

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Anrafel Fernandes Pereira

***Collaborative PL-Science: Utilizando Elementos de Colaboração em
uma Linha de Produtos de Software Científico***

Juiz de Fora

2014

Collaborative PL-Science: Utilizando Elementos De Colaboração Em Uma Linha De Produtos De Software Científico

Anrafel Fernandes Pereira

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Juiz de Fora, como parte dos requisitos necessários à obtenção do grau de Mestre em Ciência da Computação.

Orientador: José Maria Nazar David

Coorientadora: Regina Maria Maciel Braga Villela

Juiz de Fora

2014

Ficha catalográfica elaborada através do Programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Pereira, Anrafel Fernandes.

Collaborative PL-Science: Utilizando Elementos de Colaboração em uma Linha de Produtos de Software Científico / Anrafel Fernandes Pereira. -- 2014.

119 f. : il.

Orientador: José Maria Nazar David

Coorientadora: Regina Maria Maciel Braga

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação, 2014.

1. Linha de Produto de Software. 2. e-Science. 3. Workflow. 4. Colaboração. 5. Ontologia. I. David, José Maria Nazar, orient. II. Braga, Regina Maria Maciel, coorient. III. Título.

Anrafel Fernandes Pereira

***Collaborative PL-Science: Utilizando Elementos de Colaboração em
uma Linha de Produtos de Software Científico***

Dissertação apresentada ao Programa de Pós-graduação
em Ciência da Computação da Universidade Federal de
Juiz de Fora como requisito parcial à obtenção do grau
de Mestre.

Aprovada em 11 de Julho de 2014.

BANCA EXAMINADORA

Prof. D.Sc. José Maria Nazar David – Orientador
Universidade Federal de Juiz de Fora

Profa. D.Sc. Regina Maria Maciel Braga Villela – Coorientadora
Universidade Federal de Juiz de Fora

Profa. D.Sc. Fernanda Cláudia Alves Campos
Universidade Federal de Juiz de Fora

Prof. D.Sc. André Santanchè
Universidade Estadual de Campinas

"Ao contrário do que se pensa,
saber depender da ajuda dos outros não é sinal de fraqueza,
mas da mais nobre inteligência e humildade."

Daniel Godri JR

AGRADECIMENTOS

Primeiramente agradeço a Deus, por ter me dado mais essa oportunidade e vencer mais uma etapa das muitas que desejo passar ao longo da minha vida.

Aos meus pais Walmir e Cátia pela força e coragem que depositaram em mim, por estarem sempre ao meu lado firmando os meus passos e fazendo desses passos, passos vencedores.

Aos meus irmãos Walmir e Carine que sempre me incentivaram para continuar na busca dos meus sonhos e aos meus familiares que fizeram parte dessa conquista comigo, em especial, minha avó Maria da Glória, meus primos e amigos Paloma, Sandro, Manuela e Patrique.

Agradeço também, de forma especial, aos meus amigos José Moutinho e Aline Miranda pelo incentivo e apoio para encarar mais esta etapa.

A todos os professores do programa, que não citarei nomes, pois, todos, cada um com o seu jeito, contribuíram com algo para que eu pudesse chegar até aqui.

Entretanto, eu não poderia deixar de agradecer aos meus primeiros professores nessa jornada do mestrado: o Prof. Dr. Guilherme Pinto e Prof.^a Dr. Fernanda Campos por acreditarem na minha capacidade de realizar esta trajetória e também pelo conhecimento e atenção que me ofereceram desde o início do curso até este momento.

Aos meus orientadores, o Prof. Dr. José Maria N. David e a Prof.^a Dra. Regina Braga, pela atenção, dedicação, apoio, incentivo e colaboração durante a realização desta dissertação. Obrigado pelos comentários, pelas críticas sempre construtivas e pela paciência comigo!

Aos membros da Banca Examinadora pela disponibilidade e pelo trabalho de avaliação.

A todos os amigos que fiz ao longo desta jornada. Em especial, agradeço aos amigos do NEnC, principalmente ao Magnus, Jonathan e Raphael com os quais tive a oportunidade de trabalhar mais tempo. Com certeza foi um período de muito aprendizado e grandes amizades que levarei para o resto da minha vida. Vocês fazem parte da minha história meus grandes amigos!

Agradeço também aos meus colegas de turma, e ainda, de maneira especial a Gláucia, secretária do Programa, pelo carinho e atenção desde o início de toda esta trajetória. Foi muito bom encontrar e conviver com cada um de vocês nesta etapa da minha vida. Encontramo-nos por aí!

Outras pessoas que eu não poderia deixar de agradecer são os meus amigos Flávio Duarte, Ana Cláudia, Carlos Henrique e a todos os meus colegas de trabalho no Centro Vocacional Tecnológico – CVT Três Rios. Vocês me ajudaram muito para que eu pudesse chegar até aqui.

À Universidade Federal de Juiz de Fora e à CAPES, pelo apoio financeiro através da bolsa.

Com todos vocês, eu aprendi muito, e com certeza deixei algo de mim também. Este trabalho é parte das nossas histórias!

RESUMO

A necessidade de colaboração no domínio científico vem sendo discutida em alguns trabalhos. A sua relevância no contexto desta dissertação é justificada por problemas clássicos neste domínio, como a falta de apoio para a composição de *workflows* científicos, dificuldade para reutilização de aplicações científicas, dificuldade de cooperação e comunicação entre as equipes de cientistas geograficamente distribuídas, entre outros. Abordagens como Linha de Produtos de Software - LPS têm sido empregadas para apoiar os cientistas na concepção de *workflows* científicos. Entretanto muitas informações relevantes sobre estas aplicações são perdidas ou mesmo não fornecidas pelos cientistas. Esta dissertação apresenta uma abordagem denominada Collaborative PL-Science, a qual é uma extensão de uma abordagem denominada PL-Science. Na abordagem PL-Science os cientistas possuíam apenas os artefatos persistidos no núcleo da LPS como componentes de trabalho. Com isso, o cientista desenvolvia um *workflow* científico baseado apenas em seu conhecimento sobre o domínio. Nenhum histórico, ou "*rationale*", era gerado neste ambiente, ficando todo o conhecimento sobre as funcionalidades dos artefatos, as experiências do cientista e as tomadas de decisões por conta do usuário. Como proposta de solução para estes problemas, a Collaborative PL-Science utiliza elementos de colaboração, tais como informações de percepção, contexto e um mecanismo de suporte à comunicação, em uma Linha de Produtos de Software Científico - LPSC. Com isso, espera-se gerar oportunidades de interação entre os pesquisadores e contextualizá-los em sua atividade de concepção de *workflows* científicos.

Palavras-chave: Linha de Produtos de Software, *e-Science*, Percepção, Contexto, Ontologia, *Workflow* Científico.

ABSTRACT

The need for collaboration in the scientific field has been discussed in some papers. Its relevance in the context of this thesis is justified by classical problems in this area, such as lack of support for the composition of scientific workflows, difficult reuse to scientific applications, difficulty of cooperation and communication among geographically distributed teams of scientists, among others. Approaches to Software Product Line - SPL have been used to support scientists in the design of scientific workflows. However many relevant information about scientific applications are lost or even not provided by scientists. This paper presents an approach named Collaborative PL-Science, which is an extension of an approach called PL-Science. In the PL-Science approach scientists had only the artifacts persisted in the core of the SPL as components of work. No historical or "rationale" was generated in this environment. Therefore all the knowledge about the functionalities of the artifacts, the experiences of the scientist and the decisions made are attributed to the user. As a proposed solution to these problems, the Collaborative PL-Science uses collaboration elements, such as awareness and context information and a mechanism to support communication in a Scientific Software Product Line - SSPL. Thus, it is expected to generate opportunities for interaction between researchers and contextualize them in their activity of designing scientific workflows.

