g. JONES *et al.*, 1998).

A seguir são apresentadas algumas metodologias e métodos para desenvolvimento e gerenciamento de ontologias:

- ✍ **TOVE**: baseada em experiências no desenvolvimento do Tove (*Toronto Virtual Enterprise*), trata para o desenvolvimento de ontologias considerações como cenários de motivação, questões de competência informal, terminologia de especificação, questões de competência formal, especificação de veracidade e normas de integridade

(GRÜNINGER; FOX, 1994a; GRÜNINGER; FOX, 1994b; GRÜNINGER; FOX, 1995).

- ✍ **Enterprise Model Approach:** descreve um esqueleto metodológico para construção de ontologias (USCHOLD, 1996a; USCHOLD, 1996b; USCHOLD; KING, 1995). Salienta as atividades de identificação do propósito, identificação do escopo, formalização e evolução formal.
- ✍ **METHONTOLOGY:** Inicialmente descrita por Gomez-Perez *et al.*, (1996) e atualizada por Fernandez *et al.* (1997), inicia pela identificação de uma série de atividades que são envolvidas no desenvolvimento de uma ontologia. Essa metodologia originou-se de um projeto na área de química.
- ✍ **KBSI IDEF5:** O método IDEF5 foi desenvolvido para assistir na criação, modificação e manutenção de ontologias. IDEF5 é um procedimento geral com um conjunto de linhas para organização e escopo, coleção de dados, análise de dados, desenvolvimento inicial da ontologia, e validação e refinamento da ontologia – KBSI (1994)

Outras metodologias para desenvolvimento de ontologias podem ser encontradas em Jones *et al.* (1998).

A seguir é apresentada uma contextualização da proposta deste trabalho relacionando-a aos conceitos citados sobre ontologias.

Conforme já mencionado, define-se neste trabalho uma metodologia própria e integrada a uma arquitetura de governo eletrônico para definição de unidades de informação. As unidades de informação definidas a partir desta metodologia:

- a) fornecerão uma especificação explícita do seu domínio ou parte (conforme Gruber (1993) define ontologia), agregando a essa especificação as relações, propriedades, funções, os processos e ainda as regras e restrições dos objetos pertencentes a esses domínios (DACONTA *et al.*, 2003);
- b) permitirão a comunicação entre aplicações, independentemente de tecnologias de sistemas, de arquiteturas de informação e de domínios individuais de aplicação (conforme Ontology (2003) define ontologia);

- c) e ainda possibilitarão o processo de descoberta e geração de conhecimento (conforme define Pacheco (2003)).

No entanto, vale ressaltar que, para que essas unidades se tornem ontologias, requere-se que elas sejam definidas em comum acordo no seu domínio (conforme Daum e Merten (2002) descrevem ontologia) e na sua comunidade.

2.4.5 PADRONIZAÇÃO DE INFORMAÇÕES

A padronização de informações é o *kernel* da proposta metodológica deste trabalho. Considera-se como padronização a atividade que irá estabelecer a criação do padrão para as unidades de informação.

Essa atividade é avaliada por Pacheco (2003) como fundamental em um projeto de e-gov. Acredita-se que todo trabalho despendido no processo de definição de informações, baseado na proposta metodológica deste trabalho, pode ter o seu sucesso comprometido se as unidades de informação definidas não estiverem padronizadas.

O padrão, resultado da atividade de padronização, conterá informações que descreverão o metadado definido para a unidade de informação a ser representada (TURNER, 2002). Nesse padrão deverão constar as informações obtidas no processo de levantamento de requisitos abordado anteriormente.

Relacionando a atividade de padronização de informação com a arquitetura referência desta dissertação, Pacheco (2003) mostra no seu trabalho que, para cada unidade de informação, deverão ser definidos metadados específicos, os quais devem ser descritos em uma linguagem de marcação para promover a interoperabilidade dos sistemas e informações da plataforma. Ceri *et al.* (2000) relatam a relevância da linguagem de marcação, mais especificamente a XML, ao considerá-la como chave para permitir o alcance da interoperabilidade de dados.

Em síntese, essa interoperabilidade, alcançada através da padronização especificada em uma linguagem de marcação, consiste na habilidade de transferir e usar a informação de maneira uniforme e eficiente através das organizações múltiplas e dos sistemas da tecnologia de informação (NOIE, 2004).

2.4.6 EXTENSIBLE MARKUP LANGUAGE (XML)

XML consiste em uma linguagem de marcação de dados com o propósito fundamental de descrever informações. Décio (2002) afirma que essa capacidade é extremamente importante para armazenamento, recuperação e transmissão de informações.

Através da XML é possível colocarmos em um mesmo lugar dados e metadados, ou seja, a descrição dos dados representados em XML (Décio, 2000).

O desenvolvimento da linguagem XML foi iniciado em 1996 pelo *World Wide Web Consortium* - W3C visando construir uma linguagem de marcação abrangente que combinasse a flexibilidade e a capacidade do *Standard Generalized Markup Language* (SGML) com a ampla aceitação do *HyperText Markup Language* (HTML) (ANDERSON *et al.*, 2001). Em 4 de fevereiro de 1998, o Consórcio W3C estabeleceu a primeira versão recomendada da linguagem XML (W3C, 2003).

Após a aprovação da XML pelo Consórcio W3C, surgiram tecnologias baseadas na XML com finalidades complementares, como as apresentadas a seguir (W3C, 2003):

- ✎ *eXtensible Stylesheet Language* – XSL para transformação de documentos XML;
- ✎ *XML Schema*¹ para padronização de documentos XML, que até então havia sido feita com o *Document Definition Type* - DTD originário do SGML;
- ✎ *XPath* para realização de consultas dentro dos documentos XML;
- ✎ *XQuery* para realizar consultas em bancos de dados XML;
- ✎ *XPointer* e *XLink* respectivamente para apontar documentos XML e criar *links* entre diferentes partes de um mesmo documento XML;
- ✎ *XML Signature* e *XML Encryption* muito utilizadas para segurança considerando-se que XML é um padrão aberto.

¹ As linguagens *XML Schema* e *DTD*, recursos para especificação dos metadados em XML, serão abordadas com mais detalhes no próximo capítulo.

Mais informações sobre essas e outras tecnologias XML não relatadas neste trabalho podem ser obtidas diretamente no *site* do Consórcio W3C (<http://www.w3.org>).

Uma das características principais das tecnologias XML consiste em ser um padrão aberto e extensível (BAX, 2001). Um documento XML pode ser facilmente aberto em um simples editor de textos (o que não impede de também serem utilizados editores próprios para XML). Além disso, a XML detém, conforme citado anteriormente, o propósito de descrever seus dados. Essa descrição é feita de tal forma que permite que um documento XML possa ser lido por seres humanos e inclusive por máquinas através de sistemas.

Outra característica interessante do XML é a sua independência de tecnologia e plataforma (BAX, 2001). Uma das razões para isso pode ser o fato de ela ser desenvolvida e mantida por um consórcio mundial de empresas e instituições, o W3C. Isso representa que uma unidade de informação especificada em XML não está restrita a nenhuma plataforma e nem mesmo tecnologia proprietária de alguma empresa.

Essas e outras características podem ter sido o motivo que levaram a XML a se popularizar tanto em seu segmento. Segundo Khare e Rifkin (1998), os últimos anos foram caracterizados por uma evolução crescente de padronização para interoperabilidade na *Web*, com a XML sendo o padrão evolutivo de maior aceitação.

Essa aceitação da linguagem XML como padrão para padronização de informação também ocorreu no âmbito do governo eletrônico. A seguir relatamos alguns projetos governamentais entre os muitos existentes que utilizam como recurso a linguagem XML.

Segundo o relatório anual da Accenture (ACCENTURE, 2003c), a Dinamarca tem utilizado tal linguagem para representar suas informações em documentos XML e armazená-los em um banco nacional para metadados em XML, conhecido como “*XML Infostructurebase*” (isb.oio.dk). Eles esperam que essa iniciativa ajude a acelerar a integração dos sistemas e serviços do seu governo.

O governo britânico também tem investido na idéia de padronização de informações em plataformas de governo eletrônico utilizando a linguagem XML. Ele tem promovido diretrizes e políticas gerais para o setor público criar padrões e

metadados em XML para seus projetos e sistemas de informação (CABBINET OFFICE, UK GOVERNMENT, 2001).

Outro exemplo já citado na seção que abordou arquiteturas de governo eletrônico é a arquitetura para e-gov do governo americano publicada pela FEA (2002), que utiliza a linguagem XML na sua arquitetura de dados.

Completando os exemplos citados de projetos de e-gov que relatam a utilização da linguagem XML, destacamos o trabalho da *Comunidade para Ontologias em Ciência, Tecnologia e Informações de Aperfeiçoamento de Nível Superior – CONSCIENTIAS*, que estabeleceu padrões em XML para as unidades de informação da Plataforma Lattes (CONSCIENTIAS, 2004) e a arquitetura referencial desta dissertação, proposta por Pacheco (2003) e utilizada no estudo de caso da Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

3 METODOLOGIA PARA DEFINIÇÃO DE UNIDADES DE INFORMAÇÃO EM PLATAFORMAS DE E-GOV

3.1 APRESENTAÇÃO

Neste capítulo é apresentada a proposta metodológica deste trabalho para definição de unidades de informação em plataformas de sistemas de e-gov.

Essa metodologia tem aproveitamento direto na arquitetura proposta por Pacheco (2003), no componente metodológico “Padronização XML e Unidade de Informação”. O enquadramento da metodologia deste trabalho na arquitetura de sistemas de e-gov proposta por Pacheco (2003) pode ser melhor visualizada na Figura 7.

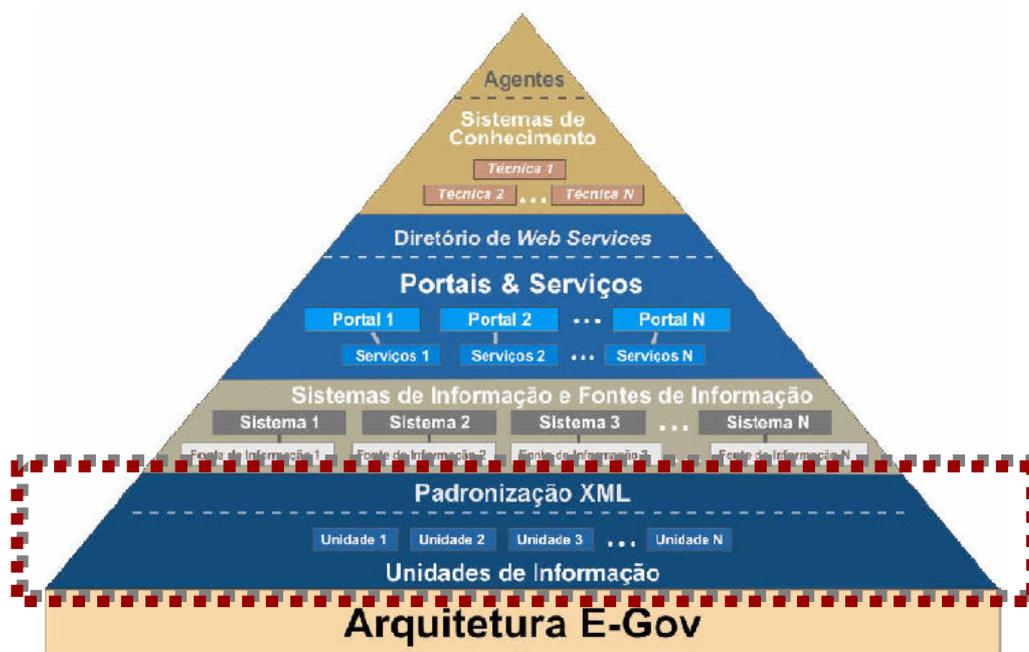


Figura 7 – Enquadramento da metodologia deste trabalho na arquitetura de sistemas de e-gov proposta por Pacheco (2003).

Serão consideradas as seguintes etapas na metodologia deste trabalho para definição das unidades de informação: a) identificação, b) especificação, c) padronização e d) gestão. A metodologia proposta para a definição de unidades de informação em plataformas de sistemas de governo eletrônico, bem como suas etapas e a ordem entre elas, pode ser mais bem visualizada na Figura 8.



Figura 8 - Metodologia proposta para a definição de unidades de informação em plataformas de sistemas de governo eletrônico

Uma descrição das etapas pertencentes à metodologia para definição e gestão de unidades de informação em plataformas de governo eletrônico é abordada nas seções a seguir. Nessas seções procura-se apresentar diretrizes, indicações, cuidados e recomendações metodológicas referentes a esse desenvolvimento, resultado do estudo realizado no capítulo anterior.

3.2 IDENTIFICAÇÃO DE UNIDADE DE INFORMAÇÃO

A etapa de identificação de uma unidade de informação surge como alicerce para as demais etapas, cujo sucesso está estritamente ligado a essa etapa, também denominada de etapa-base. A Figura 8, mostrada anteriormente, permite visualizar a representação disso, quando ilustra a sobreposição das demais etapas sobre a etapa de identificação.

O resultado da aplicação dessa fase no processo de padronização de unidades de informação consistirá na posse das informações necessárias para que se possa realizar o trabalho da fase seguinte, que é o de especificação das informações alcançadas nessa fase.

Para isso, previamente à identificação de quais são as considerações importantes da metodologia deste trabalho no processo de **identificação de uma unidade de informação**, sugere-se, com base na indicação de Turner (2002), que antes mesmo de se criar um novo modelo de informação sejam averiguados os modelos já existentes para tal finalidade. Nessa averiguação, podem ser considerados inclusive aqueles que já tenham sido especificados e padronizados na linguagem de marcação XML, de modo que não se construa algo que já tenha sido criado com as mesmas necessidades da unidade a ser definida. Segundo Daum e Merten (2002, p. 162), “a introdução da XML, em particular, ocasionou atividades espalhadas de definição de ontologias compartilhadas, o que resultou em uma rica coleção de DTDs XML¹ de setores específicos”.

Caso seja encontrado um ou mais modelos que não atendam a todas as necessidades específicas esperadas para tal unidade ser definida, verifique se não é possível aproveitar algo ou parte do que já foi definido nesses modelos. O aproveitamento de tal parte de um modelo de informação já definido surge como um dos recursos a serem considerados para a etapa de identificação na **metodologia** proposta neste trabalho.

Constatada a necessidade de construção de tal unidade de informação, através da averiguação da inexistência de um modelo que esteja de acordo com o propósito estabelecido para ela, inicia-se o processo de desenvolvimento de um novo modelo.

Com base no estudo realizado no capítulo anterior, sugere-se a realização de algumas tarefas, entre elas algumas oriundas da atividade de levantamento de requisitos pertencente à área de engenharia de requisitos. Essas tarefas deverão levantar informações que descrevam a unidade de informação para a realização das demais atividades estabelecidas na metodologia deste trabalho para definição de unidades de informação.

A seguir, são apresentadas tarefas sugestivas para o levantamento das informações:

¹ Considera-se como “DTDs XML” modelos de informação ou unidades de informação especificadas e padronizadas na linguagem de marcação XML. Serão abordadas mais informações sobre a Document Type Definition na etapa de especificação desta metodologia.

- ✍ análise de bases de dados, sejam elas do tipo, relacional, orientada a objetos, estruturada ou até mesmo textual, que representam informações de domínio do problema;
- ✍ análise das interfaces, no caso de sistemas implementados ou ainda em desenvolvimento, porém com especificação;
- ✍ análise da modelagem do sistema considerando principalmente identificação dos atores, identificação dos cenários, identificação e refinamento dos casos de uso, identificação dos relacionamentos entre classes e identificação dos requisitos não funcionais;
- ✍ entrevistas aplicadas aos usuários e outros tipos de reunião entre desenvolvedores e usuários, tais como *workshops*, *walkthrough* etc., que possibilitem uma comunicação eficaz com os usuários.

A realização dessas atividades possibilitará o levantamento das informações necessárias para a definição da unidade de informação. No entanto, outras atividades não citadas neste trabalho, mas que porventura permitam a extração de informações consideradas pelos seus desenvolvedores relevantes para a definição de uma unidade de informação, podem também ser úteis nesse processo.

Uma vez identificadas as informações de uma unidade, sugere-se na metodologia proposta a realização da etapa de especificação da unidade de informação, que será abordada na seção a seguir.

3.3 ESPECIFICAÇÃO DE UNIDADE DE INFORMAÇÃO

O processo de especificação de uma unidade de informação consiste basicamente na definição do conteúdo de uma unidade de informação e na representação e descrição desse conteúdo definido.

Para a concretização desse trabalho, essa seção foi dividida em duas subseções: “*Definição do conteúdo de uma unidade de informação*”, que apresenta informações em forma de instruções que o conduzirá no processo de definição do conteúdo de uma UI, e “*Representação da definição do conteúdo na forma de um diagrama de classes*”,

que o conduzirá para a realização do trabalho de representação e descrição do conteúdo definido na subseção anterior. Ambas as seções são apresentadas a seguir.

3.3.1 DEFINIÇÃO DO CONTEÚDO DE UMA UNIDADE DE INFORMAÇÃO

A definição do conteúdo de uma unidade de informação consiste essencialmente na identificação e descrição dos objetos encontrados no domínio de problema da unidade de informação em desenvolvimento.

Para a efetivação deste trabalho, sugere-se que ele seja realizado em duas fases, que serão descritas a seguir.

a) Fase de levantamento dos candidatos a objetos

A identificação dos candidatos a objetos para definição do conteúdo de uma unidade de informação foi concebida na metodologia desta dissertação com base no trabalho de Anderson *et al.* (2001). Segundo ele, uma boa maneira de se iniciar qualquer exercício de modelagem de informação é estabelecer os nomes das *coisas* (identificadas nesse trabalho também como candidatos a objetos) no sistema ou no domínio de problema.

Para a identificação desses candidatos a objetos, sugere-se que o trabalho seja iniciado fazendo-se uma lista das coisas relevantes para o domínio de problema da unidade de informação. Nem todas essas “coisas”, ou candidatos a objetos, tornar-se-ão objetos válidos para a unidade de informação após a fase seguinte de validação. Porém, dependendo do levantamento realizado, grande parte das “coisas” levantadas tende a se tornar objetos válidos.

No caso de uma unidade de informação em ciência e tecnologia que representará um grupo de pesquisa, podem ser identificados inicialmente nesse domínio de problema os candidatos a objetos listados na Tabela 2, a seguir.

Nome dos candidatos a objetos:	
1. Nome	2. Endereço institucional
3. E-mail para contato	4. Sítio <i>WEB</i>
5. Líder	6. CPF do Líder

7. Pesquisador	8. Estudante
9. Técnico	10. Linha de pesquisa
11. Empresa	12. Palavra-chave
13. Setor de aplicação	14. Repercussão

Tabela 2 – Lista inicial dos candidatos a objetos para a unidade de informação de grupo de pesquisa.

Depois de construída a lista inicial com os candidatos a objetos relevantes para o domínio de problema da unidade de informação, proceda à fase seguinte, que o auxiliará a validar os candidatos a objetos levantados.

b) Fase de validação dos candidatos a objetos

Antes de se partir para a definição das classes e os relacionamentos para esses objetos, é necessário primeiramente certificar-se de que esses objetos encontrados são válidos. Esta fase de validação dos candidatos a objetos visa auxiliá-lo nesse trabalho.

A partir da lista dos candidatos a objetos, criada na fase anterior, descreva as responsabilidades e os colaboradores desses candidatos. A forma proposta nesta dissertação para descrição dos objetos foi criada com base na técnica de Beck e Cunningham (1989), denominada Cartões CRC – Cartões Classe-Responsabilidade-Colaboração (*Class-Responsibility-Collaboration*). Essa técnica prevê que para cada objeto podem ser descritas informações tais como seu nome, suas responsabilidades e suas colaborações.

No entanto, o modelo proposto neste trabalho, criado com base no CRC, não visa deixá-lo limitado no que diz respeito à descrição de uma informação que vá além do que foi estabelecido para os cartões CRC, como responsabilidade e colaboração.

Caso note que ainda há informações complementares interessantes para descrição do candidato a objeto em análise, como os possíveis valores que estarão representados nesse futuro objeto, sintá-se na liberdade para descrevê-las em uma nova coluna, que poderia ser inserida após as colunas Responsabilidade e Colaboração que estão no modelo original do Cartão CRC. A partir dessa nova coluna, que pode ser denominada “Outras informações”, atributos inicialmente identificados nos candidatos a objetos podem ser descritos nos seus respectivos cartões. Apesar da metodologia deste

trabalho prever mais adiante uma fase que tratará com mais detalhes da identificação dos atributos em objetos, essa atividade pode também auxiliar na validação dos objetos.

A representação gráfica para o Cartão CRC adaptado com a nova coluna “Outras informações” pode ser visualizada na Figura 9.