Keywords: Software Product Line, e-Science, Awareness, Context, Ontology, Scientific Workflow

LISTA DE FIGURAS

2.1	Atividades essenciais de uma LPS (CLEMENTS & NORTHROP, 2002)	22
2.2	Visão geral da Abordagem PL-Science (COSTA et al., 2013)	40
2.3	Núcleo de Artefatos da Abordagem PL-Science.	41
3.4	<i>Workflow</i> do produto <i>Pipeline PhredPhrap</i>	49
3.5	Visão Geral da Collaborative PL-Science - (PEREIRA et al., 2013).....	50
3.6	Visão da Arquitetura da Abordagem Collaborative PL-Science	60
3.7	Processos envolvidos no Gerente de Artefatos	61
3.8	Interface de visualização do Núcleo de Artefatos.....	63
3.9	Projeto Lógico do Banco de Dados	64
3.10	Projeto Físico do Banco de Dados	66
3.11	Tipos de atividades modeladas na ontologia.....	67
3.12	Tipos de Artefatos modelados na ontologia.....	67
3.13	Cientistas e Comunidades modelados na ontologia.....	68
3.14	Interações modeladas na ontologia.....	70
3.15	Produtos modelados na ontologia	70
3.16	Collaborative PL-Science Ontology - Ferramenta Protegé	71
3.17	Propriedades criadas no Protegé	72
3.18	Algumas inferências sobre o indivíduo AC_Develop_a_Product01	72
3.19	Exemplos de <i>Properties Chains</i> criadas.....	73
4.20	<i>Workflow</i> para Sequenciamento e Alinhamento Genético (COSTA, 2013).....	81
4.21	Interface de desenvolvimento de novos <i>workflows</i> científicos	82
4.22	Modelo do <i>Workflow Pipeline PhredPhrap</i> (ARBEX, 2009).....	85

4.23	Desenvolvimento do Produto <i>Pipeline PhredPhrap</i>	86
4.24	Aplicação para Alinhamento Múltiplo de Sequências (COSTA, 2013).....	89
4.25	Informações sobre o produto <i>Pipeline PhredPhrap</i>	93
4.26	Informações sobre a atividade que gerou o <i>Pipeline PhredPhrap</i>	94

LISTA DE TABELAS

2.1 Tipos de Percepção (DRURY & WILLIAMS, 2002)	30
2.2 Abordagens de Representação de Contexto (VIEIRA, 2006)	32
2.3 Comparações dos Trabalhos Relacionados.....	46
3.4 Localização dos Artefatos utilizados pela Collaborative PL-Science.....	68
4.5 Artefatos utilizados na Avaliação da Solução.....	80

LISTA DE QUADROS

3.1 Modelo do Projeto Lógico do Banco de Dados.....	65
---	----

SUMÁRIO

1. INTRODUÇÃO	17
1.1 MOTIVAÇÃO	17
1.2 PROBLEMA	18
1.3 HIPÓTESE	19
1.4 OBJETIVOS	19
1.5 ESTRUTURA DA DISSERTAÇÃO	20
2. REFERENCIAL TEÓRICO	21
2.1 LINHA DE PRODUTOS DE SOFTWARE	21
2.1.1 Desenvolvimento do Núcleo de Artefatos	23
2.1.2 Desenvolvimento dos Produtos de Software	25
2.1.3 Gerenciamento da Linha de Produtos	25
2.1.4 Benefícios da Abordagem de LPS	25
2.2 PERCEPÇÃO E CONTEXTO	26
2.2.1 Percepção	26
2.2.2 Contexto	28
2.2.3 Tipos de Percepção e Formas de Representação de Contexto	29
2.3 USO DE ONTOLOGIAS PARA REPRESENTAÇÃO DE CONTEXTO	34
2.4 COLABORAÇÃO	36
2.4.1 Colaboração em e-Science	37
2.5 ABORDAGEM PL-SCIENCE	39
2.6 TRABALHOS RELACIONADOS	42

2.7 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO	46
3. A ABORDAGEM COLLABORATIVE PL-SCIENCE	48
3.1 VISÃO GERAL DA ABORDAGEM	48
3.1.1 Elementos de Colaboração.....	52
3.1.2 Memória de Grupo.....	54
3.2 ESPECIFICAÇÃO DE REQUISITOS	56
3.2.1 Requisitos Não Funcionais.....	58
3.3 ARQUITETURA DA SOLUÇÃO	59
3.3.1 Processos Adicionados pela Collaborative PL-Science.....	62
3.3.2 Memória de Grupo.....	64
3.3.2.1 Banco de Dados.....	64
3.3.2.2 Ontologia Collaborative PL-Science.....	66
3.4 IMPLEMENTAÇÃO	73
3.4.1 Desenvolvimento da Memória de Grupo.....	73
3.4.2 Aplicação Collaborative PL-Science.....	74
3.4.3 Testes Funcionais	74
3.5 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO	75
4. AVALIAÇÃO DA PROPOSTA	77
4.1 DEFINIÇÃO E PLANEJAMENTO DO MÉTODO DE PESQUISA	77
4.2 PROVAS DE CONCEITOS	79
4.2.1 Cenário e Caracterização do Grupo	79
4.2.2 Primeira PoC: <i>Workflow</i> para Sequenciamento/Alinhamento Genético	80
4.2.2.1 Evidências Observadas na Primeira PoC	84
4.2.3 Segunda PoC: Workflow Pipeline PhredPhrap	85

4.2.3.1 Evidências Observadas na Segunda PoC.....	87
4.2.4 Terceira PoC: <i>Workflow</i> para Alinhamento Múltiplo de Sequências	88
4.2.4.1 Evidências Observadas na Terceira PoC	91
4.2.5 Quarta PoC: Realização de uma Consulta Ontológica na Aplicação	92
4.2.5.1 Evidências Observadas na Quarta PoC	95
4.3 ESTUDO DE CASO	96
4.3.1 Planejamento do Estudo de Caso	96
4.3.2 Objetivo do Estudo	97
4.3.3 Caracterização do Grupo.....	97
4.3.4 Cenário do Estudo	97
4.3.5 Fontes de Evidências	98
4.3.6 Resultados do Estudo.....	99
4.4 ANÁLISES DOS ESTUDOS	100
4.5 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO	102
5. CONCLUSÕES	104
5.1 CONTRIBUIÇÕES	104
5.2 LIMITAÇÕES	105
5.3 TRABALHOS FUTUROS	107

1 INTRODUÇÃO

1.1 MOTIVAÇÃO

Abordagens computacionais têm sido utilizadas nos mais diversos cenários, incluindo o cenário de experimentos científicos. No domínio de *e-Science* estes recursos são fundamentais para o sucesso na realização de pesquisas científicas. O uso de *workflows* científicos tem sido crescente neste cenário (SEGAL & MORRIS, 2008). Um workflow científico é um mecanismo utilizado como forma de representar e/ou executar atividades ou experimentos científicos (COSTA, 2013). Entretanto, a concepção e representação desta tecnologia não são triviais, e a utilização de abordagens *ad-hoc* criam várias barreiras. Como exemplo temos, a falta de apoio para a reutilização de *workflows* concebidos por outros cientistas e o controle sobre a evolução das diferentes versões destes produtos (MATTOSO et al., 2009).

A abordagem de Linha de Produtos de Software - LPS vem sendo utilizada visando auxiliar cientistas na solução de problemas no domínio de *e-Science* (REMMEL et al., 2011) (COSTA et al., 2013). O propósito da sua utilização neste cenário tem sido aumentar a qualidade e a produtividade em experimentos e, ao mesmo tempo, reduzir os custos envolvidos através do reúso de artefatos. Nestes trabalhos, alguns autores buscam o apoio nos modelos de *features* associados a ontologias, para oferecer mais semântica para o domínio [CZARNECKI et al. (2006), JOHANSEN et al. (2010), FILHO et al. (2012), COSTA et al. (2013)]. Entretanto, muitas informações relevantes para um experimento são perdidas, ou mesmo não são fornecidas pelos participantes, o que pode dificultar a execução da pesquisa. Com isso, apesar do objetivo de facilitar o desenvolvimento de experimentos científicos, a utilização de um artefato em uma LPS continua sendo feita baseada na experiência individual do pesquisador.

Uma característica que tem sido fundamental na ciência nos dias de hoje é a colaboração [CHIN & LANSING (2004), OLSON, ZIMMERMAN & BOS (2008), MOREIRA, VIEIRA & DEL ARCO (2012)]. Este assunto não é algo recente no domínio científico. WULF (1989) já evidenciava alguns problemas clássicos desse cenário, como por exemplo a falta de serviços oferecidos para

promover a colaboração entre os pesquisadores. Assim, associado a abordagem de LPS, a utilização de elementos que gerem oportunidades de colaboração no domínio científico pode ser útil para apoiar a coleta e a disponibilização de decisões e resultados obtidos neste domínio. Um exemplo é a criação de um histórico das atividades que são realizadas pelos cientistas. Este histórico pode ser entendido como sendo os passos realizados pelos pesquisadores durante a realização de uma dada atividade, bem como as decisões que foram tomadas e as anotações feitas sobre os artefatos ou *features*. Com isso, forma-se uma memória de grupo, que compõe o conhecimento sobre o domínio. A memória de grupo se torna útil nesse cenário, por permitir que cientistas possam reutilizar em experimentos, decisões e experiências já vividas por outros pesquisadores.

Considerando este contexto, nesta dissertação é apresentada uma arquitetura desenvolvida para apoiar os cientistas durante a atividade de composição de *workflows* científicos, que neste trabalho pode ser entendido com o produto final da Linha de Produto de Software Científico. Esta arquitetura promove a interação entre os pesquisadores e os contextualiza acerca do ambiente compartilhado de trabalho no qual estão inseridos. Para isso, a arquitetura captura o histórico das atividades realizadas pelos cientistas, possibilitando que consultas e novas descobertas sejam feitas a partir desses dados. Com isso, possibilita ainda que informações de percepção e contexto possam ser geradas para auxiliá-los em suas atividades, gerando ainda a oportunidade de interação entre eles. Para esta finalidade, a abordagem PL-Science, proposta por COSTA et al. (2013), foi estendida nesta pesquisa. A abordagem apresentada nesta dissertação é denominada Collaborative PL-Science, e é descrita no decorrer deste trabalho.