Nome do objeto: _____		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>

Figura 9 – Modelo do cartão para descrição dos candidatos a objetos na fase de validação

Todas essas informações fornecidas sobre o candidato a objeto irão ajudá-lo a resolver problemas de ambigüidade e, ainda, possibilitarão que ele certifique-se de que os nomes escolhidos para esses objetos podem ser reconhecidos e interpretados corretamente.

Durante a realização deste trabalho algumas mudanças entre os candidatos a objetos podem ocorrer, tais como:

- ✍ necessitar alterar o nome dos candidatos levantados, a fim de resolver problemas de ambigüidade e também fornecer uma descrição clara dos objetos abordados ou até mesmo por nomenclatura incorreta.
- ✍ constatar que alguns candidatos se definem melhor como atributos de um objeto; e
- ✍ encontrar novos objetos. Isso é perfeitamente normal. Estamos na fase de validação dos objetos, e o detalhamento do objeto pode gerar novos objetos a partir dele.

No exemplo abordado para grupo de pesquisa após descrição dos candidatos a objetos obtiveram-se a validação dos seguintes candidatos que, a partir desta fase, tornam-se objetos pertencentes à unidade de informação em ciência e tecnologia que representará um grupo de pesquisa:

- a) **IDENTIFICAÇÃO DO GRUPO:** o cartão para esse candidato a objeto validado pode ser visualizado na Figura 10 apresentada a seguir.

Nome do objeto: IDENTIFICAÇÃO DO GRUPO		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar informações gerais de identificação sobre o Grupo de pesquisa		A representação deverá conter: Nome do grupo, Ano de criação, E-mail para contato, Sítio <i>WEB</i> , Instituição
Representar informações sobre os Líderes do grupo	Objeto “Líder”	
Representar informações sobre o Endereço institucional do grupo	Objeto “Endereço Institucional do grupo”	
Representar informações sobre as Repercussões do grupo	Objeto “Repercussão”	

Figura 10 – Cartão de identificação para o objeto “Identificação do grupo”

- b) **ENDEREÇO INSTITUCIONAL DO GRUPO:** o cartão para esse candidato a objeto validado pode ser visualizado na Figura 11 apresentada a seguir.

Nome do objeto: ENDEREÇO INSTITUCIONAL DO GRUPO		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar informações sobre o Endereço institucional físico do grupo		A representação deverá conter: Logradouro, Bairro, UF, Cidade, CEP, telefone
Representar informações sobre o Endereço institucional eletrônico do grupo		A representação deverá conter: E-mail, Sítio <i>WEB</i>

Figura 11 – Cartão de identificação para o objeto “ENDEREÇO INSTITUCIONAL DO GRUPO”

- c) **LÍDER:** o cartão para esse candidato a objeto validado pode ser visualizado na Figura 12 apresentada a seguir.

Nome do objeto: LÍDER		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar o Primeiro líder do grupo	Objeto “Pesquisador”	
Representar o Segundo líder do grupo	Objeto “Pesquisador”	

Figura 12 – Cartão de identificação para o objeto “LÍDER”

- d) **REPERCUSSÃO**: o cartão para esse candidato a objeto validado pode ser visualizado na Figura 13 apresentada a seguir.

Nome do objeto: REPERCUSSÃO		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar informações sobre as Repercussões do trabalho do grupo junto à comunidade científica ou ao seu público-alvo		A informação deve ser representada em um único campo atributo no formato texto.

Figura 13 – Cartão de identificação para o objeto “REPERCUSSÃO”

- e) **PESQUISADOR**: o cartão para esse candidato a objeto validado pode ser visualizado na Figura 14 apresentada a seguir.

Nome do objeto: PESQUISADOR		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar informações sobre os Pesquisadores do grupo		A representação deverá conter: Nome, Nacionalidade, País de nascimento, Data de nascimento, CPF.

Figura 14 – Cartão de identificação para o objeto “PESQUISADOR”

- f) **ESTUDANTE**: o cartão para esse candidato a objeto validado pode ser visualizado na Figura 15 apresentada a seguir.

Nome do objeto: ESTUDANTES		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar informações sobre os Estudantes do grupo		A representação deverá conter: Nome, Nacionalidade, País de nascimento, Data de nascimento, CPF.

Figura 15 – Cartão de identificação para o objeto “ESTUDANTE”

- g) **TÉCNICO**: o cartão para esse candidato a objeto validado pode ser visualizado na Figura 16 apresentada a seguir.

Nome do objeto: TÉCNICO		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar informações sobre os Técnicos do grupo		A representação deverá conter: Nome, atividade exercida no grupo e grau de escolaridade.

Figura 16 – Cartão de identificação para o objeto “TÉCNICO”

- h) **LINHA DE PESQUISA:** o cartão para esse candidato a objeto validado pode ser visualizado na Figura 17 apresentada a seguir.

Nome do objeto: LINHA DE PESQUISA		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar informações básicas sobre as Linhas de pesquisa		A representação deverá conter: Nome e Objetivo da linha
Representar informações sobre as Palavras-chave nas linhas de pesquisa	Objeto “Palavra-chave”	
Representar informações sobre os Pesquisadores nas linhas de pesquisa	Objeto “Linha de pesquisa”	
Representar informações sobre os Estudantes nas linhas de pesquisa	Objeto “Estudante”	
Representar informações sobre as Áreas do conhecimento nas linhas de pesquisa	Objeto “Área do conhecimento”	
Representar informações sobre os Setores de aplicação nas linhas de pesquisa	Objeto “Setor de Aplicação”	

Figura 17 – Cartão de identificação para o objeto “LINHA DE PESQUISA”

- i) **EMPRESA:** o cartão para esse candidato a objeto validado pode ser visualizado na Figura 18 apresentada a seguir.

Nome do objeto: EMPRESA		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar informações básicas sobre as Empresas		A representação deverá conter: Nome, Sigla, UF, Cidade, CNPJ, Segmento econômico e Natureza jurídica.
Representar informações sobre os Setores de atividade econômica da empresa	Objeto “Setor de atividade econômica”	
Representar informações sobre os Tipos de relação da empresa	Objeto “Tipo de relação”	
Representar informações sobre os Tipos de remuneração	Objeto “Tipo de remuneração”	

Figura 18 – Cartão de identificação para o objeto “EMPRESA”

- j) **PALAVRA-CHAVE:** o cartão para esse candidato a objeto validado pode ser visualizado na Figura 19 apresentada a seguir.

Nome do objeto: PALAVRA-CHAVE		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar as palavras-chave do objeto Linhas de pesquisa		A representação deverá conter: Nome.

Figura 19 – Cartão de identificação para o objeto “PALAVRA-CHAVE”

- k) **SETOR DE APLICAÇÃO:** o cartão para esse candidato a objeto validado pode ser visualizado na Figura 20 apresentada a seguir.

Nome do objeto: SETOR DE APLICAÇÃO		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar os setores de aplicação das Linhas de pesquisa		A representação deverá conter: Nome do setor e Nome do subsetor.

Figura 20 – Cartão de identificação para o objeto “SETOR DE APLICAÇÃO”

- l) **SETOR DE ATIVIDADE ECONÔMICA:** o cartão para esse candidato a objeto validado pode ser visualizado na Figura 21 apresentada a seguir.

Nome do objeto: SETOR DE ATIVIDADE ECONÔMICA		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar informações sobre os setores de atividade econômica das empresas		A representação deverá conter: Grande atividade, Divisão e Grupo.

Figura 21 – Cartão de identificação para o objeto “SETOR DE ATIVIDADE ECONÔMICA”

- m) **TIPO DE RELAÇÃO:** o cartão para esse candidato a objeto validado pode ser visualizado na Figura 22 apresentada a seguir.

Nome do objeto: TIPO DE RELAÇÃO		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar informações sobre os Tipos de relação das empresas		A representação deverá conter: Descrição. Os valores são preestabelecidos com base na definição da agência de fomento responsável (CNPq)

Figura 22 – Cartão de identificação para o objeto “TIPO DE RELAÇÃO”

- n) **TIPO DE REMUNERAÇÃO**: o cartão para esse candidato a objeto validado pode ser visualizado na Figura 23 apresentada a seguir.

Nome do objeto: TIPO DE REMUNERAÇÃO		
<i>Responsabilidade</i>	<i>Colaboração</i>	<i>Outras informações</i>
Representar informações sobre os Tipos de remuneração das empresas		A representação deverá conter: Descrição. Os valores são preestabelecidos com base na definição da agência de fomento responsável (CNPq).

Figura 23 – Cartão de identificação para o objeto “TIPO DE REMUNERAÇÃO”

Como se pode notar nesses cartões, diversas modificações ocorreram na lista inicial dos candidatos a objetos. A necessidade dessas modificações surgiu a partir do trabalho realizado com os cartões que foram amplamente exemplificados nesta dissertação.

Entre as modificações ocorridas pode-se notar que a primeira concerniu na criação de um novo objeto chamado “**Identificação do Grupo**”, a fim de proporcionar a representação dos candidatos “Nome”, “E-mail para contato” e “Sítio WEB”, que, após descrição, demonstraram ser responsabilidade de um objeto que tivesse um nome que proporcionasse uma representação mais ampla, como o que foi dado a ele (Identificação do grupo).

Além disso, a especificação do cartão para o objeto “Empresa” permitiu a identificação de novos objetos como “Setor de atividade econômica”, “Tipo de relação” e “Tipo de remuneração”.

Ao concluir essa fase, alcançou-se o número de 14 objetos validados entre os 14 identificados na etapa de levantamento dos candidatos a objetos. Resumindo, conforme já citado, alguns desses candidatos acabaram tornando-se responsabilidade de outros objetos e deixaram de existir. Também houve casos de novos objetos encontrados.

Além da validação dos candidatos a objetos, essa fase possibilitou a especificação de mais informações sobre os objetos, que auxiliarão o trabalho seguinte de criação de um diagrama de classes para representação e descrição desses objetos validados.

3.3.2 REPRESENTAÇÃO DA DEFINIÇÃO DO CONTEÚDO EM UM MODELO DESCRITIVO (NA FORMA DE UM DIAGRAMA DE CLASSES)

Uma vez identificado o conteúdo de uma unidade de informação, fato consolidado através da validação dos objetos pertencentes a ela, surge a necessidade de representar esse conteúdo em um modelo descritivo para a unidade de informação em desenvolvimento.

Essa representação, denominada na metodologia deste trabalho como especificação, é concretizada com base na modelagem de classes através da *Unified Model Language (UML)* e nos conceitos de orientação a objetos. A abordagem dos conceitos de modelagem através do diagrama de classes da UML e orientação a objetos transcende o escopo deste trabalho. No entanto, serão fornecidas neste trabalho informações, incluindo exemplificação através do estudo de caso para a unidade de informação “Grupo de Pesquisa”, que possibilitarão a representação de um modelo de informação através de um diagrama de classes.

Segundo Carlson (2002), uma classe da UML é um contêiner de recursos estruturais e comportamentais, porém apenas os recursos estruturais são relevantes durante a definição de um vocabulário XML. Esses recursos estruturais consistem em atributos, referências de associação, agregações e composições.

Como o foco deste trabalho está na construção de um modelo de informação, ou vocabulário XML (termo utilizado para modelos de informação representados através da linguagem XML), não será abordado nesta metodologia o estudo do comportamento de classes.

Tratando-se dos recursos estruturais abordados pela UML, Carlson (2002) afirma que as associações, as agregações e as composições não fazem parte da definição de classe em si, mas são definidas como objetos de associação independentes na UML.

Para a criação das classes para os objetos validados na etapa anterior deste trabalho serão realizadas as seguintes fases: a) organização dos objetos em uma hierarquia de classes; b) definição dos relacionamentos, cardinalidade e restrições; e c)

adicionamento de propriedades para indicar os detalhes dos valores associados com os objetos.

Durante essas três fases serão apresentados os conceitos para realização dos trabalhos e ainda o estudo de caso da unidade de grupo de pesquisa para exemplificação dos conceitos mencionados. São elas:

a) Fase de organização dos objetos em uma hierarquia de classes

Esta fase consiste essencialmente em criar tipos de dados para os objetos descritos e validados através do cartão CRC. Os tipos de dados que possuem subtipos serão representados, juntamente com seus subtipos, através de um diagrama de classes da UML. Esse diagrama conterá as classes organizadas de forma hierárquica para os tipos e subtipos.

Essa hierarquia será formalizada através da aplicação do conceito de herança. Segundo Ehrich (1997), a herança descreve como uma classe reaproveita características de uma outra. Carlson (2002) afirma que o uso da herança é uma característica fundamental na modelagem orientada a objeto, sendo essencial para a representação de conceitos abstratos e para a classificação dos termos em um vocabulário.

Mais especificamente, a utilização da herança consiste na classificação de superclasses para os tipos de dados e de subclasses para os subtipos. O relacionamento entre essas duas representações para as classes fica por conta de uma seta que consiste em uma linha simples que liga uma classe à outra, contendo uma ponta triangular cheia que aponta para a superclasse, conforme Figura 24 apresentada a seguir.

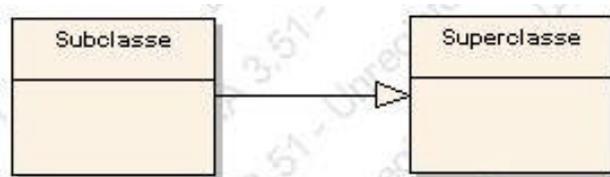


Figura 24 – Exemplo de representação do conceito de herança em um diagrama de classes da UML

Nesse momento ainda não é objetivo se preocupar com detalhamento dos atributos dessas classes. Esse trabalho será realizado mais adiante, em que as superclasses possuirão informações mais gerais, e as subclasses possuirão informações

mais específicas. Segundo Silva (2001), a herança possibilita o compartilhamento pelas classes especializadas (subclasses) dos atributos e operações de classes mais gerais (superclasse).

Para realizar o trabalho de organização esperado nessa fase, inicie analisando os objetos e verifique se há a ocorrência de objetos que possam ser tipos para outros objetos. Exemplificando: ao analisar os objetos identificados para uma unidade de informação para a área de recursos humanos, pode-se encontrar dois objetos que são “pessoa” e “funcionário”, que possivelmente teriam uma relação de herança.

Visando averiguar essa relação, é sugerida neste trabalho a utilização da pergunta: “é um?” ou “é um tipo de?”, proposta por Anderson *et al.* (2001) para organização de objetos dentro de uma hierarquia de classes.

No exemplo dado a pergunta “funcionário é uma pessoa?” permite certificar que “funcionário” pode ser um subtipo de “pessoa”. Caso isso se concretize, a classe para representar o tipo de dados “Pessoa” será uma superclasse, e a classe para representar o subtipo “Funcionario” será uma subclasse. A representação em forma de diagrama de classes da UML para esse exemplo pode ser visualizada na Figura 24 a seguir.

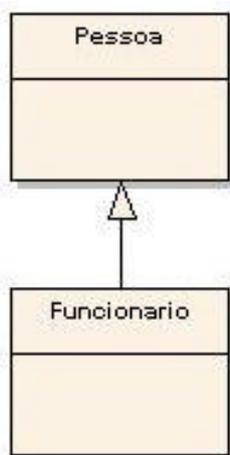


Figura 24 – Exemplo da organização de tipos e subtipos de dados em um diagrama de classes da UML

No entanto, para uma utilização correta do conceito de herança, essas classes generalizadas e especializadas precisam deter o mesmo resultado semântico entre elas.

Aplicando essas orientações ao estudo de caso para a unidade de informação de grupo de pesquisa, pode-se identificar que os objetos “Pesquisador”, “Estudante”, “Técnico” e “Líder” podem possuir informações semelhantes e que uma superclasse poderia ser criada visando dispor essas informações mais gerais. Atentando para o resultado semântico dessas relações, a sugestão ficaria na criação de uma superclasse chamada “Pessoa”.

Com isso, visando certificar essas relações, pode-se formular as seguintes questões para as subclasses e superclasse:

- ✍ “Técnico é uma Pessoa?”;
- ✍ “Estudante é uma Pessoa?”;
- ✍ “Pesquisador é uma Pessoa?”;
- ✍ “Líder é uma Pessoa?”;

Essas quatro perguntas permitem validar a herança utilizada. Dessa forma, as classes “Técnico”, “Estudante”, “Pesquisador” e “Líder” podem ser organizadas no diagrama como subclasses, e a classe “Pessoa” pode ser classificada como uma superclasse.

Ao se tentar incluir uma outra classe como “Empresa” nessa organização, o resultado não seria o mesmo. Se por algum motivo houvesse dúvida se é correto criar uma relação também da classe “Pessoa” com a classe “Empresa”, a pergunta “é um” ou “é um tipo de” agora permite rapidamente entender que o resultado semântico entre essas duas classes difere, pois “empresa não é um tipo de pessoa¹”, e que essa relação de herança não é correta.

Essas classes organizadas no diagrama de classes da UML podem ser visualizadas na Figura 25 a seguir:

¹ Nesse contexto, visando simplificar o modelo, pessoa está sendo tratado especificamente como pessoa física.

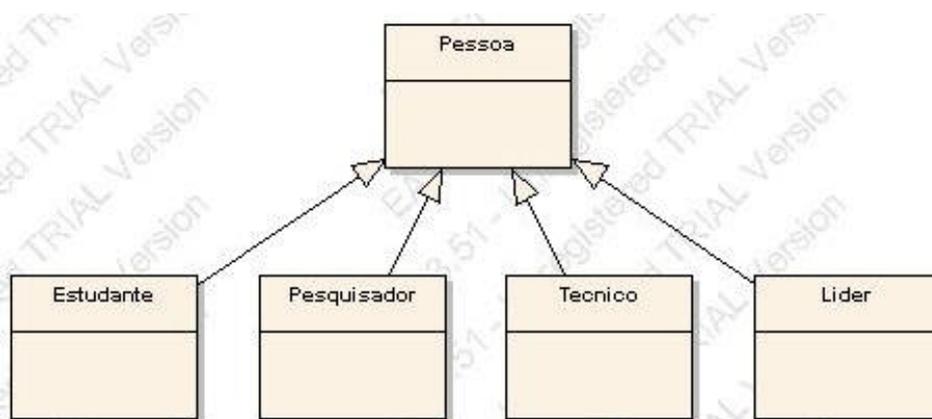


Figura 25 – Diagrama de classe que contém os objetos da unidade de informação de “Grupo de Pesquisa” organizados em uma hierarquia de classes

Com isso, a aplicação do conceito de herança no estudo de caso para a unidade de informação de “Grupo de Pesquisa” se restringe inicialmente apenas às classes que representarão os objetos “Técnico”, “Estudante”, “Pesquisador” e “Líder”.

Em relação às demais classes, como já mencionado, a princípio não foi possível e necessário aplicar o conceito de herança sobre elas, motivo pelo qual terão especificações individuais. As classes definidas até o momento são: “Pessoa”, “Tecnico”, “Estudante”, “Pesquisador”, “Líder”, “IdentificaçãoDoGrupo”, “LinhaDePesquisa”, “Empresa”, “EnderecoInstitucionalDoGrupo”, “Repercussao”, “PalavraChave”, “SetorDeAplicação”, “SetorDeAtividadeEconômica”, “TipoDeRelação” e “TipoDeRemuneração”.

A relação entre todas as classes será discutida na fase a seguir. Ainda será vista a criação da classe principal, que será necessária para agrupar e representar os relacionamentos entre as classes já criadas.

b) Fase de definição dos relacionamentos, cardinalidade e restrições

Para definir os relacionamentos, a cardinalidade e as restrições entre as classes, sugere-se que primeiramente seja identificada a classe principal da unidade de informação. Geralmente essa classe leva o nome da própria unidade de informação.

A partir desse procedimento, identifique as classes que possuem uma relação direta com essa classe principal.

No exemplo para a unidade de informação de grupo de pesquisa, criou-se a classe principal e com base na orientação proposta neste trabalho nomeou-se ela como “GrupoDePesquisa”. A partir dessa classe principal, identificou-se que as classes “IdentificacaoDoGrupo”, “Pesquisador”, “Estudante”, “Tecnico”, “LinhaDePesquisa” e “Empresa” possuíam uma relação direta com a principal.

Para a modelagem desse relacionamento é necessário também identificar o tipo de relacionamento e a cardinalidade entre as classes mapeadas. Os possíveis tipos de relacionamento a serem encontrados são herança (já visto anteriormente), associação, composição e agregação.

As composições são sempre mostradas nos diagramas UML usando-se um losango preenchido no final do relacionamento, conforme ilustrado na Figura 26, indicando que os objetos relacionados estão sob propriedade não apenas em referência, mas em valor (CARLSON, 2002). Já as agregações são apresentadas nos diagramas UML através de um losango sem preenchimento, conforme indicado na Figura 26, e também são utilizadas para indicar que os objetos relacionados estão sob propriedade em forma de valor.