1.2 PROBLEMA

O problema de pesquisa levantando nesta dissertação é a necessidade de interação entre cientistas no desenvolvimento de *workflows* científicos, permitindo também a contextualização destes pesquisadores durante a concepção destes *workflows* e proporcionando ainda a descoberta de novo conhecimento sobre o domínio no qual estão inseridos.

Como este trabalho é uma extensão da abordagem PL-Science (COSTA et al., 2013), algumas questões foram levantadas, considerando a necessidade de colaboração efetiva entre cientistas no

desenvolvimento de novas aplicações através da LPSC. São elas: (i) ausência de informações sobre o contexto que envolve as atividades que os cientistas estão realizando, como por exemplo: quem realizou o experimento, quando, como foi realizado, decisões tomadas, entre outras; (ii) ausência de funcionalidades que possibilitem que os cientistas possam interagir entre eles, trocando experiências e habilidades que possam ser úteis na concepção de *workflows científicos*; (iii) dificuldade para reutilização de artefatos do núcleo da LPS, como a ausência de informações sobre como o artefato foi utilizado, em quais produtos, por quais cientistas, entre outras; e (iv) perda de semântica ao encerrar a utilização de um artefato, como por exemplo perder o contexto no qual o artefato foi empregado.

1.3 HIPÓTESE

Assim, a hipótese desta pesquisa está relacionada aos benefícios de oferecer aos cientistas uma solução para minimizar a dificuldade de colaboração. A hipótese desta pesquisa é apresentada a seguir: *Se oferecermos aos cientistas uma arquitetura que utilize elementos de colaboração (informações de percepção, contexto e um mecanismo de suporte à comunicação) conectado em uma Linha de Produtos de Software Científico (LPSC), podemos criar oportunidades para promover e apoiar a interação, contextualizando-os sobre o ambiente compartilhado de trabalho no qual estão inseridos.*

1.4 OBJETIVOS

Conhecendo os benefícios da utilização de elementos de colaboração em outros domínios, esta dissertação tem como objetivo principal evoluir a abordagem PL-Science (COSTA et al., 2013). Para isso, utiliza elementos de colaboração, tais como, informações de percepção, contexto e um mecanismo de suporte à comunicação, conectados em uma Linha de Produtos de Software Científico. Através da utilização destes elementos, a proposta é replicar os estudos apresentados em COSTA et al. (2013) utilizando a abordagem apresentada nesta dissertação, denominada Collaborative PL-Science.

O objetivo geral desta pesquisa pode ser decomposto nos seguintes objetivos específicos:

- (i) Evoluir a arquitetura da abordagem PL-Science de maneira que ela possa atender aos propósitos desta pesquisa;
- (ii) Oferecer aos cientistas serviços que utilizem elementos de colaboração (*informações de percepção, contexto e um mecanismo de suporte à comunicação*) conectados em uma LPSC;
- (iii) Implementar estes serviços na LPSC, apresentando exemplos de uso, no domínio da bioinformática, mais precisamente na composição de *workflows* para sequenciamento/alinhamento genético;
- (iv) Verificar a viabilidade da solução proposta na criação de oportunidades de interação entre os cientistas a partir dos serviços oferecidos e na contextualização dos pesquisadores sobre o ambiente compartilhado de trabalho no qual estão inseridos.

1.5 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação está organizada nos seguintes capítulos, além desta introdução:

O capítulo 2 apresenta conceitos relacionados, tais como Linha de Produtos de Software, Percepção e Contexto, bem como os principais tipos de percepção e formas de representação de contexto. A utilização de ontologia para representação de contexto também é discutida neste capítulo. A colaboração, destacando a sua importância no domínio científico é outro assunto que também é abordado. Neste capítulo é também apresentada a abordagem PL-Science, que foi evoluída neste trabalho, e também os trabalhos relacionados com esta dissertação.

No capítulo 3, a abordagem *Collaborative PL-Science* é apresentada, bem como sua arquitetura, metodologia e projeto de desenvolvimento da solução proposta. A ontologia criada para representação do contexto e para apoiar a aquisição de conhecimento sobre o domínio também é apresentada e discutida neste capítulo.

O capítulo 4 apresenta quatro provas de conceitos e um estudo de caso realizados através da utilização da abordagem *Collaborative PL-Science* em experimentos realizados na Embrapa e na FioCruz, mais precisamente no contexto de sequenciamento / alinhamento genético. Por último, as considerações finais e os trabalhos futuros são apresentados no capítulo 5.

2 REFERENCIAL TEÓRICO

Este capítulo tem como objetivo apresentar os principais conceitos relacionados ao uso de elementos de colaboração em uma Linha de Produtos de Software Científico. A seção 2.1 apresenta os conceitos envolvidos na abordagem de Linha de Produtos de Software (LPS), sendo detalhado, entre outros, o conceito de Núcleo de Artefatos. Na seção 2.2 é discutida a utilização de Percepção e Contexto na solução proposta neste trabalho de pesquisa. São apresentados os conceitos, os principais tipos de percepção identificados na literatura e, ainda, as principais formas para representação do contexto em Sistemas Colaborativos. Na seção 2.3, são apresentados os conceitos relacionados a ontologia, e sua utilização para a representação de contexto. A seção seguinte apresenta a utilização de elementos de colaboração no domínio científico, destacando a sua importância e os desafios ainda presentes neste cenário. A seção 2.5 apresenta a abordagem PL-Science, a partir da qual se origina a abordagem Collaborative PL-Science proposta nesta dissertação. Já na seção 2.6, são discutidos os trabalhos relacionados com esta pesquisa e, por fim, tem-se as considerações finais sobre este capítulo.

2.1 LINHA DE PRODUTOS DE SOFTWARE

A necessidade de diversificar e liberar produtos rapidamente e eficientemente tem levado os pesquisadores e desenvolvedores de software a buscarem alternativas para aumentar a produtividade, reduzir o tempo para o mercado e ainda tratar a diversidade de produtos do mercado global. Neste cenário, o enfoque de Linha de Produtos de Software (LPS) surge como uma proposta de construção sistemática de software baseada em uma família de produtos (GIMENES & TRAVASSOS, 2002). Segundo CLEMENTS & NORTHROP (2002), Linha de Produtos de Software (LPS) é definida como “[...] um conjunto de sistemas que usam software intensivamente, compartilhando um conjunto de características comuns e gerenciadas, que satisfazem as necessidades de um segmento particular de mercado ou missão, e que são desenvolvidos a partir de um conjunto comum de ativos principais e de uma forma preestabelecida”.

O desenvolvimento de produtos com alto grau de similaridade entre si, empregando o enfoque de Linha de Produtos de Software pode apresentar benefícios consideráveis, como por exemplo, a facilidade na reutilização de artefatos de software, já que nessa abordagem o reuso é planejado e executado de forma sistemática (CLEMENTS & NORTHROP, 2002). Essa abordagem apresenta ainda, ganhos em termos de redução de custo, esforço, melhoria da qualidade e produtividade (BOSCH, 2005), (STEGER et. al, 2004), (ZHANG & JARZAB, 2005). De acordo com YU & SMITH (2009) a abordagem de LPS também melhora a capacidade de manutenção, pois ao contrário de copiar e modificar, os artefatos que são usados para a criação dos produtos da família de software são geridos em conjunto.

A concepção de uma LPS precisa ser desenvolvida partindo de conceitos que levem à generalização de aplicações e que permitam instanciações para aplicações específicas (GIMENES & TRAVASSOS, 2002). Os produtos gerados em uma LPS podem se diferenciar em termos de comportamento, configuração física, atributos de qualidade, plataforma, entre outros (CLEMENTS & NORTHROP, 2002). Para GIMENES & TRAVASSOS (2002) as variações presentes nos produtos de softwares são diferenças tangíveis que podem ser reveladas e distribuídas entre os artefatos da Linha de Produtos. Essas variações podem estar presentes, por exemplo, na arquitetura, nos componentes, nas interfaces entre componentes ou nas conexões entre componentes. GIMENES & TRAVASSOS (2002) asseveram que a instanciação de todas as variações possíveis deve ser consistente e ter um valor semântico que faça sentido para o produto desejado.

Dentro da definição de LPS, alguns termos devem ser destacados, pois representam os pontos mais importantes de uma Linha de Produtos, sendo estes os responsáveis por sustentarem as principais vantagens propostas por este modelo. Segundo o Software Engineering Institute (SEI) (2014) a abordagem de Linha de Produtos de Software estabelece o desenvolvimento de três atividades essenciais, são elas: o Desenvolvimento do Núcleo de Artefatos (Engenharia de Domínio); o Desenvolvimento do Produto (Engenharia de Aplicação); e o Gerenciamento da Linha de Produtos.

Na figura 1 é apresentado o relacionamento entre estas três atividades. Pode-se perceber que elas são atividades dinâmicas e que estão ligadas umas com as outras, ou seja, a ocorrência de uma alteração em qualquer uma destas atividades acarreta em mudança nas demais.

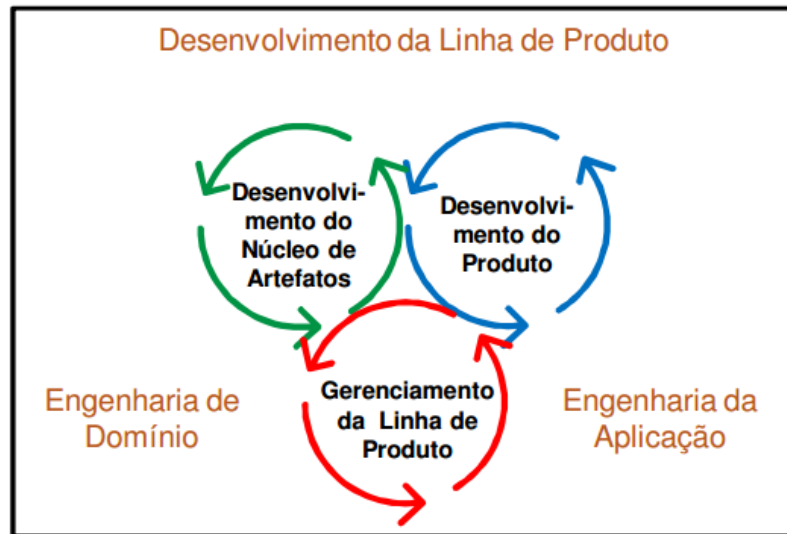


Figura 2.1: Atividades essenciais de uma LPS – adaptada (CLEMENTS & NORTHROP, 2002)

Nas subseções seguintes são apresentadas mais informações sobre estas atividades.

2.1.1 Desenvolvimento do Núcleo de Artefatos

Os artefatos são a essência da Linha de Produtos e correspondem a um conjunto de componentes que podem ser customizados, utilizados como blocos para construção dos novos produtos de software (GIMENES & TRAVASSOS, 2002). Entre os artefatos podemos citar, por exemplo: os modelos utilizados no processo de produção dos artefatos, os próprios artefatos de software, a documentação dos requisitos comuns para a família de produtos, a arquitetura da linha de produtos (base de cada produto gerado), os cronogramas, entre outros.

O desenvolvimento do núcleo de artefatos em uma LPS é um dos pontos que merecem uma atenção especial ao conceber uma Linha de Produtos. Um repositório de artefatos surge com o intuito de organizar e facilitar a busca e utilização de componentes. Entretanto, nem sempre estes artefatos estão armazenados de forma organizada, o que dificulta obter informações sobre como estes componentes podem ser, ou foram, utilizados.

O objetivo do Desenvolvimento do Núcleo de Artefatos é estabelecer uma capacidade de produção para os produtos. Segundo o SEI as três atividades essenciais nessa etapa são:

a) Definir o escopo da linha de produtos: O escopo da Linha de Produtos é uma descrição dos artefatos que constituem os produtos que serão desenvolvidos pela LPS. Em sua forma mais simples, pode consistir de uma lista enumerada de nomes de produtos, embora, geralmente, essa descrição é colocada em termos de similaridades e variabilidades dos produtos. Estas semelhanças e diferenças podem incluir os recursos ou operações que os artefatos fornecem, o desempenho ou outros atributos de qualidade, as plataformas em que são executados, entre outros.

Para que uma Linha de Produtos seja desenvolvida de forma bem sucedida, o seu âmbito deve ser definido com cuidado. Assim a definição do escopo torna-se uma etapa muito importante. Por exemplo, se o escopo é muito grande e os tipos de produtos variam muito, a economia de produção será perdida, e a Linha de Produtos cai em alguma abordagem antiga. Mas, se o alcance definido pelo escopo é muito pequeno, o núcleo de artefatos pode não ser construído de uma forma genérica o suficiente para acomodar o crescimento futuro, e a Linha de Produtos vai estagnar. O estudo de escopo é um ponto bastante importante, embora não seja o foco desta dissertação.

b) Definir a base do núcleo de ativos: esta atividade inclui todos os artefatos necessários, que são a base para o desenvolvimento dos produtos. Esta atividade diz simplesmente como (i) usar os requisitos da Linha de Produtos como requisitos de base, (ii) especificar a exigência de variação para qualquer ponto de variação aceitável, e (iii) confirmar que as alterações e extensões podem ser suportadas pela arquitetura.

c) Seguir um plano de Produção: Um plano de produção prescreve como os produtos são produzidos a partir dos artefatos essenciais que se encontram no núcleo. Ele inclui o processo para ser utilizado na construção dos novos produtos (o processo de produção). Ele descreve ainda os detalhes do projeto para permitir a execução e gestão do processo e, portanto, incluem detalhes como o cronograma de projeto, de materiais e métricas.

Como resultado, estas atividades são utilizadas com entradas para a atividade de Desenvolvimento de Produtos, o que implica em produtos que atendem a um determinado domínio.

2.1.2 Desenvolvimento dos Produtos de Software

Nesta atividade é descrito o funcionamento do processo produtivo da Linha de Produtos. É essa descrição que garante o ganho de tempo e confiabilidade no desenvolvimento dos produtos utilizando a abordagem de LPS. Para a produção de novos artefatos, um Plano de Produção é necessário. Ele define os processos para a produção de cada produto das famílias de produtos de software que será construído.

Estabelecer a relação dos artefatos que farão parte deste produto e ainda, estabelecer um vínculo aos processos anexos de cada artefato utilizado, também são tarefas que deverão ser descritas no Plano de Produção, entre outras atividades importantes.

No contexto deste trabalho, a atividade de Desenvolvimento de Produtos está sendo realizada com o intuito de apoiar os cientistas na concepção de workflows científicos. Entretanto, não está no escopo desta dissertação trabalhar com Linha de Produtos de processos de workflows científicos.

2.1.3 Gerenciamento da Linha de Produtos

A atividade de Gerenciamento da LPS garante que todas as atividades técnicas sejam coordenadas e supervisionadas. Uma das ações mais importantes desta atividade é a criação de um plano de adoção que descreva o estado desejado da organização e uma estratégia para alcançá-lo. Apesar do gerenciamento da Linha de Produtos de Software ser muito importante, ele não é tratado no escopo desta dissertação.

2.1.4 Benefícios da Abordagem de LPS

Para que possamos mapear os benefícios alcançados com a utilização da abordagem de Linha de Produtos de Software, podemos classificá-los de duas maneiras (COHEN, 2003):

- Tangíveis: benefícios que podem ser medidos diretamente, como por exemplo a redução do *time-to-Market* ou redução de defeitos.
- Intangíveis: benefícios que os desenvolvedores relatam, mas que não podem ser medidos em termos de métricas. Esses benefícios podem incluir, por exemplo, a satisfação do cliente.

Alguns benefícios que podem ser alcançados com a adoção deste paradigma de LPS são:

- Atingir ganhos de produtividade em larga escala;
- Melhorar o tempo para a disponibilização para o mercado;
- Melhorar a qualidade do produto;
- Aumentar a satisfação do cliente;
- Atingir os objetivos de reuso, entre vários outros.

No contexto deste trabalho a abordagem de LPS está sendo utilizada como forma de apoiar os cientistas no desenvolvimento de *workflows* científicos. Atualmente, estes *workflows* são utilizados em várias situações, como por exemplo, para realizar simulações ou mesmo para testar uma teoria científica. Sendo assim, associado à abordagem de LPS, oferecer aos participantes mecanismos que possibilitem conhecer e compreender melhor o ambiente de trabalho no qual estão inseridos, pode também trazer benefícios para este domínio. Assim, na seção seguinte são apresentados os conceitos de percepção e contexto, que também são utilizados como parte da solução proposta para o problema levantado nesta dissertação, e que é apresentado no Capítulo 3 deste trabalho.

2.2 PERCEPÇÃO E CONTEXTO

2.2.1 Percepção

No trabalho em grupo, perceber pode ser entendido como a compreensão e o conhecimento adquirido sobre os acontecimentos que ocorrem dentro do grupo. DOURISH & BELLOTTI (1992)

definem percepção como uma compreensão das atividades dos outros, que provê um contexto para as atividades próprias. Em um grupo de pesquisadores, por exemplo, o conhecimento sobre o término de uma atividade realizada por um membro, pode ser importante para que o grupo possa conhecer os resultados desta tarefa, e dar continuidade em um experimento. Esse conhecimento e compreensão adquiridos a partir dessa percepção podem garantir que as contribuições individuais sejam relevantes para as atividades do grupo como todo. Além disso, os participantes do grupo podem se avaliar quanto à realização das suas ações individuais em relação aos objetivos e progresso do grupo.