Ao se representar um modelo UML em XML pode ser visto que, de modo semelhante, um elemento XML possui seus elementos-filhos por valor. No caso de relacionamento do tipo composição, se o proprietário é excluído, então os objetos (ou elementos no caso de representação em XML) que estão sob propriedade também são excluídos. O mesmo não ocorre quando o relacionamento é do tipo agregação, pois nesse caso, a criação do objeto da classe agregada pode ocorrer depois de o objeto da classe superior ser criado, e ainda a destruição pode ocorrer antes do objeto proprietário ser destruído (FOWLER; SCOTT, 2000).

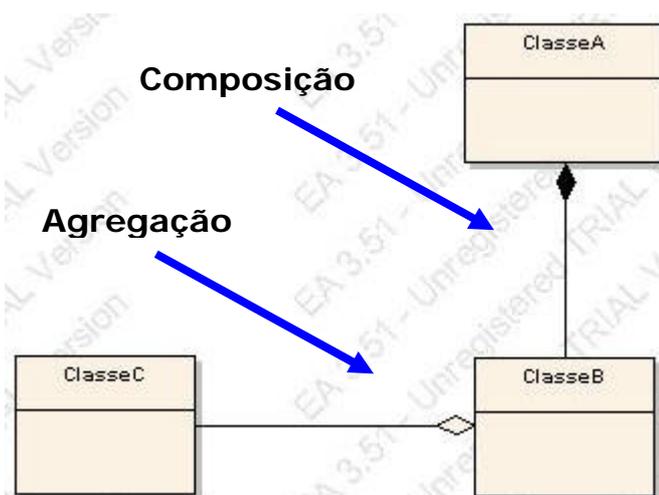


Figura 26 – Exemplo de relação do tipo composição e agregação

Já a associação, não tão comum para a modelagem de informações estruturais, surge quando um objeto usa um outro, mas não o possui. O vínculo entre esses dois objetos é apenas uma referência de um para o outro, conforme ilustrado na Figura 27. Nesse caso, quando um dos objetos é destruído, o outro que possui o relacionamento não é destruído conseqüentemente (CARLSON, 2002).



Figura 27 – Relacionamento de associação com cardinalidade de 1 para muitos, entre a classe “ClasseA” e a classe “ClasseB”

A modelagem dos relacionamentos será iniciada utilizando composição e agregação, e posteriormente será abordado o relacionamento do tipo associação. Ainda são apresentados exemplos para esses relacionamentos procurando utilizar como referência a unidade de informação de Grupo de Pesquisa.

Sugere-se que o trabalho de identificação dos relacionamentos e cardinalidade seja iniciado também através da utilização de sentenças em português que podem ser formuladas utilizando a palavra “possui”. Questões como essas, a seguir, podem ajudar na identificação:

- ✍ Um grupo de pesquisa pode possuir nenhuma, uma ou mais identificações?

- ✍ Um grupo de pesquisa pode possuir nenhum, um ou mais pesquisadores?
- ✍ Um grupo de pesquisa pode possuir nenhum, um ou mais estudantes?
- ✍ Um grupo de pesquisa pode possuir nenhum, um ou mais técnicos?
- ✍ Um grupo de pesquisa pode possuir nenhuma, uma ou mais linhas de pesquisa?
- ✍ Um grupo de pesquisa pode possuir nenhuma, uma ou mais empresas?

Devido ao fato de todos os objetos especificados nas classes que têm relação com a classe principal serem de propriedade dela, o tipo de relacionamento mais apropriado nesses casos é o de composição.

A utilização do relacionamento do tipo agregação não é muito comum para especificação de modelos de informação (CARLSON, 2002). Devido à unidade de informação de Grupo de Pesquisa não possuir esse tipo de relação e visando fornecer um exemplo mais claro também para esse relacionamento, o exemplo para ele será apresentado na etapa seguinte juntamente com o seu respectivo código XML.

Com base nas informações apresentadas a respeito dos relacionamentos, pode-se chegar às conclusões especificadas no diagrama de classes apresentado na Figura 26 a seguir.

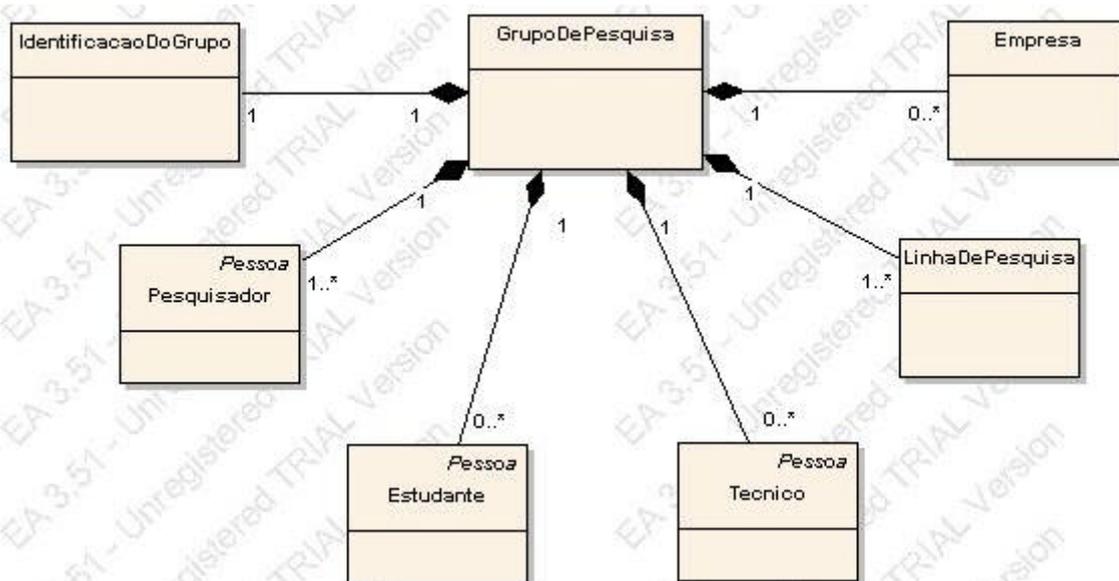


Figura 28 – Definição e apresentação dos relacionamentos de composição da classe principal da unidade de informação de Grupo de Pesquisa.

Depois de especificados os relacionamentos das classes que têm relação direta com a classe principal, denominadas neste trabalho de classes de primeiro nível, especifique, utilizando as mesmas orientações, os relacionamentos das classes que possuem relação com cada uma das classes de primeiro nível já apontadas.

A seguir, na Figura 29, são apresentadas as classes (classificadas como de segundo nível) que possuem relação com a classe “IdentificacaoDoGrupo” (classificada como de primeiro nível).

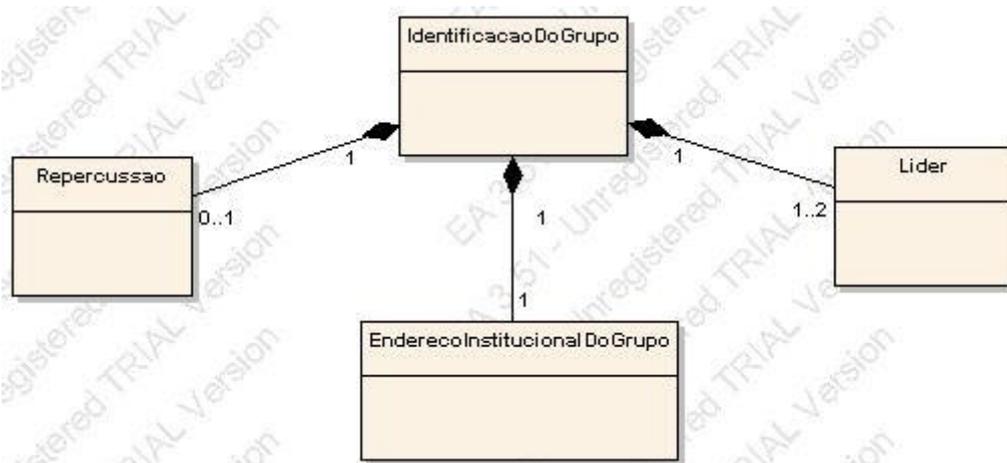


Figura 29 – Especificação das classes relacionadas com a classe “IdentificacaoDoGrupo”.

Realize esse trabalho de identificação dos relacionamentos para todas as classes definidas na unidade de informação em construção, até alcançar a especificação de todos os objetos validados.

Ao realizar a especificação da classe “Líder” apresentada na figura anterior notou-se que essa classe usa a classe “Pesquisador”. Depois de aplicada a orientação para definição dos relacionamentos, apresentada neste trabalho, chegou-se à conclusão que entre essas classes há um relacionamento de associação, conforme apresentado na Figura 30 a seguir.

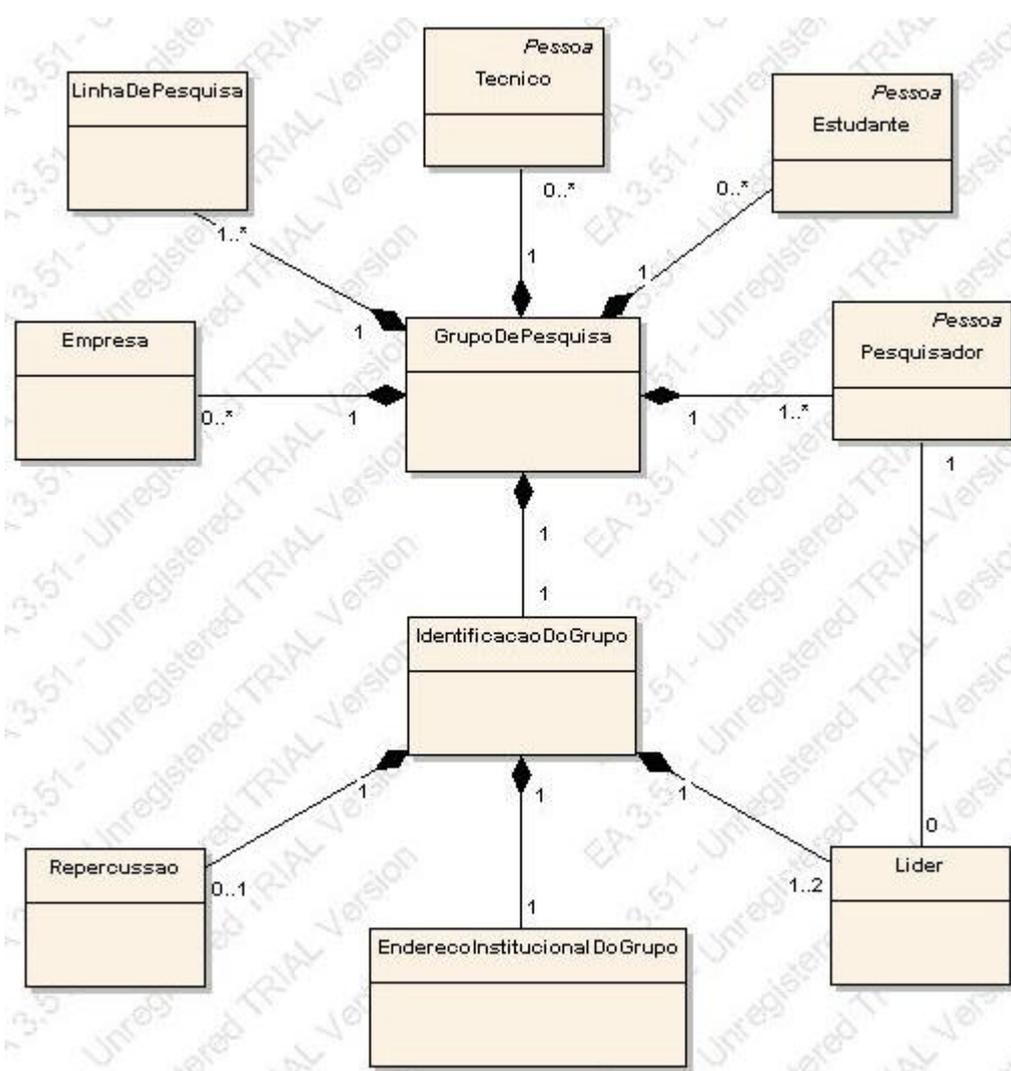


Figura 30 – Apresentação do relacionamento de associação utilizado na especificação da unidade de informação de Grupo de Pesquisa

A orientação permitiu que se chegasse a um relacionamento de associação com cardinalidade de “um” para “um”, conforme apresentado na figura anterior. Nesse tipo de relacionamento, a classe “Lider” apenas contém uma referência para a classe “Pesquisador”, que é de propriedade da classe “GrupoDePesquisa”. Conforme relatado neste trabalho, a referência da classe “Lider” ao objeto “Pesquisador” e o fato de “Proprietario” não ser de propriedade de “Lider” reafirmam as características do relacionamento do tipo associação.

Com isso, concluem-se a apresentação e a exemplificação do relacionamento do tipo associação, também utilizado para modelos de informação estruturais.

Depois de definidas as classes e seus relacionamentos, sugere-se o adicionamento das propriedades nessas classes. As orientações para realização deste trabalho se encontram na fase seguinte.

c) Fase de adicionamento das propriedades para indicar os detalhes dos valores associados com os objetos.

As classes e seus relacionamentos formam o esqueleto do modelo de informação e as propriedades completam o cenário.

Segundo Anderson *et al.* (2001), as propriedades são simplesmente valores associados com os objetos. Exemplificando: uma pessoa possui altura, peso, cor, data de nascimento, nacionalidade e outras características. Essas e outras características podem ser consideradas como propriedades de um objeto como pessoa.

Baseando-se nas informações levantadas nas etapas e fases anteriores, adicione as propriedades correspondentes para cada uma das classes modeladas.

Além do nome das propriedades, tenha o cuidado de especificar o tipo de dado de cada característica levantada para cada objeto. Verifique também se há uma faixa de valores para essas propriedades e, ainda, se essas propriedades devem possuir obrigatoriamente ou não valores. Havendo uma faixa de valores, geralmente esses valores enumerados para determinado atributo são especificados em uma outra classe com o estereótipo de <<enumeration>>.

Sugere-se que sejam adicionadas algumas propriedades para melhor identificação e manuseio dos dados da unidade de informação. Entre elas estão: a) número identificador único da UI; b) data e hora da última atualização das informações da UI; c) instituição (ou sistema) de origem para uma unidade de informação que será compartilhada por diversas instituições e que pode ser necessária em determinados momentos para identificar a origem da informação. Todas essas informações podem ser mais bem representadas na classe principal, que será a agrupadora de todas as classes.

A seguir, é apresentada na Figura 31 parte do diagrama de classes para a unidade de informação de Grupo de Pesquisa, com suas propriedades e tipos de dados definidos.

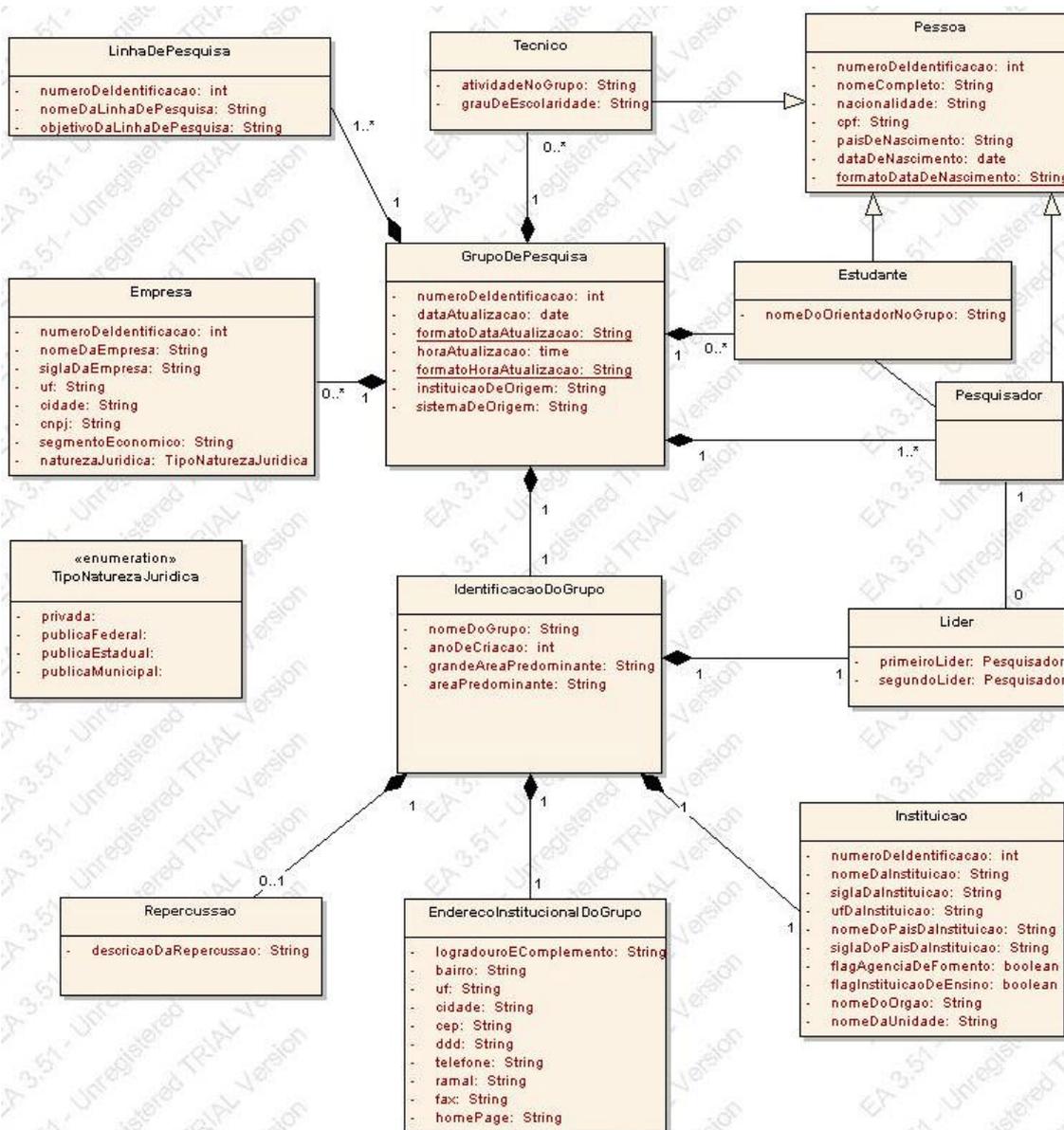


Figura 31 – Apresentação das classes para a unidade de informação de Grupo de Pesquisa com detalhamento de propriedades e tipos de dados.

Por característica do modelo de informação abordado, que é o estrutural, os valores que indicam a visibilidade das propriedades das classes, também conhecidos como atributos, não precisam ser considerados. Apesar de esses atributos estarem sendo especificados nas classes com visibilidade protegida através do sinal de negativo “-”, essa visibilidade, no caso de representação de modelos de informação, não detém importância e não será considerada na representação do modelo em uma linguagem de marcação, desenvolvida na próxima etapa.

Durante o adcionamento das propriedades nas classes, como em toda metodologia interativa de desenvolvimento, notou-se a necessidade de algumas complementações, conforme apresentado no diagrama da figura anterior. O que se nota é que por melhor que sejam trabalhadas as etapas e fases de desenvolvimento, essas modificações podem e tendem a ocorrer durante todo o processo. A consideração importante é o fato de a metodologia permitir que essas modificações e complementações possam ser ajustadas no trabalho já desenvolvido, sem grandes perdas ou danos no trabalho já realizado.

A seguir, são apresentadas e comentadas as modificações e complementações ocorridas na fase de adcionamento das propriedades às classes especificadas para a unidade de informação de Grupo de Pesquisa.

- ✍ Inserção de uma relação entre as classes “Estudante” e “Pesquisador”. Ao inserir a propriedade “nome do orientador do estudante no grupo” notou-se que esse orientador está representado na classe “pesquisador” e que há a possibilidade de realizar um vínculo entre essas duas classes.
- ✍ Inserção de uma propriedade formato de data e formato de hora para as classes que possuem informações desse tipo. Essas propriedades, conforme apresentadas no diagrama, são estáticas, ou seja, não mudam o valor, independentemente do tipo de objeto, e servem como referência indicando sempre o formato da data e da hora que estão representadas nas classes da unidade de informação.
- ✍ Criação de uma classe “Instituicao” agregada pelo tipo de composição à classe “IdentificacaoDoGrupo”. Ao descrever as propriedades da classe “IdentificacaoDoGrupo”, considerando-se a quantidade de informações sobre instituição a serem representadas, notou-se que essas seriam melhor especificadas em uma classe exclusiva para elas.
- ✍ Criação de uma classe “TipoNaturezaJuridica” para especificação dos valores possíveis enumerados do atributo “naturezaJuridica” da classe “Empresa”. Os valores possíveis para esse tipo foram especificados em forma de atributos sem indicação de tipo, pois o nome do atributo é o

próprio tipo. A próxima fase desta metodologia apresentará também uma representação na linguagem XML para os tipos enumerados.

- ✍ Ajuste no relacionamento entre as classes “IdentificacaoDoGrupo” e “Lider”. Considerando que a classe “Lider” passou a conter dois atributos “primeiroLider” e “segundoLider” que são do tipo da classe “Pesquisador”, a relação pode ser simplificada pois nesse caso específico a cardinalidade pode ser melhor representada através dos atributos.

Concluído o adição das propriedades, o modelo já está apto para representação em uma linguagem de marcação. Isso representa que a especificação, resultado dos esforços realizados nesta etapa, possibilitará a realização da etapa de padronização da unidade de informação, que será o assunto abordado na próxima seção.

3.4 PADRONIZAÇÃO DE UNIDADE DE INFORMAÇÃO

A etapa de padronização consiste na representação, através da linguagem de marcação XML, da especificação para a unidade de informação construída com base na UML. Isso permitirá a consolidação do trabalho de padronização, que tem sido desenvolvido desde a primeira etapa desta metodologia.

No entanto, esta etapa visa algo além da representação do modelo na linguagem XML. A etapa de padronização tem a finalidade também de salientar a importância de o modelo representado em XML ser uma definição de comum acordo entre os interessados.

Anderson *et al.* (2001) alertam para a importância de se ter um modelo de informação padronizado, registrado e acordado. Segundo ele, caso não se utilize um modelo nesses termos, assume-se automaticamente o risco de qualquer um ter um modelo de informação diferente em mente, com as confusões inevitáveis resultantes.

Para a viabilização dos acordos a partir das definições para as unidades de informações em desenvolvimento, sugere-se a criação de comunidades através da participação das partes interessadas. No caso da Plataforma Lattes do CNPq, os padrões XML foram acordados através de uma comunidade criada pelo CNPq para esse fim.

Criada em 2000, a Comunidade de Marcação da Plataforma Lattes, mais conhecida como Comunidade LMPL, que mais tarde se ampliou e se tornou a Comunidade CONSCIENTIAS - Comunidade para Ontologias em Ciência, Tecnologia e Informações de Aperfeiçoamento de Nível Superior (<http://lattes.cnpq.br/lmpl>), é hoje formada por agências de fomento e instituições, e atua na discussão, definição e aprovação de padrões XML para ciência e tecnologia, mais especificamente para as unidades de informação da Plataforma Lattes (CONSCIENTIAS, 2004). Mais informações sobre a Plataforma Lattes e a Comunidade CONSCIENTIAS serão apresentadas no capítulo seguinte.

Para a representação do modelo em XML, serão realizadas as atividades de desmontagem de objetos da UML para XML. Será interpretado o diagrama de classes da UML para a criação do documento XML, especificado em XML Schema, que será a linguagem utilizada nesta metodologia para especificação do metadado em XML.

Para melhor realização dessas atividades, o trabalho foi dividido nas seguintes fases:

a) Classes da UML para elementos XML

Segundo Carlson (2002), o mapeamento para XML é bastante simples quando cada instância de uma classe da UML produz um elemento em XML. Em XML, um elemento serve como um contêiner para elementos e atributos-filhos. Dessa forma, como na UML, os elementos XML (a instância de classe) servem como um contêiner para atributos e regras de associação.

É apresentado a seguir um elemento XML que será uma instância da classe “Empresa” pertencente à especificação de exemplo deste trabalho para a unidade de informação de Grupo de Pesquisa.

```
<EMPRESA>  
</EMPRESA>
```

O nome para a tag XML pode ser o mesmo que o nome da classe da UML, embora a XML imponha algumas regras adicionais sobre nomenclatura de elementos e nomes. Segundo Anderson *et al.* (2001), todos os nomes devem começar com uma letra, traço “_”, ou dois pontos “:”, e continuar usando caracteres de nomes válidos.

Caracteres de nomes válidos incluem o precedente, mais dígitos, hífen “-”, ou pontos “.”.

A sugestão deste trabalho para nomenclatura de elementos e atributos do XML consiste na representação dos nomes especificados nas classes utilizando todos os caracteres em maiúsculo, e quando houver uma classe ou atributo com nome composto, a separação pode ser feita através de um hífen. Dessa forma, o elemento XML para classe “GrupoDePesquisa” ficaria assim:

```
<GRUPO-DE-PESQUISA>
</GRUPO-DE-PESQUISA>
```

Segundo Carlson (2002), a especificação do “Modelo XML para Intercâmbio de Metadado” – XMI (*XML Metadata Interchange*), versão 1.1, permite, mas não exige, o uso de espaços de nomes (também conhecidos como *namespaces*) de XML como parte de um nome de *tag*. Exemplificando: o elemento “GRUPO-DE-PESQUISA” pode começar com um prefixo de espaço de nome, que o associa com o vocabulário XML (modelo de informação representado na linguagem XML) criado para a unidade de informação de Grupo de Pesquisa. A tag prefixada pode ser semelhante a:

```
<LMPL:GRUPO-DE-PESQUISA xmlns:LMPL="http://www.cnpq.br/lmpl">
</LMPL:GRUPO-DE-PESQUISA>
```

O prefixo servirá para distinguir esse elemento “GRUPO-DE-PESQUISA” de um elemento com o mesmo nome com origem em um vocabulário diferente. O atributo “xmlns” associa o prefixo “LMPL:” com um único *Uniform Resource Identifier* (URI – Indicador de Recurso Uniforme), que identifica o espaço de nome do vocabulário para grupo de pesquisa. Como teremos uma representação em XML para cada vocabulário ou unidade de informação, não há necessidade da utilização deste recurso.

b) Herança

A herança empregada nos diagramas de classes da UML terá sua utilização essencialmente no documento de regras que especifica o vocabulário XML para a unidade de informação. Esse documento pode ser especificado através de duas linguagens: DTD ou XML Schema, conforme citado no capítulo anterior na seção específica sobre XML.

Tratando-se do vocabulário XML em específico, a utilização ou não de herança não implica em modificações a princípio. O que pode ocorrer é a modificação no nome

dos atributos e/ou elementos que representam uma subclasse para fins de indicação explícita da herança, quando for de interesse de quem estiver criando modelo de informação (ou vocabulário) em XML.

Fato como esse, é regra na especificação do XMI (CARLSON, 2002). No entanto, na metodologia deste trabalho não se encontraram motivos para indicar no vocabulário XML a representação de que determinado elemento XML foi especificado por uma subclasse e que possui conseqüentemente informações da sua superclasse.

A seguir, apresentam-se considerações sobre as duas linguagens de especificação de vocabulários XML citadas para aplicação de herança.

Segundo Carlson (2002), um DTD não pode representar a herança entre definições de elementos, podendo apenas representar a estrutura de agregação dos elementos contidos em outros elementos.

No entanto, a linguagem mais recente para definição de tipos de documento, aprovada pelo Consórcio W3C – a XML *Schema*, permite a criação e derivação (ou extensão) de tipos que irão servir para a aplicação do recurso de herança através da representação das superclasses e subclasses. De acordo com Tesch Jr. (2002), a derivação de tipos é um mecanismo que permite a reutilização de tipos já declarados no documento XML *Schema*.

Com isso, a representação em XML *Schema* para a unidade de informação de Grupo de Pesquisa conterà tipos para as classes “Estudante”, “Tecnico” e “Pesquisador” que derivarão do tipo criado para a classe “Pessoa”. A representação do XML *Schema* para essa unidade será apresentada mais adiante, quando já se terá o vocabulário ou o modelo do documento XML definido.

c) Atributos UML para elementos XML

Dada uma instância de uma classe da UML, comece mapeando cada um dos seus atributos em XML. Segundo Carlson (2002), é comum definir cada valor de dados como um elemento-filho separado durante o projeto de vocabulários XML.

Entretanto, a metodologia deste trabalho recomenda a utilização de atributos XML em vez de elementos para representar os atributos da classe UML, por algumas

características que serão apresentadas na próxima fase, a qual trata “elementos *versus* atributos”.

A seguir, é apresentado o código XML com atributos para representação da classe “Empresa”¹:

```
<EMPRESA
  NUMERO-DE-IDENTIFICACAO="1"
  NOME-DA-EMPRESA="MANAGEMENT'S SOLUTIONS SOFTWARES S.A."
  SIGLA-DA-EMPRESA="MSS"
  UF="SC"
  CIDADE="FLORIANÓPOLIS"
  CNPJ="15.541.848/0001-87"
  SEGMENTO-ECONOMICO="PRIMARIO"
  NATUREZA-JURIDICA="PRIVADA" >
</EMPRESA>
```

Para uma classe que possui um atributo que pode ter mais de um valor, a recomendação é utilizar excepcionalmente um elemento, devido ao fato de a especificação da XML não permitir que se tenha mais de um atributo com o mesmo nome em um mesmo elemento XML.

d) Elementos versus atributos

Segundo Carlson (2002), a distinção entre elementos e atributos pode não ser importante quando o vocabulário é usado principalmente para intercâmbio de dados. No entanto, estima-se que a utilização de um ou de outro apresenta diferenças no documento, principalmente em relação à quantidade de texto que estará contida no documento XML e, conseqüentemente, ao tamanho do arquivo em bytes para esse documento.

A utilização de atributos reduz o tamanho do texto do documento e, como conseqüência, o tamanho desse arquivo em bytes será menor. Isso se deve ao fato de que, para representar um atributo, é necessário especificar o nome dele uma única vez e o seu valor entre aspas (“”). Já para o elemento, é necessário ter o nome dele uma vez para indicar a sua abertura, e outra para indicar o fechamento. Ambas as vezes (abertura e fechamento) têm-se ainda que indicar para o elemento os sinais de maior e menor (“<”, “>”). A seguir, essa comparação é exemplificada através da representação XML:

¹ Os dados apresentados no trecho XML para empresa são ilustrativos e fictícios, e não possuem valor comercial.

Representação XML para a informação “NomeDaEmpresa” da classe “Empresa” utilizando atributo:

```
<EMPRESA NOME-DA-EMPRESA="MANAGEMENT'S SOLUTIONS SOFTWARES S.A.">
</EMPRESA>
```

Ou ainda, como o elemento empresa não possui nenhum texto ou outro elemento, a representação poderia ser da seguinte forma:

```
<EMPRESA NOME-DA-EMPRESA="MANAGEMENT'S SOLUTIONS SOFTWARES S.A."/>
```

Agora, a representação XML para a informação “NomeDaEmpresa” da classe “Empresa” utilizando elemento:

```
<EMPRESA>
  <NOME-DA-EMPRESA>MANAGEMENT'S SOLUTIONS SOFTWARES S.A.
</NOME-DA-EMPRESA>
</EMPRESA>
```

No caso de representação da informação “NomeDaEmpresa”, não há possibilidade de representar o XML da segunda forma, ou seja, omitir a tag de fechamento e utilizar a barra “/” no final da tag de início. Isso se deve ao fato de o elemento “EMPRESA” neste exemplo possuir o elemento “NOME-DA-EMPRESA”.

Outro fator importante para decisão na escolha entre representar em XML as informações em forma de atributos ou elementos é a possibilidade de especificar o tipo dessa informação. As especificações atuais, tanto para DTD ou XML *Schema*, não fornecem opção para detalhamento de tipo de dado, a não ser que esse esteja representado na forma de atributo.

Carlson (2002) considera também importante para a escolha entre optar por representar em XML as informações como atributo ou como elemento o fato de como as informações são apresentadas visualmente em editores ou ferramentas XML. Segundo ele, quando um usuário humano é um dos processadores das informações do vocabulário, o fato de se ter uma melhor apresentação visual da informação nessas ferramentas, seja com atributo ou com elemento, pode ser algo representativo para essa escolha.

Contudo, acredita-se que comparadas às demais características apresentadas, esse fator é relativo e não deve ser necessariamente considerado. O simples fato de novas versões de ferramentas ou editores XML fornecerem uma nova forma de visualização pode desbancar a justificativa por determinada escolha.

Foram apresentadas algumas características comparativas sobre a representação das informações em XML com atributos e com elementos. Seja qual for a opção escolhida para representar em seu vocabulário, procure utilizá-la durante a definição de todo o vocabulário. Carlson (2002) afirma que o mais importante é ser consistente no uso dos elementos e atributos XML para um dado vocabulário.

e) Valores de atributos enumerados

Um atributo enumerado exige que o valor de atributo UML seja atribuído a partir de uma lista finita de valores possíveis. A Figura 31 apresenta a classe “Empresa”, que contém um atributo enumerado. Esse atributo consiste em uma entidade nomeada, que foi especificada na classe “TipoNaturezaJuridica” estereotipada como “*enumeration*”. Exemplificando: de acordo com a especificação, a natureza jurídica de uma empresa na unidade de informação de Grupo de Pesquisa pode ser pública, privada federal, privada estadual ou privada municipal.

A especificação realizada para atributos enumerados pode ser feita de duas formas em XML. A primeira é através da forma de atributo, e a segunda, na forma de elementos.

A seguir, apresentam-se as representações em XML para o atributo “naturezaJuridica” da classe “Empresa”. A primeira consiste na representação na forma de atributo, como se pode visualizar abaixo.

```
<EMPRESA NATUREZA-JURIDICA="PUBLICA" />
```

E a segunda na forma de elemento:

```
<EMPRESA>
  <NATUREZA-JURIDICA>PUBLICA</NATUREZA-JURIDICA>
</EMPRESA>
```

A representação em forma de atributo não parece ser de modo diferente do que outros atributos não enumerados, mas o DTD e XML *Schema* fornecem recursos para especificação em XML dos valores permitidos para esses atributos que contêm valores enumerados. O mesmo não ocorre para os elementos. Com isso, a metodologia deste trabalho salienta a recomendação para que sejam utilizados atributos para representação dos atributos das classes UML, principalmente quando esses detiverem valores enumerados.

f) Representação de composições da UML

A representação de composições da UML consiste basicamente na inserção de um elemento para cada instância das classes agregadas dentro do elemento que representa a classe proprietária.

A representação em XML para as classes agregadas à classe principal da unidade Grupo de Pesquisa, denominada “GrupoDePesquisa”, é apresentada a seguir.

```
<GRUPO-DE-PESQUISA>
  <IDENTIFICACAO-DO-GRUPO/>
  <PESQUISADOR/>
  <ESTUDANTE/>
  <TECNICO/>
  <LINHA-DE-PESQUISA/>
  <EMPRESA/>
</GRUPO-DE-PESQUISA>
```

Neste exemplo está sendo apresentado um relacionamento de “um” para “um”. De acordo com o modelo de classes UML para essa unidade, “GrupoDePesquisa” deve ter no mínimo um, podendo ainda ter mais de um “Pesquisador”. A seguir, apresenta-se o exemplo anterior alterado para representar o relacionamento de “um” para “três” para as classes “GrupoDePesquisa” e “Pesquisador”:

```
<GRUPO-DE-PESQUISA>
  <IDENTIFICACAO-DO-GRUPO/>
  <PESQUISADOR/>
  <PESQUISADOR/>
  <PESQUISADOR/>
  <ESTUDANTE/>
  <TECNICO/>
  <LINHA-DE-PESQUISA/>
  <EMPRESA/>
</GRUPO-DE-PESQUISA>
```

Em alguns casos de ocorrência de elementos com o mesmo nome, pode-se encontrar documentos XML que detêm elementos com função específica de agrupamento desses elementos que possuem o mesmo nome. Essa forma é utilizada simplesmente para fornecer uma melhor organização desses elementos repetidos, agrupando-os dentro de um elemento específico para eles. Geralmente, o nome desse elemento é o mesmo nome dos elementos que estão sendo agrupados nele, com alteração para o nome ficar no plural e representar o grupo. A seguir, apresenta-se o exemplo em XML para essa forma:

```

<GRUPO-DE-PESQUISA>
  <IDENTIFICACAO-DO-GRUPO/>
  <PESQUISADORES>
    <PESQUISADOR/>
    <PESQUISADOR/>
    <PESQUISADOR/>
  </PESQUISADORES>
  <ESTUDANTE/>
  <TECNICO/>
  <LINHA-DE-PESQUISA/>
  <EMPRESA/>
</GRUPO-DE-PESQUISA>

```

No entanto, como a especificação de classes em UML para o exemplo abordado não contemplou a existência de classes com função específica de agrupar e separar os elementos com o mesmo nome, neste trabalho será utilizada a primeira forma apresentada para representação em XML.

Sugere-se que, qualquer que seja a forma utilizada, essa mesma sistemática seja empregada em todos os casos do modelo que possuir o relacionamento de composição de “um” para “muitos”.

g) Representação de associações da UML

A representação do relacionamento de associação de classes UML em XML pode ser feita através de atributos que terão a função de indicar e referenciar os elementos relacionados, estabelecidos com base nos recursos adicionais da especificação XMI.

A seguir, apresenta-se na Figura 32, parte do diagrama de classes UML do exemplo da unidade de informação de Grupo de Pesquisa, que contém um caso de relação do tipo associação.

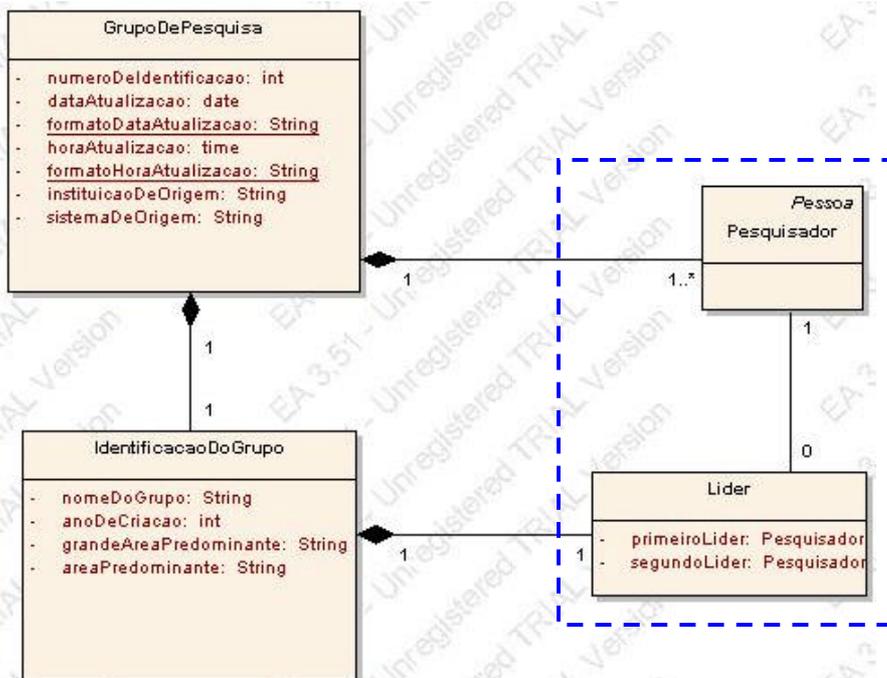


Figura 32 – Parte do diagrama de classes UML, criado para a unidade de informação de Grupo de Pesquisa, que contém classes com relacionamento do tipo associação

Como se pode notar na figura anterior, as classes que possuem relação de associação estão delimitadas por uma linha pontilhada. Os elementos para representação em XML para essas classes, “Pesquisador” e “Líder”, deverão possuir os atributos necessários para permitir a identificação e a indicação de referência entre eles.

A seguir, apresenta-se a representação em XML para essas duas classes que conterão nelas atributos que irão indicar o relacionamento, denominados respectivamente “INDICADOR” no elemento “PESQUISADOR”, e “REFERENCIA” no elemento “LIDER”:

```

<GRUPO-DE-PESQUISA>
  <PESQUISADOR IDENTIFICADOR="1"
    NOME-COMPLETO="Roberto Carlos dos Santos Pacheco"/>
  <PESQUISADOR IDENTIFICADOR="2"
    NOME-COMPLETO="Paulo Henrique de Souza Bermejo"/>
  <IDENTIFICACAO-DO-GRUPO>
    <LIDER REFERENCIA-IDENTIFICADOR-PRIMEIRO-LIDER="1"
      REFERENCIA-IDENTIFICADOR-SEGUNDO-LIDER="2"/>
  </IDENTIFICACAO-DO-GRUPO>
</GRUPO-DE-PESQUISA>
  
```

Conforme apresentado, a classe “Líder” que possui atributos que são do tipo referência “Pesquisador” terão na representação em XML um atributo referência. O outro elemento, que representa a classe “Pesquisador”, a qual é associada, detém um atributo identificador que conterá um valor único e permitirá assim a sua referência.