Perceber e entender as atividades realizadas por outros participantes são requisitos fundamentais para que haja comunicação e interação do grupo em direção a um objetivo comum (DAVID, 2004). Percebe-se então, que a percepção é um fator determinante para uma colaboração efetiva, sendo considerada uma das principais funcionalidades das aplicações colaborativas (VIEIRA, TEDESCO & SALGADO, 2011). Através dela os participantes do grupo podem diminuir as sensações de trabalho impessoal, de isolamento e de distância, características comuns em ambientes virtuais (GEROSA et al., 2001). Sendo assim, dentre os benefícios que a percepção pode oferecer para trabalho em grupo, destacam-se: (i) a oportunidade de ampliar de memória de grupo; (ii) reduzir o isolamento entre os participantes; (iii) gerar a oportunidade do participante medir a qualidade do seu trabalho; e (iv) colaborar de maneira mais efetiva; entre outras.

Entretanto, para assimilar o que ocorre dentro do grupo, o usuário deve ter à sua disposição informações que possibilitem a percepção. Para isso, a disponibilização destas informações de percepção devem auxiliar os participantes a montar o contexto de suas atividades em relação às dos demais, coordenando-se para que os esforços de comunicação e de trabalho sejam transformados em cooperação (VIEIRA, TEDESCO & SALGADO, 2011). Diante deste cenário dinâmico, um participante do grupo pode valer-se dessas informações para obter uma visão mais ampla do desenrolar das atividades e do estado de cada uma delas. Com isso, o coordenador do grupo, por exemplo, pode organizar os demais participantes dentro do grupo incentivando a cooperação entre eles (GEROSA et al., 2001). Sem tais informações, os indivíduos não podem, por exemplo, medir a qualidade de seu próprio trabalho referente aos objetivos e progressos definidos pelo grupo (DOURISH & BELLOTTI, 1992).

Apesar disso, não se pode garantir que o participante tenha a percepção do que ocorre dentro do grupo, já que a percepção é um “estado mental” (VIEIRA, TEDESCO & SALGADO, 2011). O que se faz então, é oferecer mecanismos na aplicação que possam facilitar a assimilação das informações de percepção disponibilizadas para os membros do grupo. Para isso, os participantes necessitam de mecanismos adequados de percepção que processem as informações geradas. Como exemplo, pode-se citar um mecanismo de linha do tempo apresentado em uma aplicação. Através deste recurso, um participante pode conhecer as atividades que estão sendo realizadas por outros membros e, a partir destas informações, decidir se interage ou não com um membro do grupo. Outro exemplo, pode ser o momento em que uma interação é iniciada. Geralmente, uma notificação sobre o evento ocorrido é apresentada ao participante, seja através de uma mensagem ou através de algum outro recurso visual, por exemplo. Assim, a percepção é alcançada quando os participantes de um grupo visualizam o mecanismo utilizado, processam e compreendem as informações oferecidas.

2.2.2 Contexto

Diferentes definições são encontradas na literatura para o termo contexto [GREENBERG (2001), RITTENBRUCH (2002), ALARCÓN et al. (2005), BAZIRE e BRÉZILLON (2005)]. DEY (2001), por exemplo, define contexto como qualquer informação que caracteriza a situação de uma entidade, onde uma entidade é uma pessoa, lugar ou objeto considerados relevantes para a interação entre um usuário e uma aplicação, incluindo o próprio usuário e a aplicação. O contexto é tipicamente a localização, a identidade, e o estado das pessoas, grupos e objetos físicos e computacionais.

Em sistemas computacionais, a maior parte das abordagens geralmente não levam em consideração o contexto do usuário com quem estão interagindo. Com isso, estas aplicações agem de maneira pré-programada, e não conseguem discriminar as necessidades e as restrições do ambiente no qual estão envolvidas. Em aplicações científicas, que é o foco deste trabalho, a ação de contextualizar os cientistas envolvidos em uma atividade pode resultar em contribuições relevantes para o grupo. Por exemplo, se um mecanismo de busca, utilizado em uma abordagem, conhecesse o contexto do cientista, poderia personalizar a busca e priorizar artefatos mais adequados para o produto no qual o pesquisador está trabalhando. Mas o que geralmente

acontece não é exatamente isso. Por isso, conhecer o contexto que envolve os participantes, pode permitir que as aplicações ofereçam serviços e informações mais relevantes para os seus usuários. Caso contrário, a ausência deste conhecimento pode acarretar alguns problemas para o ambiente, como retrabalho, inconsistências e contradições, bem como impedir que comunicações espontâneas entre os participantes sejam iniciadas, entre outros.

Assim, torna-se imprescindível que estas informações sejam constantemente atualizadas no cenário onde os participantes estão inseridos. No contexto desta dissertação a utilização de informações de percepção e contexto pode trazer ganhos para cenário científico, principalmente por se tratar de um ambiente dinâmico e muitas vezes complexo. Com isso, identificar os tipos de percepção existentes, as principais formas para a representação de contexto, bem como a forma como estas informações devem ser coletadas e apresentadas para os participantes torna-se necessário para auxiliá-los na realização das suas atividades. Estes tópicos são abordados na seção seguinte.

2.2.3 Tipos de Percepção e Formas de Representação de Contexto

Nesta seção são apresentados os diferentes tipos e conceitos de percepção encontrados na literatura. O objetivo desta identificação está relacionado à seleção dos tipos que possam ser utilizados pela abordagem proposta nesta dissertação. Além disso, nesta seção são também apresentadas as principais formas para representação de contexto e destacada a forma escolhida para ser utilizada nesta dissertação. Uma revisão na literatura foi realizada com objetivo ampliar a tabela apresentada por DAVID (2004) com os principais conceitos de percepção. Entretanto, outra revisão da literatura apresentada por STEINMACHER, CHAVES & GEROSA (2012) também permitiu que esse propósito pudesse ser alcançado.

Na Tabela 1 esses conceitos são apresentados de forma resumida, procurando ressaltar os tipos de percepção, os autores e uma breve descrição do seu objetivo. Outros tipos foram identificados, tais como: Percepção Pessoal, Percepção de Atividade e Percepção Contextual, permitindo assim a ampliação da Tabela 1.

Tabela 1. Tipos de Percepção - adaptado de DRURY & WILLIAMS (2002)

PERCEPÇÃO	AUTORES	DESCRIÇÃO
Percepção	DOURISH e BELLOTTI (1992)	<ul style="list-style-type: none"> ○ Compreensão das atividades dos outros participantes, capaz de fornecer o contexto para as suas próprias atividades.
Percepção Orientada a Tarefa	PRINZ (1999)	<ul style="list-style-type: none"> ○ Percepção focada em atividades realizadas para a conclusão de uma tarefa compartilhada.
Percepção Social	GUTWIN et al.(1995) GUTWIN et al.(1996) PRINZ (1999)	<ul style="list-style-type: none"> ○ Compreensão que os participantes têm sobre as conexões sociais dentro do seu grupo. ○ Informação sobre a presença e atividades de pessoas no espaço de trabalho compartilhado.
Percepção Grupo-Estrutural	GUTWIN et al. (1996)	<ul style="list-style-type: none"> ○ Conhecimento sobre os papéis e responsabilidades dos participantes, suas posições em relação a uma questão, seus status e processos do grupo.
Percepção de Tarefa	GUTWIN et al. (1995)	<ul style="list-style-type: none"> ○ Compreensão dos participantes de como as suas tarefas serão finalizadas.
Percepção de Conceito	GUTWIN et al. (1995)	<ul style="list-style-type: none"> ○ Compreensão dos participantes de como as suas tarefas são realizadas.
Percepção do Espaço de Trabalho	GUTWIN et al.(1995) VERTEGAAL et al.(1997)	<ul style="list-style-type: none"> ○ Conhecimento instantâneo das interações de outros participantes com o espaço de trabalho compartilhado. ○ Conhecimento de quem está trabalhando com o que.