A nomenclatura desses elementos pode ser estabelecida de acordo com o exemplo apresentado ou, ainda, utilizando-se os próprios atributos do XMI, como *xmi.id* e *xmi.idref*. De acordo com Carlson (2002), ambas as formas são aceitas pela especificação XMI.

Seja qual for a nomenclatura dos atributos, a forma de garantir que eles estejam referenciando valores existentes e de maneira correta é indicar no documento de especificação para esse modelo XML, XML *Schema* ou DTD o atributo identificador sendo do tipo “ID”, e o atributo de referência, do tipo “IDREF”. E ainda, essa certificação de valores está estritamente ligada ao fato da existência e utilização desse documento de especificação detentor das regras de validação.

Outra forma de representar um relacionamento de associação consiste na utilização das tecnologias XLink e XPointer (CARLSON, 2002). Com essas tecnologias XML, é possível indicar uma associação mesmo quando o atributo que será referenciado não possui um atributo identificador. Esse mecanismo de referência é possível através da tecnologia XPath, que possibilita a indicação do caminho do documento XML, utilizada pelo XPointer.

Com isso, a representação XML para o elemento “Pesquisador” não precisaria conter o atributo identificador para poder ser referenciada. E ainda, é possível estabelecer a associação entre elementos XML mesmo quando eles pertencem a documentos físicos distintos (isso nos casos em que uma unidade de informação está subdividida em diferentes documentos XML).

Essas tecnologias comentadas para indicar no XML a associação, como XLink e XPointer, são comuns na representação em XML de relacionamentos do tipo agregação, assunto que será abordado a seguir.

h) Representação de agregação da UML

Conforme citado na etapa de especificação, o relacionamento de agregação entre classes da UML surge quando uma unidade de informação é formada por partes de outras unidades, ou até mesmo quando uma unidade de informação é subdividida em partes entre si. No caso de representação em XML, essas partes seriam documentos XML separados fisicamente. Nesse segundo caso, isso pode ocorrer quando a quantidade de informação for muito extensa para ser representada em um único modelo de classes e, conseqüentemente, em um único documento XML, e necessitar de ser segmentada.

O Consórcio W3C oferece tecnologias que permitem representar essas relações de agregação em XML e ainda possibilitar a realização da Web Semântica (CARLSON, 2002). Essas tecnologias, já citadas para representação da associação, consistem em XLink e XPointer (W3C, 2003). Através delas é possível indicar que determinado elemento de um documento XML contém um ou mais elementos que fazem parte de outro documento XML

Devido ao tamanho do modelo para a unidade de informação de Grupo de Pesquisa, modelo este que não é extenso, não houve a necessidade de segmentá-lo. Com isso, a aplicabilidade do relacionamento de agregação nesse caso não demonstrou-se ser adequada.

Entretanto, considerando-se a intenção de exemplificar também esse tipo de relacionamento em XML, poder-se-ia considerar, com critérios apenas de ilustração, que especificamente para o elemento “Pesquisador” da unidade de informação de Grupo de Pesquisa há a necessidade de ele conter as produções bibliográficas que estaria em um outro modelo de informação denominado “Currículo”, definido e padronizado pela Comunidade CONSCIENTIAS/LMPL – CONSCIENTIAS (2004).

A seguir, apresenta-se na Figura 33 parte do modelo de classes UML para a unidade de informação de Grupo de Pesquisa, que contém relação com o modelo de classes UML criado com base na especificação em *XML Schema* da Comunidade CONSCIENTIAS (CONSCIENTIAS, 2004) para a unidade de informação de Currículo.

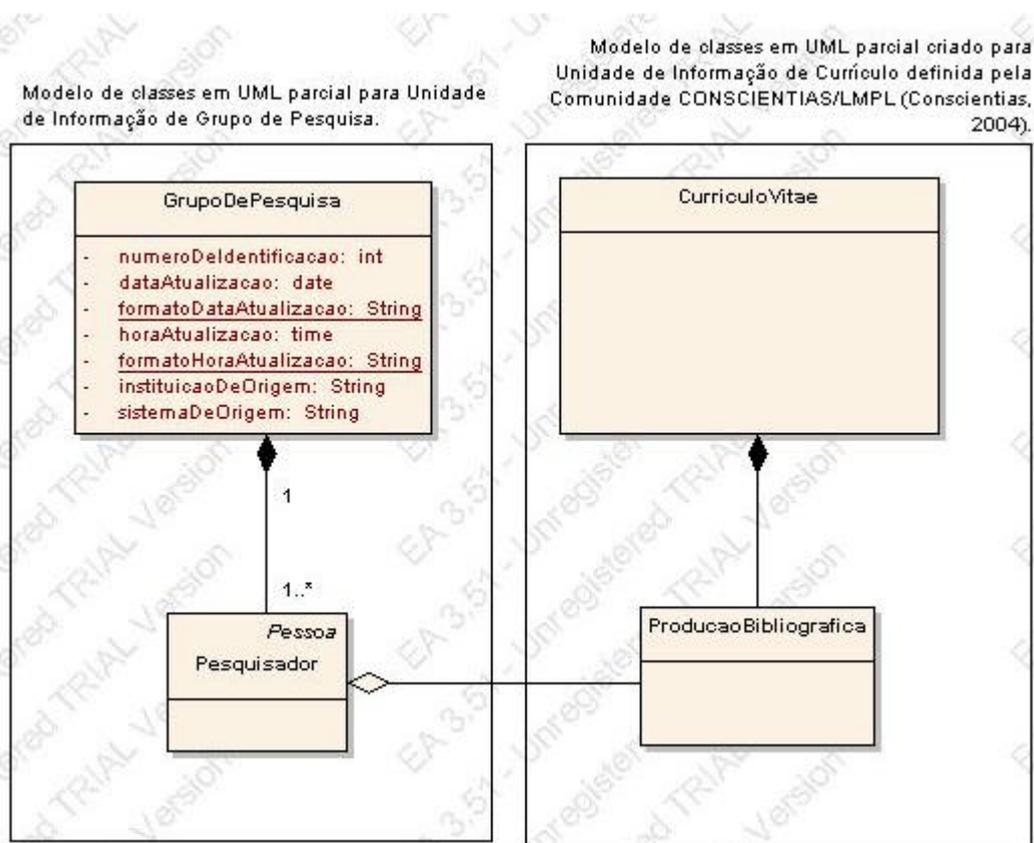


Figura 33 – Relacionamento de agregação entre as unidades de informação Grupo de Pesquisa e Currículo

A representação XML para o modelo de classe parcial para a unidade de informação de Currículo, que terá parte dela agregada à classe Pessoa da unidade de informação de Grupo de Pesquisa, é apresentada a seguir.

```

<CURRICULO-VITAE>
  <PRODUCAO-BIBLIOGRAFICA/>
</CURRICULO-VITAE>
  
```

Na seqüência, apresenta-se a representação XML para o elemento “PESQUISADOR”, que conterá uma relação de agregação ao elemento “PRODUCAO-BIBLIOGRAFICA” da unidade de informação de Currículo, de acordo com o modelo de classe apresentado na Figura 33.

```

<GRUPO-DE-PESQUISA>
  <PESQUISADOR IDENTIFICADOR="1"
  NOME-COMPLETO="Roberto Carlos dos Santos Pacheco">
    <PRODUCAO-BIBLIOGRAFICA>
      Currículo.xml#xpointer(/CURRICULO-VITAE/PRODUCAO-BIBLIOGRAFICA)
    </PRODUCAO-BIBLIOGRAFICA>
  </PESQUISADOR>
  
```

</GRUPO-DE-PESQUISA>

Tendo-se concluído a forma de representação em XML da agregação e demais relacionamentos da especificação para uma UI realizada na etapa anterior, salienta-se a importância de se ter um mecanismo que possa descrever o conteúdo da unidade de informação padronizada e ainda garantir que esteja no formato esperado, através das tecnologias *DTD* ou *XML Schema*.

3.5 GESTÃO DE UNIDADE DE INFORMAÇÃO

A etapa de gestão de unidade de informação é a etapa final da metodologia proposta neste trabalho. Ela está baseada na proposição de diretrizes visando proporcionar um melhor gerenciamento de uma unidade de informação e de suas etapas de desenvolvimento.

As diretrizes sugeridas nesta etapa estão baseadas no modelo de gestão de ontologias da Comunidade CONSCIENTIAS, conforme apresentado na Figura 34, que estabelece um ciclo para desenvolvimento de uma ontologia.

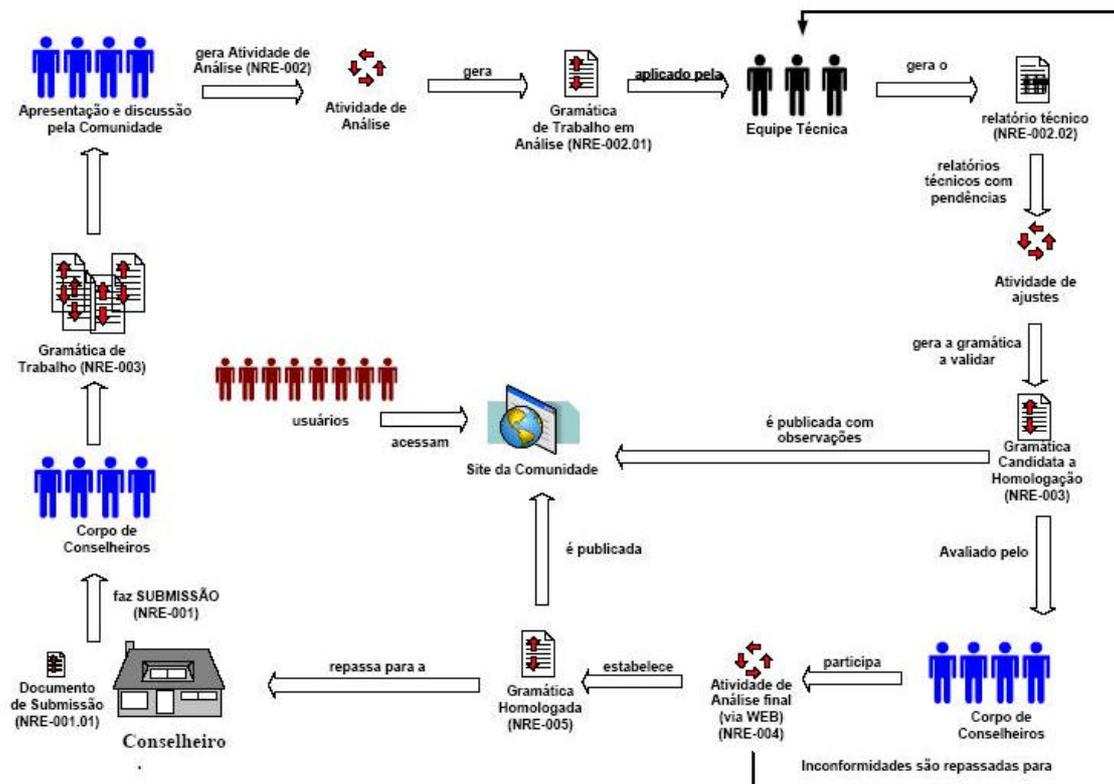


Figura 34 – Ciclo da Comunidade CONSCIENTIAS para desenvolvimento de ontologias (CONSCIENTIAS, 2004)

O principal objetivo desse ciclo de desenvolvimento está em possibilitar a discussão, o desenvolvimento e a homologação de versões de uma unidade de informação visando assegurar que a versão final do trabalho realizado esteja de acordo com o esperado pelos seus interessados.

As atividades principais deste ciclo, conforme apresentado na figura anterior, são:

- i. submissão de uma proposta. É realizada por um conselheiro (pessoa representante de uma instituição ou agência que possui participação no conselho da Comunidade);
- ii. análise da proposta pelo corpo de conselheiros e criação de uma “gramática de trabalho”;
- iii. apresentação e discussão pela Comunidade para geração de um “documento de análise”;
- iv. modificação da gramática proposta com base no “documento de análise” para definição da “gramática de trabalho em análise”;
- v. avaliação por uma equipe técnica, que pode ser indicada ou formada pelos conselheiros da Comunidade. A avaliação será relatada em um “documento técnico”;
- vi. criação da “gramática candidata à homologação”, na qual constarão as modificações relatadas no “documento técnico”. A partir dessa versão, a gramática passa a estar disponível de forma pública; e
- vii. atividade de análise final. Pode ser feita via *Web* através de ferramenta específica no *site* da Comunidade. Caso seja acordada alguma definição incorreta, essa gramática volta para a equipe técnica (“atividade v”), de modo que possa verificar as pendências. Caso contrário, a gramática obtém o status de “gramática homologada”, é publicada no *site* da Comunidade e é repassada ao conselheiro solicitante.

Depois de homologada a gramática XML para a unidade, sugere-se que seja adotado o recurso de controle de versões da especificação para o modelo de informação em XML. Esse recurso visa fornecer uma garantia de que o padrão para a unidade de informação não será alterado, a menos que seja em uma nova versão.

A informação de versão pode ser facilmente disponibilizada em uma especificação DTD ou XML Schema na forma de um comentário ou através do elemento *documentation* do XML Schema. Para uma melhor identificação ao usuário do modelo que está sendo utilizado, sugere-se que sejam consideradas no controle de versões informações como data de atualização ou publicação da versão, número da versão, descrição sucinta da versão e descrição detalhada das modificações realizadas.

Apresenta-se a seguir, na Figura 35, a ferramenta *Web* da Comunidade CONSCIENTIAS para acompanhamento das versões das unidades de informação publicadas.

The screenshot shows a web browser window with the URL <http://lattes.cnpq.br/imp/index.jsp?go=histVersao.jsp?ontologia=cur>. The page title is 'Histórico das versões de Ontologia'. The main content area contains the following text:

Informações:
Ontologia XML para a Unidade de Informação de Currículo

Mais informações sobre esta ontologia podem ser obtidas em:
<http://lattes.cnpq.br/imp/index.jsp?go=cur.jsp>

Para download do XML Schema da Ontologia clique em:
<http://lattes.cnpq.br/imp/gramaticas/Curriculo/WS/Fontes/LMPLCurriculo.xsd>

Para download da documentação do XML Schema da Ontologia clique em:
<http://lattes.cnpq.br/imp/gramaticas/Curriculo/WS/Documentacao/HTML/>

Para download do DTD da Ontologia clique em:
<http://lattes.cnpq.br/imp/gramaticas/Curriculo/DTD/Fontes/LMPLCurriculo.DTD>

Para download da documentação do DTD da Ontologia clique em:
<http://lattes.cnpq.br/imp/gramaticas/Curriculo/DTD/Documentacao/DTDCurriculo.pdf>

Versão	Data	Descrição	Detalhes
14	01/03/2004	Ajustes de tipos de dados de alguns atributos	Com o trabalho de inserção desta gramática no Sistema de Currículo Lattes, verificou-se a necessidade de ajustar alguns tipos de dados implementados em XML Schema para uma melhor validação das informações e também para possibilitar ao Sistema de Currículo Lattes realizar a importação sem erros de currículos em XML que detêm dados vindos de versões antigas do Lattes ou até mesmo de outros sistemas (ambos os casos os dados se originaram das bases de dados do CNPq).
14.1	03/03/2004	Indução de atributo SEQUENCIA	Alterações: a) Indução do atributo SEQUENCIA-AREA-DE-OTIACAO no elemento AREA-DE-OTIACAO. b) Alteração do

Figura 35 – Ferramenta da Comunidade CONSCIENTIAS/LMPL para acompanhamento das versões das unidades de informação.

3.6 FASE DE APLICAÇÃO À PLATAFORMA DE E-GOV

A metodologia proposta neste trabalho foi desenvolvida visando ser aplicada como etapa-base no desenvolvimento de uma plataforma de e-gov. A aplicação na fase inicial de desenvolvimento foi estabelecida sendo fundamentada na proposta metodológica de e-gov de Pacheco (2003).

Para os casos de uma plataforma já desenvolvida ou em estágio de desenvolvimento mais adiantado, a metodologia proposta neste trabalho pode ser aplicada visando alcançar a padronização das unidades de informação completadas nos seus sistemas. O simples fato de se considerarem as interfaces dos sistemas e o modelo de base de dados para identificar quais são as informações para uma unidade, na etapa de Identificação, elucida essa possibilidade de se aplicar a metodologia em outras fases.

Entretanto, acredita-se que, caso a metodologia seja aplicada no início do desenvolvimento de determinada plataforma, isso pode trazer mais benefícios do que se sua aplicação for realizada posteriormente em outras fases, pois as unidades definidas e padronizadas podem servir como essência na definição de sistemas, portais e descoberta de conhecimento.

4 APLICAÇÃO: GESTÃO DE UNIDADES DE INFORMAÇÃO EM C&T – PLATAFORMA LATTES

4.1 APRESENTAÇÃO

Nesse capítulo é apresentada a aplicação na Plataforma Lattes da metodologia proposta nesta dissertação. Esta Plataforma constitui um projeto do Conselho Nacional de Desenvolvimento Científico e Tecnológico do Brasil - CNPq para promover Ciência, Tecnologia e Inovação.

Mas antes disso, serão citados a plataforma europeia denominada EuroCris e o seu trabalho no âmbito da C&T.

Em seguida, comenta-se a Plataforma Lattes, que detém padronização XML para diversas unidades de informação, tais como Currículo, Grupo de Pesquisa, Instituição em C&T, Projetos de Pesquisa, Periódicos Científicos (CONSCIENTIAS, 2004). Essas unidades de informação em C&T também são conhecidas como unidades de análise, nome este mais específico para as unidades de informação na área de C&T. Essa nomenclatura pode variar de acordo com a área de aplicação para a taxonomia de subdomínios na qual se baseia. Por exemplo, caso seja aplicada ao setor de energia elétrica, essa unidade pode ser denominada como unidade de controle.

Ainda no contexto da Plataforma Lattes são citados os esforços da Comunidade CONSCIENTIAS/LMPL no processo de padronização das unidades de análise, englobando as etapas propostas nessa metodologia, tais como identificação, especificação, padronização e gestão dessas no domínio de conhecimento para a qual foi projetada, isto é, Ciência e Tecnologia. Atentando para a internacionalização das unidades padronizadas, é comentada a solução existente hoje proposta por Pacheco, Kern e Bermejo (2003), que dispõe de mecanismos para tradução das unidades de informação representadas em XML em diferentes idiomas. Essa tradução é proposta por uma ferramenta complementar para internacionalização das especificações em XML Schema, que além de ser utilizada de forma independente, pode ser acoplada à existente (que também será comentada mais adiante), visando dispor de uma solução de internacionalização de unidades de informação padronizadas em XML desde a

especificação em XML Schema ate a integração de informações que estejam nesse formato em sistemas de diferentes idiomas.

Concluindo, são apresentados os recursos da Plataforma Lattes viabilizados através do trabalho de padronização das unidades de informação.

4.2 PLATAFORMA EUROPÉIA – EUROCRIS

Antes mesmo de se abordar com mais detalhes a Plataforma Lattes neste capítulo, cita-se uma plataforma de sistemas e pesquisa de grande alcance na Europa. Conhecida também como entidade, essa plataforma é a EuroCRIS (*Europe Current Research Information Systems*).

A EuroCRIS é uma entidade sem fins lucrativos e objetiva ser ponto de referência internacional para todas as matérias relacionadas ao *Current Research Information Systems – CRIS* (Sistemas de Informação de Pesquisas Correntes). Com condução de seus trabalhos enfatizados no continente europeu, a EuroCRIS serve seus membros e a comunidade global de pesquisa, avançando os campos do CRIS através do desenvolvimento profissional de longo prazo e do compartilhamento de conhecimento, e promovendo um sentido de comunidade verdadeira (EUROCRIS, 2004).

O CRIS pode ser organizado tematicamente ou em torno de linhas de disciplinas científicas, podendo ser regional, nacional ou internacional. Apesar de não limitadas, algumas de suas tarefas são: bases de dados globais; definição de tipos de informação; padrões e linhas de orientação; melhores práticas; acesso a dados e mecanismos de intercâmbio; e adereços de outras tarefas de padronização dentro de ambiente de pesquisa, treinamento e desenvolvimento (pesquisa e desenvolvimento), no prazo e de uma maneira eficiente (EUROCRIS, 2004).

Na realização de suas missões, a EuroCRIS capitaliza no desenvolvimento internacional na ciência e ciência política, particularmente com referência para a Área de Pesquisa Européia – ERA (*European Research Area*) e o envolvimento necessário para os grupos usuários (EUROCRIS, 2004). Entre os seus projetos destacam-se o DRIS, o ERGO e o CERIF.