Tabela 1. Tipos de Percepção - adaptado de DRURY & WILLIAMS (2002) – Continuação

PERCEPÇÃO	AUTORES	DESCRIÇÃO
Percepção Informal	GUTWIN, STARK e GREENBERG(1995)	<ul style="list-style-type: none"> ○ Informação sobre quem está ao redor, e o que eles são capazes de realizar. ○ O sentido geral de quem está ao redor.
	GUTWIN et al.(1996)	
Percepção Passiva	DOURISH e BLY (1992)	<ul style="list-style-type: none"> ○ Informação passivamente coletada e apresentada no espaço de trabalho.
Percepção Periférica	GUTWIN et al.(1996)	<ul style="list-style-type: none"> ○ Apresentação da localização das pessoas no contexto global.
	BAECKER et al.(1993)	<ul style="list-style-type: none"> ○ Informação que permite que as pessoas saibam o que os outros participantes estão realizando.
Percepção Conversacional	VERTEGAAL et al. (1997)	<ul style="list-style-type: none"> ○ Conhecimento sobre quem está se comunicando com quem.
Percepção Mútua	SCHMIDT (1998)	<ul style="list-style-type: none"> ○ Conhecimento sobre as atividades que acontecem no grupo, e o que os outros fizeram ou estão fazendo.
Percepção Pessoal	LICCARDI et. al.(2008)	<ul style="list-style-type: none"> ○ A informação que os usuários possuem sobre si mesmos e seus papéis no grupo.
Percepção de Atividade	VIEIRA et. al. (2011)	<ul style="list-style-type: none"> ○ Conhecimento focado nas atividades desempenhadas para se atingir um trabalho específico e os objetivos do grupo. Sua importância é evidenciada em interações assíncronas.
Percepção Contextual	LIECHTI (2000)	<ul style="list-style-type: none"> ○ Utiliza o conceito de contextos para determinar que tipo de informação os usuários desejam estar atentos e de que forma eles podem estar atentos, de modo a separar informações de percepção das informações de “perturbação”.

LICCARDI et al. (2008) define a *Percepção Pessoal* como a informação que os usuários possuem sobre si mesmos e seus papéis no grupo. Segundo o autor, esta percepção pode ser

síncrona, como o atual paradeiro do usuário dentro do sistema, por exemplo, ou assíncrona. A *Percepção de Atividade* é definida por VIEIRA et al. (2011) como sendo o conhecimento que os participantes têm sobre as atividades desempenhadas pelo grupo para se atingir um trabalho específico e os objetivos da equipe. Sua importância é evidenciada em interações assíncronas, podendo o desconhecimento das atividades do participante e de sua posição na estrutura de atividades do grupo incorrer em duplicidade de trabalho ou em alto nível de conflitos.

Outro tipo de percepção identificado foi a *Percepção Contextual*, apresentada por LIECHTI (2000). O autor discute que este tipo de percepção utiliza o conceito de contexto para determinar que tipo de informação os usuários desejam estar atentos e de que forma eles podem estar atentos, de modo a separar informações de percepção das informações de “perturbação”.

Através da Tabela 1 foi possível identificar os tipos de percepção que poderiam ser utilizados inicialmente na abordagem Collaborative PL-Science, proposta neste estudo. Estas informações são apresentadas e discutidas no capítulo 3 desta dissertação.

Conforme mencionado acima, formas para representar o contexto dos participantes também tem sido discutida na literatura. Abaixo é apresentada uma tabela com as principais tecnologias que vêm sendo utilizadas como forma de representação de contexto para abordagens computacionais, destacando as vantagens e desvantagens de cada tecnologia.

Tabela 2. Abordagens de Representação de Contexto - adaptado de VIEIRA (2006)

Modelos	Vantagens	Desvantagens
Modelo Baseado em Pares de Chave-Valor	Modelo simples, de fácil implementação e uso.	Não considera hierarquia. Inadequado para aplicações com estruturas complexas.
Modelo Baseado em Linguagens de Marcação	Baseado no padrão XML. Prevê hierarquia entre os elementos. O esquema de marcação implementa o próprio modelo. Utilização típica em perfis.	Incompletude e ambiguidade na informação contextual deve ser tratada pelo sistema. Inadequado para representar estruturas complexas.

Tabela 2. Abordagens de Representação de Contexto - adaptado de VIEIRA (2006) –
Continuação

Modelos	Vantagens	Desvantagens
Modelo Baseado em Representações Gráficas	Ajuda a estruturar o contexto e a modelar o sistema ciente de contexto. Apoia a compreensão do contexto por humanos	Não utilizado para instanciar a informação. Falta formalismos computacionais que permitam a compreensão do contexto por máquinas.
Modelo Orientado a Objetos	Usufri das propriedades de encapsulamento e reusabilidade. Permite estruturar a informação contextual e abstrair tratamento do contexto	A invisibilidade no tratamento do contexto, pelo encapsulamento, dificulta o formalismo do modelo.
Modelos Baseados em Lógica	Contextos são definidos como fatos, expressões e regras. Alto grau de formalismo.	Difícil compreensão por humanos. Difícil modelagem da estrutura do contexto.
Modelo Baseado em Grafos Contextuais	Adequado para modelar contextos associados a processos	Não permite formalizar a estrutura do contexto.
Modelo Baseado em Mapas de Tópicos	Facilita a navegação entre os contextos. Facilita modelagem dos contextos por humanos.	Estágio inicial. Tecnologia imatura. Faltam exemplos reais. Baixo formalismo.
Modelo Baseado em Ontologias	Contextos modelados como conceitos e fatos. Viabiliza formalização e compartilhamento por humanos e computadores.	Muitos recursos estão sendo desenvolvidos e testados. Carece do desenvolvimento de alguns recursos para uma aplicação prática (dependendo do contexto).

Nesta dissertação, a forma escolhida e utilizada para representação do contexto envolvido é através de ontologia. As vantagens e os benefícios na sua utilização são apresentados na seção seguinte. Nela, também é apresentado como ontologias podem ser utilizadas para a representação do contexto que envolve os participantes de um grupo, e ainda auxiliá-los na realização do seu trabalho.

2.3 USO DE ONTOLOGIAS PARA REPRESENTAÇÃO DE CONTEXTO

Prover o entendimento e o compartilhamento comum sobre um determinado domínio com o intuito de facilitar a comunicação entre as pessoas e os sistemas de aplicação, tem colaborado para que as ontologias se tornem mais populares (DAVIES et al., 2006). Entretanto, ontologias já vêm sendo utilizadas pela Ciência da Computação há vários anos, visando criar representações que fossem além da descrição de simples instâncias do domínio (GAVA & MENEZES, 2003).

Atualmente, diversas áreas utilizam ontologias, buscando desenvolver um vocabulário contendo os conceitos relativos ao domínio de aplicação. A área de Ciência da Computação define, em geral, ontologia como uma especificação formal e explícita de uma conceituação compartilhada [GRUBER (1993), GAVA & MENEZES (2003)]. Isso quer dizer que, através do uso de ontologias, é possível estabelecer uma compreensão comum sobre objetos e os relacionamentos existentes entre eles em um determinado domínio. Isso porque, segundo GRUBER (1995), as ontologias fornecem um modelo formal e manipulável deste domínio.

A grande vantagem na sua utilização está na possibilidade de fixar de forma rigorosa o significado pretendido para o vocabulário do domínio, podendo assim ser utilizada para a representação do conhecimento. Além disso, as ontologias possibilitam ainda que essa representação de um determinado domínio possa ser lida e processada por sistemas computacionais. Com isso, máquinas de inferências (*reasoners*) também podem ser utilizadas para derivar novos conhecimentos e relacionamentos, que anteriormente estavam implícitos (COSTA et al., 2013), (FILHO et al., 2012).

Sendo assim, pode-se perceber que a utilização de ontologias para descrição semântica de um vocabulário proporciona um entendimento amplo das características e propriedades de suas classes e dos relacionamentos entre elas. Além disso, pode-se acrescentar o fato de serem extensíveis, pois novas classes, regras ou vocabulários podem ser adicionados para descrição de um novo domínio de aplicação (MATOS, 2008). Para VIEIRA (2006), “o uso de ontologias é muito útil para checar a validade da informação contextual e para facilitar a especificação do comportamento de aplicações cientes de contexto”.

Além disso, as ontologias podem ser compartilhadas para uso em conjunto com outras ontologias ou mesmo com outras ferramentas. Sabendo que o contexto de um domínio evolui, a

aplicação pode, em um determinado momento, se mover para um novo ambiente, e assim apresentar um contexto diferente. RANGANATHAN et al. (2003) asseveram que se a diferença neste novo contexto for terminológica, as regras apresentadas pela ontologia podem ser traduzidas e funcionarem corretamente neste novo ambiente. Caso não seja, outras metodologias podem ser aplicadas, por exemplo, através da utilização de redes de ontologias, onde outras ontologias podem ser conectadas umas com as outras possibilitando a ampliação do conhecimento que se tem sobre o domínio (CAMPOS et al., 2012).