O DRIS (*Directory of Research Information Systems*), ou *Research Information Systems Worldwide*, como também é conhecido, é um diretório *Web* de pesquisa de sistemas de informação. Os sistemas de informação cobertos pelo DRIS fornecem informações sobre organizações de pesquisa, pesquisadores e cientistas, e programas e/ou projetos de pesquisa. Alguns sistemas de informação podem conter também publicações, patentes e ou artigos publicados (DRIS, 2004).

ERGO - *European Research Gateways On-line* é um portal on-line de pesquisa e desenvolvimento de projetos europeus de informação. Em fase piloto de desenvolvimento, ele dará acesso a um catálogo de projetos de pesquisa e desenvolvimento de cerca de 20 a 30 sistemas de informação europeus de pesquisa (ERGO, 2004).

Já o CERIF - *Common European Research Information Format* é um projeto que define um formato comum de informações de pesquisa usado pela Europa. Ele foi recomendado para os Estados-Membros da Comunidade Européia como um mecanismo para proporcionar harmonia em bases de dados de projetos de pesquisa e também para disponibilizar recursos que possibilitem a comparação de um lado das informações que foram compartilhadas e linhas gerais para a construção de bases de dados de pesquisa (CERIF, 2004).

Entre outros projetos que estão relacionados com o EuroCRIS, pode-se citar o Projeto *SCHEMAS*. Esse projeto forneceu um fórum para desenvolvedores de padrões de metadados envolvidos no programa *Information Society Technologies* – IST e iniciativas nacionais na Europa. O projeto *SCHEMAS* tem informado aos seus usuários sobre o status e a propriedade de uso de novos e emergentes metadados. Possui suporte de desenvolvimento de linhas gerais de boas práticas (*good-practice guidelines*) para o uso de padrões em implementações locais. E ainda, tem investigado sobre como metadados registrados podem suportar esses objetivos (SCHEMAS, 2004).

4.3 PLATAFORMA LATTES

Lattes, nome de um renomado físico brasileiro, originou o nome de uma plataforma brasileira de ciência e tecnologia de sistemas de informação e portais *Web*

voltados para a gestão de Ciência e Tecnologia, conhecida como Plataforma Lattes (LATTES, 2004).

Lançada em 16 de agosto de 1999 com o surgimento do sistema de Currículo Lattes, a Plataforma Lattes nasceu de uma iniciativa do CNPq em unificar os sistemas de informação curriculares, Programa de Apoio ao Desenvolvimento Científico e Tecnológico - PADCT, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Capes, Financiadora de Estudos e Projetos – Finep e outros sistemas das fundações de apoio dos estados a começar pela Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP, além da sua agência mantenedora o CNPq (PEREIRA, 2002). Através dessa unificação os pesquisadores brasileiros puderam gerar seus currículos em um único local e, por sua vez, submeter através do sistema seus dados ao CNPq (LATTES, 2004).

De acordo com as informações disponibilizadas no *site* do CNPq na primeira quinzena de julho de 2004, a Plataforma Lattes, que tem como seu principal desenvolvedor o Grupo Stela, detém em seu repositório de informações cerca de 403 mil currículos. Além das informações curriculares, existem também no repositório do CNPq informações sobre aproximadamente 18 mil e 300 grupos de pesquisa certificados, que possibilitam ao CNPq realizar a classificação desses grupos para o seu censo anual (LATTES, 2004).

Essas e outras informações constantes na Plataforma Lattes não têm servido apenas para as agências de fomento. Instituições de ensino e pesquisa, grupos de desenvolvimento, empresas, profissionais como jornalistas e outros, e até mesmo usuários em geral podem extrair através do portal *Web* da plataforma inúmeras informações relacionadas à ciência e tecnologia, através do acesso aos currículos e dos grupos de pesquisa disponíveis no repositório do CNPq. A Plataforma ainda dispõe de uma grande quantidade de indicadores por área de concentração, instituição, linha de pesquisa e outros, que podem possibilitar aos diversos atores a tomada de determinadas decisões de acordo com os seus interesses.

O Jornal da Ciência (CIÊNCIA, 2004a) classifica a Plataforma Lattes como uma verdadeira vitrine e memória da C&T brasileira e ainda como a ante-sala, do mais completo mapeamento de uma comunidade científica tecnológica existente no mundo.

Alcançando a marca dos 7,5 milhões de acessos ao seu portal *Web*, a Plataforma passou a ser internacional desde 2002.

No âmbito brasileiro, a Plataforma Lattes recebe mais uma marca de reconhecimento da sua importância no campo de e-gov. Em sua terceira edição, o Prêmio e-gov que é Organizado pela Associação Brasileira de Empresas Estaduais de Processamento de Dados – ABEP e pelo Ministério do Planejamento, Orçamento e Gestão, premiou a Plataforma Lattes como primeira colocada na categoria “Governo para Cidadão” – G2C no ano de 2004 (PRÊMIO E-GOV, 2004).

Essas considerações têm chegado como incentivos a essa plataforma de sistemas e gestão do conhecimento na busca contínua da excelência no campo de governo eletrônico. Recentemente, a Plataforma Lattes iniciou o lançamento, através do sistema InterLattes CV-Resume, de uma linha de sistemas para tratamento da qualidade da informação e gestão de conhecimento. Esses sistemas foram concebidos também seguindo a linha de desenvolvimento de sistemas de governo eletrônico estabelecida na arquitetura de Pacheco (2003). Conforme já mencionado sobre essa arquitetura, que é a arquitetura referência do trabalho desta dissertação, ela prevê a realização de sistemas de conhecimento, como os citados, depois de ter alcançado (a) definição e padronização de unidades de informação, (b) sistemas de informação e fontes de informação, (c) publicação de portais e serviços.

Disponível desde dezembro de 2003 no portal da Plataforma Lattes, o sistema CV-Resume, que já é utilizado por mais de 17 mil usuários, permite obter um resumo em um único parágrafo de um currículo, independentemente do seu tamanho. Considera-se que para os casos em que há necessidade se obter uma visão geral do usuário, isso pode ser extremamente útil, de preferência nos casos em que o número de páginas do currículo é extenso, podendo chegar até mesmo 70 ou 80 (CIÊNCIA, 2004b).

Outro recurso de conhecimento a ser disponibilizado aos usuários da Plataforma Lattes é o Sistema InterLattes CV-Perfil. Em fase de desenvolvimento, esse sistema produzirá informações conclusivas privativas ao usuário. Com base em *hiperlinks*, cada sentença de conclusão sobre o currículo apresentada no CV-Perfil, permitirá que o usuário reveja informações cadastradas no Currículo Lattes.

Sintetizando, esse sistema possibilitará ao usuário o encontro de informações analíticas sobre seus dados e ou ainda permitirá a identificação de dados preenchidos erroneamente bem como campos ainda nem preenchidos (CIÊNCIA, 2004b).

De acordo com Ciência (2004b), todas essas ferramentas da Plataforma Lattes permitem a descoberta de novos conhecimentos escondidos na massa de informações sobre Ciência, Tecnologia e Inovação do País. Essas tecnologias tornam disponível para a comunidade científica uma nova linha de recursos informacionais, inviáveis sem a Plataforma Lattes, e possibilitam a criação de novos espaços de colaboração e de gestão.

Entretanto, isso pode ser alcançado também graças aos esforços da Plataforma Lattes para possuir unidades de informação padronizadas, a fim de que possa contribuir na construção dos seus sistemas de informação, seus portais *Web* e, por fim, de seus sistemas de conhecimento e gestão, comentados há pouco.

Seguindo a proposta arquitetural de e-gov de Pacheco (2003), a Plataforma Lattes possui unidades de informação padronizadas em XML, por meio do trabalho realizado pela Comunidade CONSCIENTIAS. De acordo com Conscientias (2004), essa comunidade é constituída por três agências de fomento e por dez instituições. As agências de fomento são CNPq, CAPES e FAPESP, e as instituições constituem Fundação Oswaldo Cruz - Fiocruz, Universidade Federal da Bahia - UFBA, Universidade Federal de Minas Gerais - UFMG, Universidade Federal do Rio Grande do Sul - UFRGS, Universidade Federal do Rio de Janeiro - UFRJ, Universidade Federal do Rio Grande do Norte - UFRN, Grupo Stela em conjunto com a Universidade Federal de Santa Catarina - UFSC, Universidade Estadual de Campinas – Unicamp e Universidade de São Paulo – USP.

Essa comunidade tem a responsabilidade específica de definir e homologar entre os seus participantes um padrão XML para cada unidade de informação pertencente a essa plataforma de sistemas. A referida comunidade é uma extensão da Comunidade LMPL (Linguagem de Marcação da Plataforma Lattes), estabelecida no ano 2000 para ser responsável pela criação e manutenção das gramáticas XML da Plataforma Lattes. Sua criação coroa o processo de aproximação entre agências federais e estaduais, em um movimento de padronização de informações e racionalização de

procedimentos, envolvendo fornecimento e intercâmbio de informações em benefício das comunidades científicas, tecnologias e de educação superior (CONSCIENTIAS).

4.4 UNIDADES DE INFORMAÇÃO DA PLATAFORMA LATTES

A Plataforma Lattes detém diversas unidades de informação que foram ou estão passando pelo processo de padronização. Conforme já mencionado, todo esse trabalho é criado e homologado pela Comunidade Conscientias, que, através de seus membros, apresenta e discute as propostas apresentadas para esse fim.

Atualmente, a Comunidade Conscientias é responsável pela padronização de seis unidades de informação, sendo que dentre esse grupo, duas já foram homologadas e passam por processo de acompanhamento. As demais estão em desenvolvimento. Entre as homologadas, a unidade de Currículo foi a primeira e, mais tarde, a de Grupo de Pesquisa passou pela homologação da Comunidade. Ambas as duas unidades possuem hoje disponíveis no *site* da comunidade as suas respectivas especificações em XML *Schema* e documentação. Além dessas, a Conscientias trabalha na padronização das unidades Projeto, Instituição, Produção C&T e Periódicos Científicos.

A seguir, são comentados os esforços realizados através da aplicação da metodologia desta dissertação para padronização de uma nova unidade de informação da Plataforma Lattes. Essa padronização será para o Programa de Fomento em Iniciação Científica do CNPq.

O trabalho para padronização dessa nova etapa foi realizado conforme a metodologia proposta nesta dissertação e será submetido, de acordo com o modelo de gestão da própria comunidade, para discussão entre seus membros visando obter a homologação da proposta de padronização.

4.4.1 IDENTIFICAÇÃO: PROCESSO DE DEFINIÇÃO DA UNIDADE DE INFORMAÇÃO PARA O PROGRAMA DE FOMENTO EM INICIAÇÃO CIENTÍFICA

A etapa de definição, aplicada para a unidade de informação do Programa de Fomento em Iniciação Científica (IC) permitiu a identificação de uma série de informações válidas e necessárias para a realização do trabalho de especificação.

Durante essa fase, foram realizadas as atividades de:

- ✍ análise do formulário eletrônico do Programa Institucional de Bolsas de Iniciação Científica do CNPq, também conhecido como PIBIC, disponível no endereço eletrônico do CNPq. Esta análise possibilitou: (a) obter o levantamento completo das informações solicitadas pelo CNPq para avaliação das propostas submetidas a esse programa de fomento pelas instituições; (b) fundamentar a estruturação das informações levantadas para serem dispostas em forma hierárquica; e (c) avaliar a obrigatoriedade de cada informação solicitada pelo sistema bem como a multiplicidade exigida ou permitida para determinadas informações;
- ✍ análise de outras unidades de informação padronizadas em XML da Plataforma Lattes, tais como Currículo e Grupo de Pesquisa, que possibilitaram a complementação na identificação das informações sobre instituições e áreas do conhecimento; e
- ✍ avaliação de uma proposta inicial implementada em XML *Schema* que foi enviada pelo CNPq ao Grupo Stela para análise.

Com base nesse trabalho, partiu-se para a etapa de especificação da unidade de informação em questão, que será comentada na seção seguinte.

4.4.2 ETAPA DE ESPECIFICAÇÃO DA UNIDADE DE INFORMAÇÃO PARA O PROGRAMA DE FOMENTO EM INICIAÇÃO CIENTÍFICA

O trabalho de especificação da unidade de informação para o Programa de Fomento em IC (comentado anteriormente) foi realizado com base nas informações levantadas na etapa de identificação.

Seguindo a etapa de “especificação” descrita na metodologia desta dissertação, inicialmente foi definido o conteúdo da unidade através da identificação e da descrição dos objetos. Porém, para o alcance disso, primeiramente foi criada uma lista dos possíveis objetos, que são os candidatos a objetos. A partir dessa lista, foram aplicadas as orientações da metodologia, para validação de cada candidato citado na lista para a obtenção dos objetos válidos da unidade que está sendo padronizada. Nisso, foi dada uma atenção especial para que esses objetos tivessem uma nomenclatura pertinente a fim de evitar (a) interpretações errôneas de nomes e também (b) casos de ambigüidade.

Depois disso, iniciou-se a criação do diagrama de classes UML a partir dos objetos válidos, com vistas a obter a representação desses em um modelo descritivo.

Para a criação do diagrama, foram-se realizadas as tarefas de:

- ✍ organização dos objetos em uma hierarquia de classes, visando encontrar supertipos e subtipos de objetos para aplicação de herança;
- ✍ definição dos relacionamentos, cardinalidade e restrições, que possibilitaram: (a) a hierarquização das classes iniciada a partir da classe principal. Nesse caso, a unidade de Fomento em IC do CNPq levou a sigla de seu programa de fomento conhecido como PIBIC; (b) a modelagem dos relacionamentos de composição e associação¹; e (c) a indicação da multiplicidade e de restrições nos relacionamentos modelados;
- ✍ adição dos atributos de cada classe, que envolveu a especificação do tipo apropriado, e da obrigatoriedade ou não obrigatoriedade para cada atributo adicionado. Além disso, para os atributos que possuíam valores enumerados, foram criados tipos específicos para eles, de acordo com a recomendação da metodologia. Esses tipos foram representados no diagrama de classes como uma nova classe, estereotipada por <<enumeration>>, e os valores permitidos nesses tipos foram definidos como atributos nas classes estereotipadas.

¹ O relacionamento de agregação não foi utilizado devido à unidade de informação ser representada em um único documento XML.

Com a realização dessas tarefas, a especificação para a unidade do Programa de Fomento em IC foi concluída e possibilitou o alcance da etapa de padronização, mais comentada na seção a seguir.

4.4.3 ETAPA DE PADRONIZAÇÃO DA UNIDADE DE INFORMAÇÃO PARA O PROGRAMA DE FOMENTO EM INICIAÇÃO CIENTÍFICA

A aplicação da etapa de padronização da metodologia para padronização da unidade de informação do Programa de Fomento em IC foi fundamentada na especificação criada para ela. O resultado consistiu essencialmente na representação do modelo de classes UML em um modelo de informação representado através da linguagem XML e especificado em XML *Schema*, e será finalizado com a submissão da proposta para a Comunidade Conscientias para homologação do trabalho realizado.

Durante a representação do modelo de classes UML para XML foram criados elementos e atributos XML para cada classe e atributo modelado, respectivamente. Os elementos criados para as classes que possuem outras classes relacionadas por tipo composição tiveram um subelemento para cada classe composta, ou seja, um elemento contido dentro de outro elemento.

A nomenclatura das *tags* XML foi estabelecida de acordo com a recomendação da metodologia que sugere que elas sejam especificadas todas em maiúscula, e para aquelas que possuem nomes compostos, que esses sejam separados por hífen.

Depois de definido o modelo XML, foi utilizada a linguagem XML *Schema* para especificação. Nessa especificação, as superclasses do diagrama de classes se tornaram tipos complexos no *Schema* e serviram para serem herdadas pelos tipos criados para as subclasses. Os atributos do diagrama UML que possuíam tipos de classe estereotipada por <<enumeration>> também tiveram sua representação apropriada em XML *Schema*.

A codificação em XML *Schema* proveniente da aplicação desta etapa para a unidade do Programa de Fomento em IC pode ser encontrada no Anexo I desta dissertação. Juntamente com a codificação em XML *Schema*, nesse anexo são

apresentados também os diagramas visuais que foram construídos através da ferramenta de desenvolvimento XML conhecida como *XML Spy*.

O procedimento para efetivação do trabalho realizado nesta etapa para a unidade do Programa de Fomento de IC é comentado na etapa de gestão, apresentada na seção a seguir.

4.4.4 GESTÃO: COMO SÃO GERIDAS AS UNIDADES DE INFORMAÇÃO PADRONIZADAS EM TERMOS DA PLATAFORMA LATTES

O processo de gestão para efetivação do trabalho de padronização realizado para a unidade do Programa de Fomento em IC se inicia com a submissão da proposta de padronização para análise e discussão entre os membros e a equipe técnica da Comunidade Conscientias, de acordo com os procedimentos no modelo de gestão de ontologias da Conscientias, também discutidos nesta metodologia.

De acordo com o trabalho que foi realizado para as unidades de informação padronizadas hoje pela Comunidade Conscientias, como Currículo e Grupo de Pesquisa, uma proposta depois de homologada passa a disponibilizar à comunidade acadêmica, através do *site* da Conscientias, a especificação em XML Schema e sua respectiva documentação.

Outra característica comum que a unidade padronizada do programa de fomento terá, caso seja aceita, será o controle de versões que será disponibilizado juntamente com a especificação no *site* da Comunidade, mesmo esta sendo a primeira versão do padrão.

Em se tratando da gestão dessa unidade quanto à ampliação de sua utilização em nível internacional, espera-se que, caso surja o interesse da comunidade internacional, assim como ocorreu com o padrão XML para o Currículo, essa possa ser útil.

No entanto, para um melhor aproveitamento tanto do padrão XML para o Programa de Fomento em IC quanto para o Currículo, considera-se que esses seriam mais bem entendidos caso estivessem no idioma do país que estiver utilizando.

Pensando nisso, é proposta para complementar a metodologia apresentada nesta dissertação uma sistemática para tradução das *tags* desses padrões, especificados em *XML Schema*.

A sistemática foi construída com base no trabalho de tradução de tags XML apresentado por Pacheco, Kern e Bermejo (2003), e consiste essencialmente em duas etapas: a primeira consiste em dispor para as agências dos países interessados a especificação do padrão em *XML Schema* no idioma deles, conforme ilustrado na Figura 36. A segunda etapa, apresentada graficamente na Figura 37, consiste em criar um mecanismo para que os sistemas que esperam as unidades de informação em XML em seu idioma original possam se comunicar com as unidades em XML em diferentes idiomas.

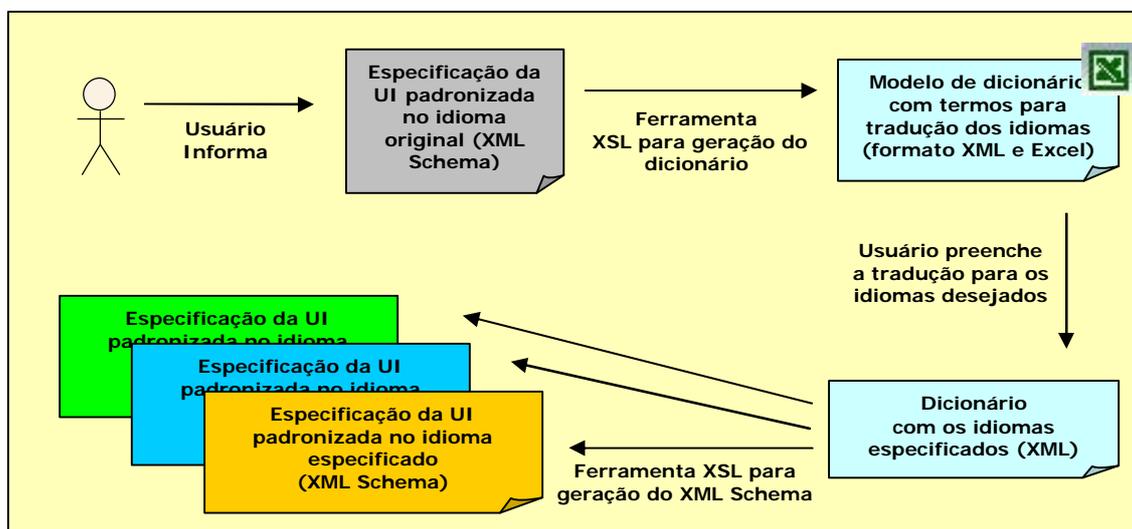


Figura 36 – Ilustração da etapa de tradução da especificação de uma unidade de informação padronizada para diferentes idiomas

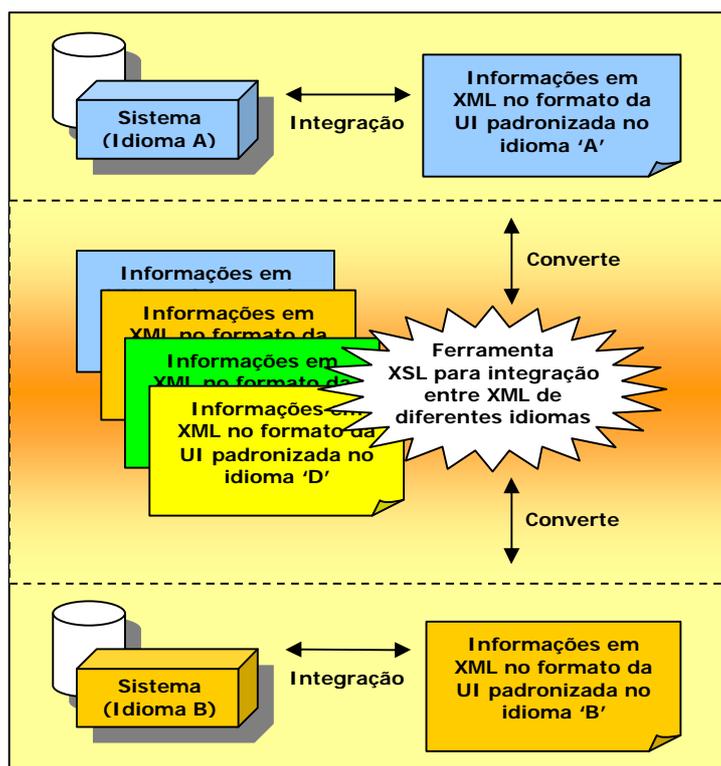


Figura 37 – Mecanismo de conversão para integração das informações entre os sistemas que mantenham compatibilidade com o padrão XML especificado em diferentes idiomas.

4.5 RECURSOS DA PLATAFORMA LATTES VIABILIZADOS PELA GESTÃO DE SUAS UNIDADES DE INFORMAÇÃO

Os comentários desta seção que se baseiam nas unidades de informação já passaram pelo processo de padronização, obtendo homologação por parte da Conscientias,. Hoje as unidades de informação são utilizadas pela comunidade acadêmica e pelos sistemas da Plataforma Lattes.

Entre os inúmeros recursos da Plataforma Lattes viabilizados com o trabalho de padronização findado pela gestão das unidades de informação destacam-se a abertura da base de currículos do CNPq para as instituições através da aplicação Lattes Extrator, que disponibiliza para as instituições as informações em XML, no padrão XML de currículos dos seus alunos, professores, pesquisadores e funcionários.

Além da recuperação de informações do CNPq por meio do Extrator, as instituições podem, através do XML de currículos, possibilitar a integração de

informações entre os seus sistemas curriculares proprietários e o sistema Lattes. Dessa forma, o usuário torna-se um grande beneficiário, pois não terá a necessidade de informar uma mesma informação em mais de um sistema.

Outro recurso viabilizado pelo padrão XML da Comunidade Conscientias, que está disponível em cerca de 12 instituições no Brasil e com previsão de adoções por parte de outras instituições ainda no ano de 2004, é a Plataforma Lattes Institucional, que é um conjunto de sistemas de informação e gestão curricular. Concebida com base nas informações em XML disponibilizadas pelo CNPq através do Lattes Extrator, a Plataforma Lattes Institucional possibilita aos gestores e interessados da instituição utilizadora consultarem uma série de indicadores sobre as informações curriculares dos usuários Lattes agregados à instituição.

Além desses, diversos outros recursos têm sido viabilizados com a gestão das unidades de informação da Plataforma Lattes. Os Sistemas InterLattes CV-Resume e CV-Perfil, ambos comentados anteriormente, também têm utilizado a unidade padronizada da Plataforma Lattes para extração de conhecimento e tratamento da qualidade da informação.

5 CONCLUSÕES E FUTUROS TRABALHOS

5.1 CONCLUSÕES

As unidades de informação padronizadas demonstram grande utilidade para a construção de sistemas de governo. A pesquisa realizada neste trabalho possibilitou constatar que mesmo as metodologias de desenvolvimento de governo eletrônico que reconhecem essa relevância não propõem uma metodologia integrada de identificação, especificação, padronização e gestão da camada de unidades de informação em plataformas de e-gov.

O trabalho realizado nesta dissertação permitiu a criação de uma metodologia para padronização de unidades de informação em governo eletrônico. Tal trabalho foi desenvolvido de forma a estar articulado com a arquitetura para construção de sistemas de governo Stela e-gov descrita no trabalho de Pacheco (2003).

Com base em estudos de arquiteturas de sistemas de governo eletrônico e em áreas de engenharia de software relacionadas com unidades de informação, a metodologia desenvolvida apresenta etapas de identificação, especificação, padronização e gestão de unidades de informação, que possibilitam a construção de unidades padronizadas a fim de permitir a viabilidade dessa camada em plataformas de e-gov.

Assim como a arquitetura em que essa metodologia está embasada, essa metodologia teve a sua aplicação na plataforma de sistemas do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, conhecida como Plataforma Lattes.

A aplicabilidade permitiu a definição de uma unidade de informação para o Programa de Fomento em IC padronizada, fornecendo uma especificação completa e explícita do seu domínio, conseguida através da concretização de cada etapa contemplada na metodologia, de forma a permitir a comunicação entre aplicações independentemente de tecnologias de sistemas, arquiteturas de informação e de domínios individuais de aplicação, e ainda possibilitar a descoberta e geração de conhecimento.

No caso da Plataforma Lattes, isso resulta no aumento do seu grau de interoperabilidade, promove a utilização e disseminação da Plataforma em âmbito nacional e internacional e, ainda, possibilita que novos recursos como sistemas, ferramentas e serviços para a área de Ciência e Tecnologia sejam adicionados a ela com base nos seus padrões.

As demais unidades de informação da Plataforma Lattes que estão sendo submetidas à Comunidade Conscientias já terão à disposição da comunidade científica a metodologia proposta, a fim de possibilitar a sua definição. Além disso, outras plataformas de sistemas de governo também poderão fazer uso deste trabalho considerando-se o alto grau de independência de sistemas e tecnologias que é apresentado na metodologia proposta.

5.2 TRABALHOS FUTUROS

Diante dos resultados obtidos envoltos à metodologia para padronização de unidades de informação, considera-se que um nicho de novas oportunidades para trabalhos futuros venha a surgir.

Ferramentas funcionais construídas com tecnologias na linha de software livre e XML poderão ser desenvolvidas com vistas a possibilitar a criação de uma versão inicial do XML *Schema* para a unidade de informação a partir do modelo de classes criado na etapa de especificação, no intuito de agilizar o trabalho realizado na etapa de padronização.

Esforços ainda podem ser realizados visando à criação de uma ferramenta que gere mecanismos e estruturas de classes para divulgações das informações contidas nas unidades padronizadas através de serviços web (*web services*).

REFERÊNCIAS BIBLIOGRÁFICAS

ACCENTURE – Management Consulting and Technology Services. **Country Reports.** The Government Executive Series. 2003c. Disponível em: <http://www.accenture.com/xdoc/en/newsroom/epresskit/egovernment/egov_epress_country.pdf>. Acesso em: 18 jan. 2004.

ACCENTURE – Management Consulting and Technology Services. **eGovernment Leadership: Engaging the Customer.** The Government Executive Series. 2003b. Disponível em: <http://www.accenture.com/xdoc/en/newsroom/epresskit/egovernment/egov_epress.pdf>. Acesso em: 18 jan. 2004.

ACCENTURE – Management Consulting and Technology Services. **Governments Closing Gap Between Political Rhetoric and eGovernment Reality,** 2003a. Disponível em: <http://www.accenture.com/xd/xd.asp?it=enWeb&xd=industries\government\gove_study.xml> Acesso em: 11 jan. 2004

ALDRICH, Duncan; BERTOT, John Carlo e MCCLURE, Charles R. **E-government: initiatives, developments, and issues.** *Government Information Quarterly*, Volume 19, Issue 4, pg. 349-355. Elsevier Science, 2002.

ALVAREZ, Rosío, **Confessions of an information worker: a critical analysis of information requirements discourse,** *Information and Organization*, Volume 12, Issue 2, April 2002, Pages 85-107.

ANDERSON, Richard *et al.* **Professional XML.** Rio de Janeiro: Editora Ciência Moderna, 2001.

BAX, Marcello Peixoto. **Introdução às linguagens de marcas.** *Revista Ciência da Informação*, Brasília, v. 30, n. 1, p. 32-38, jan./abr. 2001.

BECK, Kent; CUNNINGHAM, Ward. **A Laboratory For Teaching Object-Oriented Thinking.** OOPSLA'89 Conference Proceedings. Disponível em: <<http://c2.com/doc/oopsla89/paper.html>>. Acesso em: 31 mar. 2004. The special issue of SIGPLAN Notices, New Orleans, Louisiana, v. 24, n. 10, out. 1989.

BOURQUE, Pierre. *et al.* **Fundamental principles of software engineering – a journey.** *The Journal of Systems and Software*. Article in Press, Corrected Proof. © 2002 Elsevier Science.

BRUEGGE, Bernd e DUTOIT, Allen. H. **Object-Oriented Software Engineering: Using UML, Patterns, and Java.** Upper Saddle River – NJ: Prentice Hall, 2004.

CABINET OFFICE, UK GOVERNMENT. *Towards e – Government: E-government Metadata framework*, London, UK, May 2001. ISBN: 0711504083. Disponível em:

<<http://www.govtalk.gov.uk/documents/UK%20Metadata%20Framework%20v1%202001-05.pdf>>. Acesso em: 18 fev. 2004.

CARLSON, David. **Modelagem de aplicações XML com UML** – Aplicações Práticas de e-Business. São Paulo: Pearson Education, 2002.

CASTOLDI, André Vinícius. **Uma ontologia para enlaces de unidades de informação em plataformas de governo eletrônico**. 2003. 91 f. Dissertação (Mestrado em Engenharia de Produção) - Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2003.

CERI, Stefano; FRATERNALI, Piero e PARABOSCHI Stefano. **XML: Current Developments and Future Challenges for the Database Community**. *7th Int'l Conference on Extending Database Technology*, Konstanz (Germany), 2000. Disponível em: <<http://www.db.ucsd.edu/cse132B/XMLDB.pdf>>. Acesso em: 18 fev. 2004.

CERIF. **Common European Research Information Format**. Disponível em: <<http://www.cordis.lu/cerif/>>. Acesso em: 16 maio 2004.

CIÊNCIA, Jornal da. **Nova safra de instrumentos da Plataforma Lattes facilita a vida do pesquisador permitindo geração de resumos e práticas correções**. 2004b. *Orgão da Sociedade Brasileira para o Progresso da Ciência - SBPC*. Disponível em: <<http://www.jornaldaciencia.org.br/Detailhe.jsp?id=18476>>. Acesso em: 21 maio 2004.

CIÊNCIA, Jornal da. **Plataforma Lattes: o auto-retrato da C&T**. 2004a. *Orgão da Sociedade Brasileira para o Progresso da Ciência - SBPC*. Disponível em: <<http://www.jornaldaciencia.org.br/Detailhe.jsp?id=17132>>. Acesso em: 21 maio 2004.

CNPQ – Conselho Nacional de Desenvolvimento Científico e Tecnológico, 2003. Disponível em: <<http://lattes.cnpq.br>>. Acesso em: 18 fev. 2003.

CONSCIENTIAS - Comunidade para Ontologias em Ciência, Tecnologia e Informações de Aperfeiçoamento de Nível Superior. **Comunidade CONSCIENTIAS**. Disponível em: <<http://www.lattes.cnpq.br/lmpl>>. Acesso em: 18 fev. 2004.

COSTAKE, Nicolae. JENSEN, Finn H. **Towards an Architectural Framework of e-government Information Systems**. International Federation for Information Processing (IFIP): Knowledge Management in e-government - KMGov-2002. 2002. Disponível em: <<http://falcon.ifs.uni-linz.ac.at/KMGov2002/kmgov2002.zip>>. Acesso em: 13 jan. 2004.

DACONTA, Michael C., OBRST, Leo J., SMITH, Kevin T. **The Semantic Web: A guide to the Future of XML, and Knowledge Management**. Indianapolis: Wiley Publishing, Inc., 2003.

DAUM, Berthold. MERTEN, Udo. **Arquitetura de sistemas com XML: conteúdo, processo e apresentação**. Rio de Janeiro, Campus, 2002.

DÉCIO, Otávio C. **XML – Guia de consulta rápida**. São Paulo: Novatec Editora Ltda., 2000.

DEVADOSS, Paul Raj. PAN, Shan L. HUANG, Jimmy C. 2002. **Structurational analysis of e-government initiatives: a case study of SCO**. *Decision Support Systems* 34 (2002) 253– 269. Disponível em: <http://www.ekt.gr/links/egov_docs/egov_paperspp_ft/csa_recent2.pdf>. Acesso em: 15 jan. 2004.

DRIS. **Research Information Systems Worldwide**. Disponível em: <<http://www.niwi.knaw.nl/en/oi/dris/toon>>. Acesso em: 16 mai 2004.

EHRICH, Hans-Dieter. **Object Specification**. In: *H.-J. Kreowski et al, editor, IFIP 14.3 Volume on Foundations of System Specification*, Chapter 11. To appear in Springer LNCS, 1997.

ERGO. **European Research Gateways On-line**. About. Disponível em: <<http://www.cordis.lu/ergo/src/MainInformation.html>>. Acesso em: 16 maio 2004.

EUROCRIS – Current Research Information Systems. Disponível em: <<http://www.eurocris.org/>>. Acesso em: 11 maio 2004.

FEA, Working Group. 2002. **E-gov Enterprise Architecture Guidance - Common Reference Model**. Disponível em: <http://feapmo.gov/resources/E-gov_Guidance_Final_Draft_v2.0.pdf>. Acesso em: 19 jan. 2004.

FERNANDEZ, M., GOMEZ-PEREZ, A. and JURISTO, N. **METHONTOLOGY: From Ontological Art Towards Ontological Engineering**. *AAAI-97 Spring Symposium on Ontological Engineering*, Stanford University, March 24-26th, 1997.

FOWLER, Martin e SCOTT, Kendal. **UML Essencial: um breve guia para a linguagem-padrão de modelagem de objetos**. 2ª ed. Porto Alegre: Bookman, 2000.

GOMEZ-PEREZ, A., FERNANDEZ, M. and DE VICENTE, A.J. **Towards a Method to Conceptualize Domain Ontologies**. *ECAI-96 - Workshop on Ontological Engineering*, Budapest, 1996.

GOUSCOS, Dimitris. MENTZAS, Gregory e GEORGIADIS, Panagiotis. **Planning and Implementing e-government Service Delivery: Achievements and Learnings from On-line Taxation in Greece**. 2001. Apresentado em Workshop sobre e-government no 8th Panhellenic Conference on Informatics, Nicosia, Cyprus, de 08 a 10 de novembro de 2001. Disponível em: <<http://imu.iccs.ntua.gr/Papers/C51-EPY-e-gov.pdf>>. Acesso em: 14 jan. 2004.

GRUBER, Thomas R. **A translation Approach to portable ontology specifications**. *Knowledge Acquisition*. Vol. 5. 1993. 199-220. London: Academic Press Ltd. 1993.

GRÜNINGER, M. e FOX, M.S. **Methodology for the Design and Evaluation of Ontologies**. 1995. *IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing*, Montreal, August 19-20th. 1995.

GRÜNINGER, M. e FOX, M.S. **The Design and Evaluation of Ontologies for Enterprise Engineering**. 1994. *Workshop on Implemented Ontologies*, European Conference on Artificial Intelligence 1994, Amsterdam, NL. 1994a.

GRÜNINGER, M. e FOX, M.S. **The Role of Competency Questions in Enterprise Engineering**. 1994. *IFIP WG5.7 Workshop on Benchmarking - Theory and Practice*, Trondheim, Norway. 1994b.

HOESCHL, Hugo César (Org). **Introdução ao Governo eletrônico**. Florianópolis, 09 fev 2003. Disponível em: <<http://www.phoenix-library.org/download.php?ref=7022&format=LIT>>. Acesso em: 10 fev. 2003.

HOLMES, Douglas. **EGov eBusiness Strategies for Government**, London: Nicholas Brealey Publishing, 2001.

ITC, Internacional Teledemocracy Centre. **SmartGov - A Governmental Knowledge-based Platform for Public Sector Online Services**. 2004. Disponível em: <http://itc.napier.ac.uk/ITC_Home/Projects/Projects_SmartGov.asp>. Acesso em: 17 jan. 2004.

JACKSON, Michael. **The meaning of requirements**. *Annals of Software Engineering* 3. J.C. Baltzer AG, Science Publishers, 1997.

JACOBSON, Ivar. GRISS, Martin. JONSSON, Patrik. **Software Reuse: Architecture, Process and Organization for Business Success**. Addison Wesley, 1997.

JONES, D.M., BENCH-CAPON, T.J.M. e VISSER, P.R.S. **Methodologies for Ontology Development**. 1998. *Proceedings IT&KNOWS Conference of the 15th IFIP World Computer Congress*, Budapest, Hungary. Disponível em: <<http://gunther.smeal.psu.edu/6232.html>>. Acesso em: 10 maio 2003.

KAINDL, Hermann. **A practical approach to combining requirements definition and object-oriented analysis**. *Annals of Software Engineering* 3. J.C. Baltzer AG, Science Publishers, 1997.

KBSI. **IDEF5 Method Report**. *Information Integration for Concurrent Engineering - IICE*, KBSI Report, Texas, 1994. Disponível em: <<http://www.idef.com/Downloads/pdf/Idef5.pdf>>. Acesso em: 10 dez. 2003.

KHARE, Rohit e RIFKIN, Adam. **The origin of (document) species**. *Computer Networks and ISDN Systems*, v. 30, Issues 1-7, p. 389-397, April 1998.

MELLON, Carnegie. Software Engineering Institute – SEI. 2004. Disponível em: <<http://www.sei.cmu.edu>>. Acesso em: 2 fev. 2004.

MINISTÉRIO DO PLANEJAMENTO ORÇAMENTO E GESTÃO, REPÚBLICA FEDERATIVA DO BRASIL. **Governo eletrônico Brasileiro é Reconhecido Mundialmente**, notícia veiculada em 29 de novembro de 2002. Disponível em: <http://www.planejamento.gov.br/tecnologia_informacao/conteudo/noticias/governo_eletronico_bras_reconhecido_mundialmente.htm>. Acesso em: 11 fev. 2003.

MINISTÉRIO DO PLANEJAMENTO, Portal de Serviços e Informações de Governos. **Histórico**. 2003. Disponível em: <<http://www.governoeletronico.e.gov.br/historico.asp>>. Acesso em: 19 fev. 2003.

NATURE, Team. **Defining Visions In Context: Models, Processes And Tools For Requirements Engineering**. *Information Systems*, Vol. 21, No. 6, pp. 515-547. Elsevier Science, 1996.

NOIE, The National Office for the Information Economy. **Interoperability Technical Framework for the Australian Government**. Disponível em: <http://www.noie.gov.au/publications/NOIE/interop_frame/intro.htm>. Acesso em: 18 fev. 2004.

ONTOLOGY, Org. **Enabling Virtual Business**. Disponível em: <<http://www.ontology.org/>>. Acesso em: 10 jul. 2003.

PACHECO, Roberto C. S.; KERN, Vinícius. M. **Uma ontologia comum para a integração de bases de informações e conhecimento sobre ciência e tecnologia**. *Ciência da Informação*, v. 30, n. 3, p. 56-63, set/dez. 2001. Disponível em: <<http://www.ibict.br/cionline/300301/3030801.pdf>>. Acesso em: 11 fev. 2004.

PACHECO, Roberto Carlos dos Santos. **Uma metodologia de desenvolvimento de plataformas de governo para geração e divulgação de informações e de conhecimento**, 2003. Projeto de Pesquisa.

PACHECO, Roberto Carlos dos Santos; KERN, Vinícius Medina; BERMEJO, Paulo Henrique de Souza. **Interoperability and Information Integration in an Early Online Academic Digital Library of Theses and Dissertations: The Case of BTD**. *The International Information & Library Review*, Volume 35, Issues 2-4, Pages 319-333 (June-December 2003).

PACHECO, Roberto; KERN, Vinícius. **Arquitetura conceitual e resultados da integração de sistemas de informação e gestão da ciência e tecnologia**. 2003. *Revista de Ciência da Informação*, v.4, n.2. Disponível em: <http://www.dgz.org.br/abr03/Art_03.htm>. Acesso em: 7 maio 2004.

PAPAIIOANNOU, Vaios; THEODOULIDIS, Babis. **HERE: A web based environment for Requirements Engineering**. 1996. Disponível em: <<http://www.crim.co.umist.ac.uk/download/Files/Papers/ifip98.pdf>>. Acesso em: 9 fev. 2004.