Diversos trabalhos encontrados na literatura vêm abordando a utilização de ontologias para representação de contexto (CHEN et al., 2003; CHRISTOPOULOU et al., 2004; GAUVIN et al., 2004; PREUVENEERS et al., 2004; WANG et al., 2004; BULCÃO NETO E PIMENTEL, 2005; VIEIRA et al., 2005; OLIVEIRA, 2009). Algumas formas sobre como as ontologias têm sido utilizadas para representar o contexto são descritas a seguir (RANGANATHAN et al., 2003): *i*) ontologias são utilizadas para permitir que usuários possam obter uma melhor compreensão do ambiente e de como os diferentes componentes se relacionam uns com os outros; *ii*) para permitir a descoberta semântica de entidades; *iii*) são utilizadas também para permitir que novas entidades possam interagir com a aplicação, bem como permitir que diferentes ambientes interajam uns com os outros; *iv*) seu uso tem por objetivo ainda facilitar e permitir que agentes humanos e de software possam realizar buscas em diferentes componentes, interajam com diferentes entidades e especifiquem regras para comportamentos sensíveis a contexto nessas entidades; entre outras formas de utilização que também podem ser encontradas na literatura.

Um dos grandes interesses na construção e no uso de ontologias é justamente tornar o conhecimento sobre o mundo real processável por máquinas, uma vez que, uma representação explícita de uma ontologia consiste de um modelo estruturado de vocabulários (classes e propriedades) e sua semântica associada (relacionamentos entre as diferentes classes e propriedades), bem como funções, axiomas e instâncias (WANG et al., 2004).

Buscando minimizar alguns dos problemas enfrentados pela *e-Science*, domínio no qual a abordagem apresentada nesta dissertação está inserida, diferentes recursos computacionais podem ser associados, para juntos serem utilizados no tratamento de questões ainda pendentes na ciência. Por exemplo, minimizar as barreiras de comunicação entre equipes de cientistas geograficamente distribuídos, gerando oportunidades de interação entre eles. Com isso, espera-se

que através da utilização de informações de percepção e contexto e ad abordagem de LPS, seja possível gerar oportunidades de colaboração entre estes participantes do domínio. Para isso, a seção seguinte discute a colaboração e ressalta os desafios e as suas vantagens no domínio de *e-Science*.

2.4 COLABORAÇÃO

Para solucionar problemas complexos, a formação de grupos de pessoas com conhecimentos e habilidades diferentes é um fator importante. Isso permite que haja divisão dos objetivos entre os participantes do grupo com o propósito de buscar possíveis soluções para o problema levantado. O trabalho em grupo promove interações entre os participantes, gerando ainda discussões sobre alternativas, e auxiliando na tomada de decisões (FUKS, RAPOSO & GEROSA, 2003). Através desta oportunidade, os participantes podem colaborar uns com os outros e obter resultados melhores do que se estivessem trabalhando individualmente (PIMENTEL & FUKS, 2011). Quando colaboram, os membros de um grupo se ajudam objetivando o sucesso das tarefas e atuam em conjunto com objetivos comuns e compartilhados pelo grupo.

No ambiente de trabalho, a colaboração é de grande valia, pois possibilita ao grupo tratar tarefas complexas que requerem habilidades multidisciplinares. Com isso, a colaboração permite que uma equipe possa complementar a sua capacidade, através do auxílio mútuo. Além disso, participantes são preparados o tempo todo para se relacionar, negociar, se expor, liderar, ter responsabilidade, se comunicar, coordenar e cooperar (FUKS et al., 2007).

Diante disso, as práticas de colaboração contribuem para agregar valores organizacionais para a equipe e, conseqüentemente, proporcionar uma maior vantagem competitiva perante outros grupos (SANTANEN, KOLFSCHOTEN & GOLLA, 2006). Entretanto, para que isso seja possível, faz-se necessário oferecer ferramentas que apoiem o trabalho colaborativo e que também potencialize a colaboração entre os participantes. Para isso, uma área de estudo denominado CSCW (*Computer-Supported Cooperative Work*) tem sido a responsável pela investigação de como ocorrem as atividades em grupo, e a partir daí oferecer meios para o desenvolvimento de tecnologias mais adequadas para estas equipes.

Vários modelos de colaboração são propostos na literatura com o objetivo de fornecerem uma visão sobre como e porque as pessoas trabalham em grupo [ELLIS, GIBBS E REIN (1991), LAURILLAU & NIGAY (2002), FUKS et al. (2007), OLIVEIRA, (2009)]. Um dos mais discutidos e utilizados atualmente tem sido o modelo 3C de Colaboração (FUKS et al. 2007) originalmente definido por ELLIS, GIBBS E REIN (1991). Este modelo é baseado na concepção de que para colaborar, os participantes de um grupo se comunicam, coordenam e cooperam. Sobre o modelo 3C, pode-se dizer ainda que ele é equivalente ao modelo *Clover* apresentado por LAURILLAU & NIGAY (2002). Neste modelo são definidas três classes de funcionalidades: comunicação, coordenação e produção. O que é chamado de produção no modelo *Clover* corresponde ao conceito de cooperação no modelo 3C. Outros modelos [OLIVEIRA, (2009), LIMA & DAVID (2013)] também são apresentados na literatura os quais, na maioria das vezes, são evoluções de outros modelos. Um exemplo é o modelo apresentado por LIMA & DAVID (2013) o qual adapta o modelo 3C para incluir os relacionamentos entre os seus elementos e os elementos de riscos de projeto e de colaboração. No contexto de e-Science, os modelos de colaboração podem ser úteis, por exemplo, para oferecer aos cientistas, mecanismos para comunicação, informações de percepção e contexto, instrumentos apoiar a coordenação e a cooperação entre a equipe de pesquisadores, de maneira que estes recursos possam auxiliar o trabalho destes participantes em seu grupo gerando ainda, por exemplo, oportunidades de interação entre eles.

A colaboração no domínio científico é discutida na subseção seguinte, sendo ressaltado alguns dos problemas clássicos encontrados neste cenário.

2.4.1 Colaboração em e-Science

A ciência moderna é cada vez mais colaborativa, embora a pesquisa conduzida de forma individual por um cientista ainda continue sendo importante e continue tendo o seu papel no domínio científico. Entretanto, a colaboração tornou-se uma característica fundamental da ciência (WRAY, 2002). Várias evidências encontradas na literatura abordam que a colaboração: (i) aumenta a qualidade de pesquisa; (ii) possui um papel importante na formação de novos pesquisadores; e ainda (iii) contribui para o rápido crescimento do conhecimento científico; (CHIN & LANSING, 2004), (OLSON, ZIMMERMAN & BOS, 2008), (MOREIRA, VIEIRA & DEL ARCO, 2012).

Apesar disso, a necessidade de colaboração no domínio científico não é algo recente. Em 1989, um grupo de pesquisadores da Rockefeller University (WULF, 1989) apresentou o termo *Collaboratory*, uma combinação das palavras em inglês para os termos *Collaboration* + *Laboratory*. Este termo começou a ser utilizado então para se referir à “centro (s) sem paredes em que os pesquisadores possam trabalhar juntos, independentemente da localização geográfica, para atingir uma meta científica” (OLSON, ZIMMERMAN & BOS, 2008).

Segundo OLSON, ZIMMERMAN & BOS (2008), os pesquisadores da Rockefeller University tinham a visão de que a tecnologia em 1989 já permitia pensar nas atividades rotineiras da ciência ocorrendo através da internet. Assim, soluções poderiam ser encontradas para tratar alguns dos problemas clássicos do domínio científico, como por exemplo a dificuldade de comunicação e cooperação entre equipes de cientistas geograficamente distribuídas (WULF, 1989).

Atualmente os termos *e-Science* ou *e-Ciência* (utilizado recentemente no Brasil) são os mais utilizados para se referir às ideias do *Collaboratory* (WULF, 1989). Segundo o National e-Science Center, o termo *e-Science* refere-se à “ciência em larga escala, que será cada vez mais, realizada por colaborações globais e distribuídas através da internet”. E-Science é então, em síntese, a união da Tecnologia da Informação e Comunicação (TIC) e a ciência. Com isso, o trabalho científico passa a ser facilitado por ferramentas computacionais, permitindo a colaboração em longas distâncias e ainda, tornando possível abordar questões que não podem ser realizadas de forma individual. Por exemplo, a maior parte dos problemas científicos atuais são mais complexos, e consideram mais de uma disciplina ou pesquisador para resolvê-los. A crescente necessidade de especialização dos pesquisadores, devido ao crescimento do conhecimento científico é outro fator que agrava ainda mais essa situação (CORDEIRO et al., 2013).