PEREIRA, Luiz Carlos Bresser Gonçalves. **50 anos do CNPq**: contados pelos seus presidentes. Org: Shozo Montoyama. São Paulo: FAPESP, 2002.

PRÊMIO E-GOV, 2004. Disponível em: <<http://www.premio-e.gov.br>>. Acesso em: 25 maio 2004.

RELYEA, Harold C. **E-gov: Introduction and overview**. Library of Congress, Congressional Research Service, Washington, DC 20540-7470, USA. 2002. Disponível em: <http://www.ekt.gr/links/egov_docs/egov_papers_ft/GovInfQua_article2.pdf>. Acesso em: 13 jan. 2004.

RICHARDSON, Roberto Jarry. **Pesquisa Social: Métodos e Técnicas**, São Paulo: Editora Atlas S.A., 1999.

RIPPER FILHO, José Ellis. 1994. In: **Ciência e Tecnologia**: alicerces do desenvolvimento. 1. ed. São Paulo: Cobram.

SALM JÚNIOR, José Francisco; PACHECO, Roberto Carlos dos Santos; KERN, Vinícius Medina. **From digital repository to knowledge management system: the theses and dissertations database of PPGEP/UFSC**. In: 6TH INTERNATIONAL CONFERENCE ON CURRENT RESEARCH INFORMATION SYSTEMS, 2002, Kassel, Alemanha. CRIS 2002. 2002

SBC – Sociedade Brasileira de Computação. **Estudo revela melhora do Brasil em tecnologia**. 2003. Disponível em: <<http://www.sbc.org.br/notitia/servlet/newstorm.notitia.apresentacao.ServletDeNoticia?codigoDaNoticia=1274&dataDoJornal=atual>>. Acesso em: 19 fev. 2003.

SCACCHI, Walt. **Understanding the Social, Technological, and Public Policy Implications of Open Source Software Development**. 2002. Disponível em: <<http://www.ics.uci.edu/~wscacchi/Papers/New/OSS-Policy.pdf>>. Acesso em: 12 jan. 2004.

SCHUBERT, P., HAUSLER, U., **E-government meets E-business: a portal site for startup companies in Switzerland**, Proceedings of the 34th Hawaii International Conference on System Sciences, IEEE Computer Society, Los Alamitos, California, 2001, pp. 1700– 1708. Disponível em: <<http://csdl.computer.org/comp/proceedings/hicss/2001/0981/05/09815005.pdf>>. Acesso em: 16 jan. 2004.

SF/BNDES, Secretaria para Assuntos Fiscais do Banco Nacional de Desenvolvimento Econômico e Social. **E-gov no Brasil**. 2003. Disponível em <http://federativo.bndes.gov.br/bf_bancos/estudos/e0001364.pdf>. Acesso em: 11 fev. 2003.

SILVA, Douglas Marcos da. **UML – Guia de consulta rápida**. São Paulo: Novatec Editora Ltda., 2001.

SILVA, Edna Lúcia, MENEZES, Estera Muszkat. **Metodologia da pesquisa e elaboração de dissertação**, Florianópolis, 2000.

SMARTGOV, 2004. **A Governmental Knowledge-based Platform for Public Sector Online Services**. Disponível em: <<http://www.smartgov-project.org/>>. Acesso em: 17 jan. 2004.

STRACKE, Jürgen. **E-government: Supporting Knowledge and Information Flows with Supply Chain Management**. International Federation for Information Processing (IFIP): Knowledge Management in e-government - KMGov-2002. 2002. Disponível em: <<http://falcon.ifs.uni-linz.ac.at/KMGov2002/kmgov2002.zip>>. Acesso em: 12 jan. 2004.

TAKAHASHI, Tadao (Org). **Sociedade da Informação no Brasil**: Livro Verde, Brasília: Ministério da Ciência e Tecnologia, 2000. 195p.

TAMBOURIS, E. *et al.* **SmartGov: A Governmental Knowledge-based Platform for Public Sector Online Services**. International Federation for Information Processing (IFIP): Knowledge Management in e-government - KMGov-2002. 2002. Disponível em: <<http://falcon.ifs.uni-linz.ac.at/KMGov2002/kmgov2002.zip>>. Acesso em: 12 jan. 2004.

TAMBOURIS, E. GORILAS, S. BOUKIS, G. **Investigation of Electronic Government**. 2001. Disponível em: <http://www.euro-citi.org/publications/papers/eGOV_EuroCiti_EPY01.zip>. Acesso em: 11 jan. 2004.

TESCH JUNIOR, José Roberto. **XML Schema**. Florianópolis: Visual Books, 2002.

TRACZ, W., COGLIANESE, L., YOUNG, P. **A domain-specific software architecture engineering process outline**. *SIGSOFT Software Engineering Notes*, 18, 40-49, 1993. Disponível em: <<http://citeseer.ist.psu.edu/article/tracz93domainspecific.html>>. Acesso em: 1º out. 2003.

TURNER, T. **What is metadata?** *Kaleidoscope*, [S. l], p. 71, 10 Feb. 2002. Disponível em: <<http://www.library.cornell.edu/staffweb/Kaleidoscope/feb02.pdf>>. Acesso em: 12 fev. 2004.

UNHELKAR, Bhuvan. **Process Quality Assurance for UML-Based Projects**. Boston: Addison Wesley, 2002.

USCHOLD, M. and KING, M. **Towards A Methodology for Building Ontologies**. *IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing*, Montreal, Canada. 1995.

USCHOLD, M. **Building Ontologies: Towards A Unified Methodology**. *Proc. Expert Systems 96*, Cambridge, December 16-18th. 1996b.

USCHOLD, M. **Converting an Informal Ontology into Ontolingua: Some Experiences.** *ECAI-96 Workshop on Ontological Engineering*, Budapest, August 13th. 1996a.

W3C. **World Wide Web Consortium.** Disponível em: <<http://www.w3.org>> Acesso em: 7 jul. 2003.

WATSON, R.T., MUNDY, B.. **A strategic perspective of electronic democracy,** *Communications of the ACM*, vol. 44, Pag 27 a 30 , 2001, ISSN: 0001-0782. Disponível em: <<http://portal.acm.org/citation.cfm?doid=357489.357499>>. Acesso em: 16 jan. 2003.

WIMMER, Maria. TRAUNMÜLLER, Roland. LENK, Klaus. **Electronic Business Invading the Public Sector: Considerations on Change and Design.** 2001. *Proceedings of the 34th Hawaii International Conference on System Sciences – 2001.* Disponível em: <<http://csdl.computer.org/comp/proceedings/hicss/2001/0981/05/09815006.pdf>>. Acesso em: 12 jan. 2004.

ANEXO I – CODIFICAÇÃO EM XML SCHEMA PARA A UNIDADE PARA O PROGRAMA DE FOMENTO EM IC PROVENIENTE DA APLICAÇÃO DA ETAPA DE PADRONIZAÇÃO

Para melhor visualização do XML Schema criado para a unidade de informação para o Programa de Fomento em Iniciação Científica, proveniente da aplicação da etapa de padronização da metodologia proposta nesta dissertação, apresenta-se na Figura 38 uma imagem gerada a partir do editor de XML, chamado *XML Spy*, que permite visualizar a estrutura do XML Schema criado.

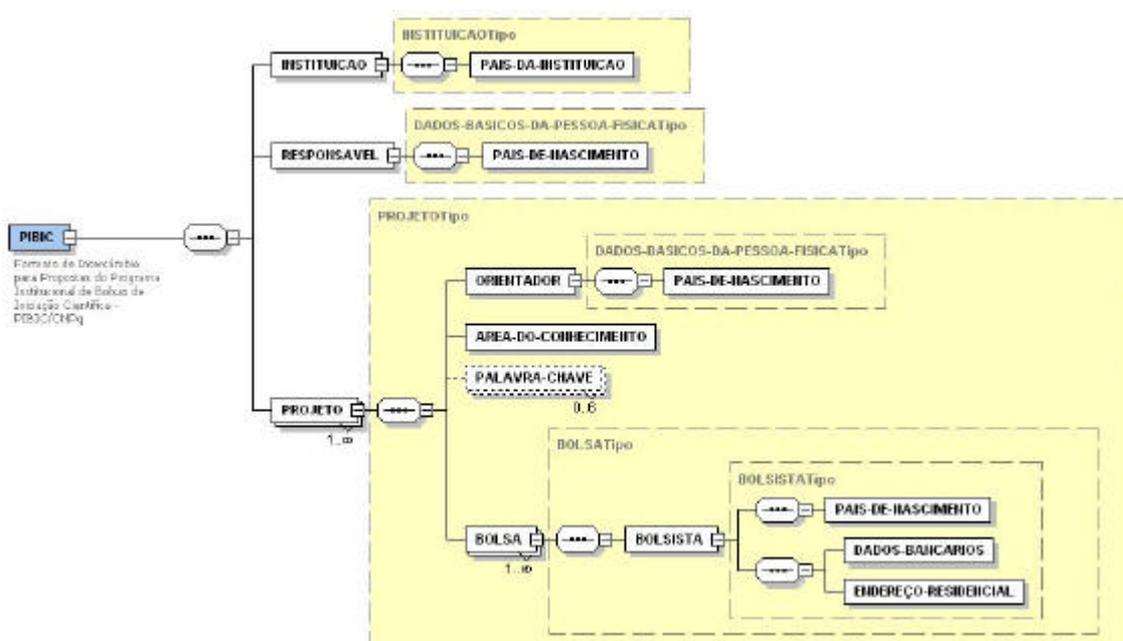


Figura 38 – Estrutura do XML Schema criado para a unidade de informação para o Programa de Fomento em Iniciação Científica

Na seqüência, é apresentado o código em XML Schema criado para essa unidade de informação.

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified"
attributeFormDefault="unqualified">
  <xs:complexType name="DADOS-BASICOS-DA-PESSOA-FISICATipo">
    <xs:annotation>
      <xs:documentation>Tipo que contém definição para os dados básicos da pessoa
fisica</xs:documentation>
    </xs:annotation>
    <xs:sequence>
      <xs:element name="PAIS-DE-NASCIMENTO" type="PAISTipo"/>
    </xs:sequence>
    <xs:attribute name="CPF" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:length value="11"/>
          <xs:pattern value="[0-9]{11}(EX[0-9]{8})"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="NOME-COMPLETO" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:maxLength value="60"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="FORMATO-DATA-DE-NASCIMENTO" type="xs:NMTOKEN" use="optional"
fixed="DDMMAAAA"/>
    <xs:attribute name="DATA-DE-NASCIMENTO" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:pattern value="\d{8}"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="NACIONALIDADE" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="BRASILEIRA"/>
          <xs:enumeration value="ESTRANGEIRA"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
  </xs:complexType>
  <xs:complexType name="PAISTipo">
    <xs:annotation>
      <xs:documentation>Tipo que contém definição para o país</xs:documentation>
    </xs:annotation>
    <xs:attribute name="SIGLA-DO-PAIS" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:length value="3"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="NOME-DO-PAIS" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:maxLength value="20"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
  </xs:complexType>
  <xs:complexType name="INSTITUICAOTipo">
    <xs:annotation>
      <xs:documentation>Tipo para o elemento Instituição</xs:documentation>
    </xs:annotation>
    <xs:sequence>

```

```

    <xs:element name="PAIS-DA-INSTITUICAO" type="PAISTipo"/>
  </xs:sequence>
  <xs:attribute name="NOME-DA-INSTITUICAO" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:maxLength value="75"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
  <xs:attribute name="SIGLA-DA-INSTITUICAO" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:maxLength value="16"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
  <xs:attribute name="UF-DA-INSTITUICAO" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:length value="2"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
  <xs:attribute name="FLAG-AGENCIA-DE-FOMENTO" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:NMTOKEN">
        <xs:enumeration value="SIM"/>
        <xs:enumeration value="NAO"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
  <xs:attribute name="FLAG-INSTITUICAO-DE-ENSINO" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:NMTOKEN">
        <xs:enumeration value="SIM"/>
        <xs:enumeration value="NAO"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
</xs:complexType>
<xs:complexType name="PROJETOTipo">
  <xs:sequence>
    <xs:element name="ORIENTADOR" type="DADOS-BASICOS-DA-PESSOA-FISICATipo"/>
    <xs:element name="AREA-DO-CONHECIMENTO" type="AREA-DO-CONHECIMENTOTipo"/>
    <xs:element name="PALAVRA-CHAVE" type="PALAVRA-CHAVETipo" minOccurs="0"
maxOccurs="6"/>
    <xs:element name="BOLSA" type="BOLSATipo" maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="TITULO-DO-PROJETO" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:maxLength value="255"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
</xs:complexType>
<xs:complexType name="AREA-DO-CONHECIMENTOTipo">
  <xs:annotation>
    <xs:documentation>Tipo para o elemento de Área do Conhecimento</xs:documentation>
  </xs:annotation>
  <xs:attribute name="SEQUENCIA-AREA-DO-CONHECIMENTO" type="xs:int" use="required"/>
  <xs:attribute name="NOME-GRANDE-AREA-DO-CONHECIMENTO" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:NMTOKEN">
        <xs:enumeration value="OUTROS"/>
        <xs:enumeration value="LINGUISTICA-LETRAS-E-ARTES"/>
        <xs:enumeration value="CIENCIAS-HUMANAS"/>
        <xs:enumeration value="CIENCIAS-SOCIAIS-APLICADAS"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>

```

```

        <xs:enumeration value="CIENCIAS-AGRARIAS"/>
        <xs:enumeration value="CIENCIAS-DA-SAUDE"/>
        <xs:enumeration value="ENGENHARIAS"/>
        <xs:enumeration value="CIENCIAS-BIOLÓGICAS"/>
        <xs:enumeration value="CIENCIAS-EXATAS-E-DA-TERRA"/>
    </xs:restriction>
</xs:simpleType>
</xs:attribute>
<xs:attribute name="NOME-DA-AREA-DO-CONHECIMENTO" use="optional">
    <xs:simpleType>
        <xs:restriction base="xs:string">
            <xs:maxLength value="85"/>
        </xs:restriction>
    </xs:simpleType>
</xs:attribute>
<xs:attribute name="NOME-DA-SUB-AREA-DO-CONHECIMENTO" use="optional">
    <xs:simpleType>
        <xs:restriction base="xs:string">
            <xs:maxLength value="85"/>
        </xs:restriction>
    </xs:simpleType>
</xs:attribute>
<xs:attribute name="NOME-DA-ESPECIALIDADE" use="optional">
    <xs:simpleType>
        <xs:restriction base="xs:string">
            <xs:maxLength value="85"/>
        </xs:restriction>
    </xs:simpleType>
</xs:attribute>
</xs:complexType>
<xs:complexType name="PALAVRA-CHAVETipo">
    <xs:annotation>
        <xs:documentation>Tipo para o elemento de Palavra-chave</xs:documentation>
    </xs:annotation>
    <xs:attribute name="SEQUENCIA" type="xs:int" use="required"/>
    <xs:attribute name="DESCRICAO" use="required">
        <xs:simpleType>
            <xs:restriction base="xs:string">
                <xs:maxLength value="50"/>
            </xs:restriction>
        </xs:simpleType>
    </xs:attribute>
</xs:complexType>
<xs:complexType name="BOLSATipo">
    <xs:annotation>
        <xs:documentation>Tipo para o elemento bolsa</xs:documentation>
    </xs:annotation>
    <xs:sequence>
        <xs:element name="BOLSISTA" type="BOLSISTATipo"/>
    </xs:sequence>
    <xs:attribute name="MES-DE-INICIO" use="required">
        <xs:simpleType>
            <xs:restriction base="xs:gMonth">
                <xs:pattern value="[0-9]{2}"/>
            </xs:restriction>
        </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="ANO-DE-INICIO" type="xs:gYear" use="required"/>
    <xs:attribute name="MES-DE-FIM" use="required">
        <xs:simpleType>
            <xs:restriction base="xs:gMonth">
                <xs:pattern value="[0-9]{2}"/>
            </xs:restriction>
        </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="ANO-DE-FIM" type="xs:gYear" use="required"/>
</xs:complexType>
<xs:complexType name="BOLSISTATipo">

```

```

<xs:annotation>
  <xs:documentation>Tipo para o elemento Bolsista</xs:documentation>
</xs:annotation>
<xs:complexContent>
  <xs:extension base="DADOS-BASICOS-DA-PESSOA-FISICATipo">
    <xs:sequence>
      <xs:element name="DADOS-BANCARIOS" type="DADOS-BANCARIOTipo"/>
      <xs:element name="ENDEREÇO-RESIDENCIAL" type="ENDERECO-RESIDENCIALTipo"/>
    </xs:sequence>
    <xs:attribute name="NUMERO-DA-CARTEIRA-DE-IDENTIDADE" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:maxLength value="15"/>
          <xs:pattern value="([A-Z][0-9])([0-9]{1,13})([A-Z][0-9])/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="ORGAO-EMISSOR-DA-CARTEIRA-DE-IDENTIDADE" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:maxLength value="8"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="UF-DE-EMISSAO-DA-CARTEIRA-DE-IDENTIDADE" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:length value="2"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="FORMATO-DATA-DE-EMISSAO-DA-CARTEIRA-DE-IDENTIDADE"
type="xs:NMTOKEN" use="optional" fixed="DDMMAAAA"/>
    <xs:attribute name="DATA-DE-EMISSAO-DA-CARTEIRA-DE-IDENTIDADE" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:pattern value="\d{8}"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="SEXO" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="FEMININO"/>
          <xs:enumeration value="MASCULINO"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="NUMERO-DO-PASSAPORTE" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:maxLength value="20"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="FORMATO-DATA-DE-VALIDADE-DO-VISTO-DO-PASSAPORTE"
type="xs:NMTOKEN" use="optional" fixed="DDMMAAAA"/>
    <xs:attribute name="DATA-DE-VALIDADE-DO-VISTO-DO-PASSAPORTE" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:pattern value="\d{8}"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="E-MAIL" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:maxLength value="70"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
  </xs:extension>
</xs:complexContent>

```

```

        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="SITUACAO-DO-BOLSISTA" use="required">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="BOLSISTA-NOVO"/>
          <xs:enumeration value="RENOVACAO-DE-BOLSA"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
  </xs:extension>
</xs:complexContent>
</xs:complexType>
<xs:complexType name="DADOS-BANCARIOSTipo">
  <xs:annotation>
    <xs:documentation>Tipo para o elemento dados bancários</xs:documentation>
  </xs:annotation>
  <xs:attribute name="CODIGO-DO-BANCO" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:int">
        <xs:length value="3"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
  <xs:attribute name="NOME-DO-BANCO" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:maxLength value="30"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
  <xs:attribute name="NOME-DA-AGENCIA" use="optional">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:maxLength value="20"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
  <xs:attribute name="NUMERO-DA-AGENCIA" type="xs:int" use="required"/>
  <xs:attribute name="NUMERO-DA-CONTA" type="xs:string" use="required"/>
</xs:complexType>
<xs:complexType name="ENDERECO-RESIDENCIALTipo">
  <xs:annotation>
    <xs:documentation>Tipo para o elemento endereço residencial</xs:documentation>
  </xs:annotation>
  <xs:attribute name="LOGRADOURO" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:maxLength value="180"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
  <xs:attribute name="BAIRRO" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:maxLength value="20"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
  <xs:attribute name="CEP" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:length value="9"/>
        <xs:pattern value="\d\d\d\d\d-\d\d\d"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>

```

```

<xs:attribute name="CIDADE" use="required">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:maxLength value="45"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
<xs:attribute name="UF" use="required">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:length value="2"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
<xs:attribute name="DDD" use="optional">
  <xs:simpleType>
    <xs:restriction base="xs:int">
      <xs:length value="2"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
<xs:attribute name="TELEFONE" use="optional">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:maxLength value="10"/>
      <xs:pattern value="\d{1,10}"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
<xs:attribute name="RAMAL" use="optional">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:maxLength value="5"/>
      <xs:pattern value="\d{1,5}"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
<xs:attribute name="FAX" use="optional">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:maxLength value="10"/>
      <xs:pattern value="\d{1,10}"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
</xs:complexType>
<!-- Elemento principal da Unidade de Informação do PIBIC-->
<xs:element name="PIBIC">
  <xs:annotation>
    <xs:documentation>Programa de Fomento em Iniciação Científica</xs:documentation>
  </xs:annotation>
  <xs:complexType>
    <xs:sequence>
      <xs:element name="INSTITUCAO" type="INSTITUCAOTipo"/>
      <xs:element name="RESPONSAVEL" type="DADOS-BASICOS-DA-PESSOA-FISICATipo"/>
      <xs:element name="PROJETO" type="PROJETOTipo" maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:attribute name="NUMERO-DO-PROCESSO-INSTITUCIONAL">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:pattern value="[0-9]{11}"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
  </xs:complexType>
</xs:element>
</xs:schema>

```