Entretanto, a dificuldade para conseguir alinhar os objetivos e interesses do grupo, e fazer uma divisão apropriada de tarefas e dos recursos, torna a colaboração uma tarefa difícil de se realizar neste cenário. Além disso, diferentes estudos apontam que as colaborações que envolvem participantes geograficamente dispersos têm uma maior probabilidade de falha ou desempenho insuficiente [CORLEY, BOARDMAN, & BOZEMAN (2006), BOH et al. (2007), CUMMINGS & KIESLER (2007)]. Apesar disso, a colaboração é importante e desejável no cenário científico. A utilização de elementos de colaboração em *e-Science* pode ser útil para apoiar a coleta e a disponibilização de decisões e resultados obtidos em experimentações científicas. Como resultado, oportunidades de

interações podem ser oferecidas. Essas informações podem ser integradas com o conhecimento existente no grupo para manter um senso de percepção da atividade realizada. Assim, os benefícios das ações executadas pelos cientistas podem ser encontrados, por exemplo, capturando os passos de interações que ocorrem, ou ocorreram, no espaço de trabalho compartilhado. Esses passos podem ser detalhados quando os cientistas atuam sobre suas tarefas deixando o histórico de suas atividades. Esse histórico se torna útil nesse cenário, por permitir que cientistas possam reutilizar em suas tarefas, decisões e experiências já vividas por outros cientistas previamente.

A necessidade destes elementos de colaboração, por exemplo, informações de percepção e contexto, e mecanismos de comunicação, para o domínio científico pode ser observada na seção seguinte. Nela é apresentada a abordagem PL-Science da qual a solução proposta neste trabalho é uma extensão.

2.5 ABORDAGEM PL-SCIENCE

A abordagem PL-Science foi proposta por COSTA et al. (2013), com o propósito de auxiliar os cientistas no processo de escolha e definição de aplicações científicas baseadas em *workflows*. Para isso, os autores exploram as vantagens da abordagem de Linha de Produtos de Software, buscando o apoio em modelos de *features* associados a ontologias para facilitar o desenvolvimento de *workflows*. A partir daí, propõem a criação de uma Linha de Produtos de Software Científico (LPSC).

A abordagem PL-Science propõe uma metodologia para o desenvolvimento destes novos produtos científicos. A figura 2 ilustra os processos envolvidos nesta abordagem. Na etapa de Engenharia de Domínio, que compreende a fase de Desenvolvimento do Núcleo de Artefatos da LPSC, os autores definem os seguintes passos para serem realizados pelos usuários da abordagem: (i) a definição do escopo da LPSC; (ii) a definição do artefato modelo de *features*; (iii) a definição de uma ontologia de domínio, a qual será associada ao modelo de *features*. É importante ressaltar que a ontologia utilizada aqui representa o conhecimento envolvido no domínio dos produtos que serão construídos pela LPSC, como, por exemplo, uma ontologia no domínio de bioinformática para auxiliar na construção de *workflows* científicos para sequenciamento / alinhamento genético; (iv) a construção de um arquivo XML que mapeie os termos da ontologia com o modelo de

features; (v) a definição de um modelo abstrato de *workflow* o qual é utilizado como base para a construção dos novos produtos; e (vi) a definição e a conexão de outros artefatos, neste caso algoritmos e/ou *web services* para a serem utilizados na realização de uma tarefa do *workflow*.

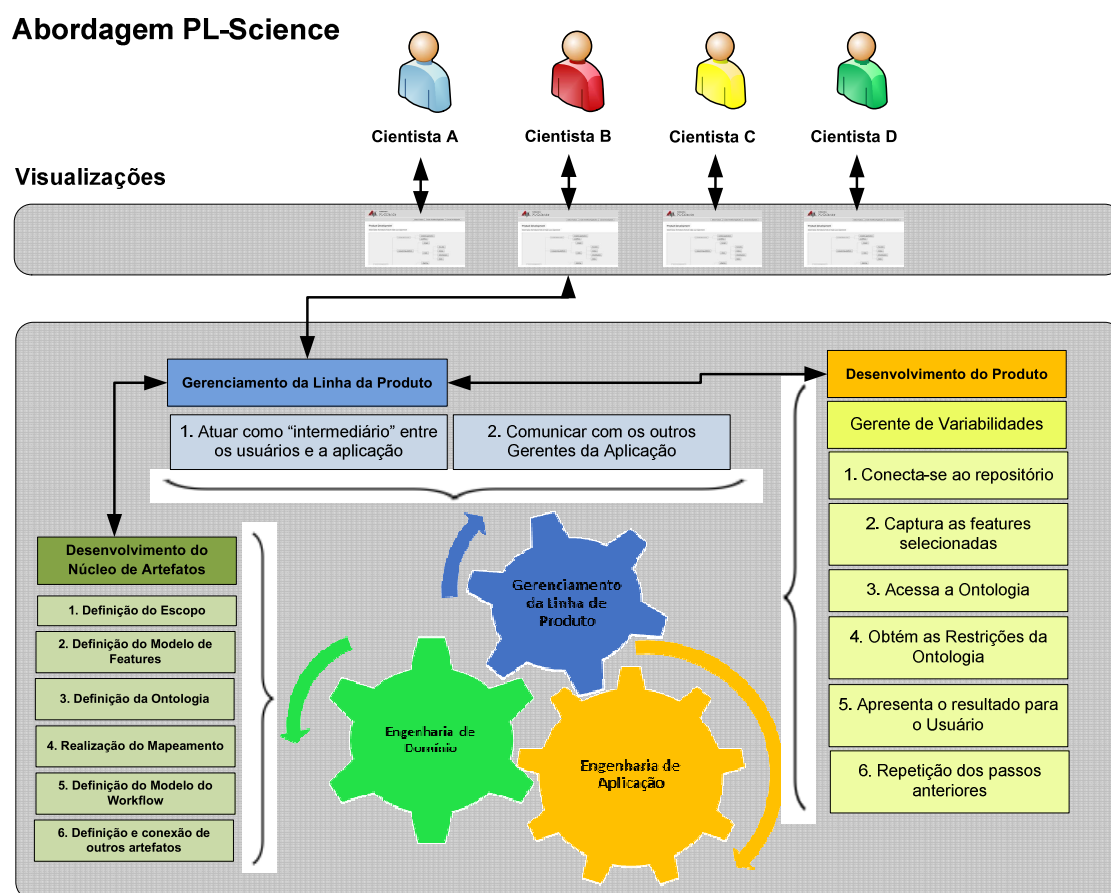


Figura 2.2: Visão geral da Abordagem PL-Science – adaptado de COSTA et al. (2013)

Atualmente, no núcleo de artefatos da abordagem PL-Science estão presentes os seguintes artefatos: modelo de *features*, arquivo de mapeamento, ontologia e os algoritmos e/ou *web services* que hoje estão armazenados como indivíduos na ontologia de domínio que é utilizada para na construção dos novos produtos da LPSC. A atividade de Gerenciamento da Linha de Produtos, desempenha o papel de um “intermediário” entre os demais gerentes da LPSC e os usuários da aplicação. Além disso, tem o papel também de se comunicar com os demais gerentes.

A atividade de Desenvolvimento do Produto (Engenharia de Aplicação) engloba os seguintes passos: (i) o usuário se conecta ao núcleo de artefatos da aplicação, e escolhe os artefatos que serão utilizados para a concepção do novo produto; (ii) o usuário escolhe uma *feature* no modelo de *features* para a construção do produto; (iii) através da leitura do arquivo de mapeamento a aplicação realiza a leitura da ontologia; (iv) as restrições apresentadas por esta ontologia são obtidas; (v) o sistema apresenta o resultado para que o usuário decida pelo próximo passo a ser dado; e (vi) os itens II ao item V são realizados novamente até que se chegue ao término da concepção do produto, a partir do qual os cientistas (usuários da aplicação) farão a escolha dos algoritmos e/ou *web services* que irão compor o produto que está sendo construído.

A figura 3 ilustra a organização do Núcleo de Artefatos da abordagem PL-Science, bem como apresenta o relacionamento entre estes artefatos.

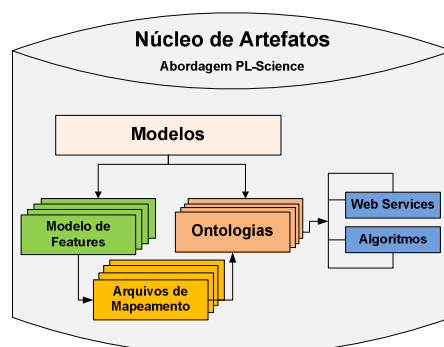


Figura 2.3: Núcleo de Artefatos da Abordagem PL-Science.

Conforme é apresentado na Figura 3, na abordagem PL-Science os cientistas possuíam apenas os artefatos persistidos no núcleo da LPSC como componentes de trabalho. Com isso, o cientista desenvolvia um *workflow* científico baseado apenas em seu conhecimento sobre o domínio. Nenhum histórico, ou "*rationale*¹", era gerado neste ambiente, ficando todo o conhecimento sobre as funcionalidades dos artefatos, as experiências do cientista e as tomadas de decisões por conta do usuário. Neste contexto algumas lacunas são deixadas pela abordagem PL-Science, tais como: (i) a ausência de informações sobre o contexto que envolve as atividades que

¹ Rationale pode ser entendido como uma lista explícita de decisões tomadas durante o processo de projeto e/ou desenvolvimento de um artefato, por exemplo, e as razões pelas quais as decisões foram tomadas. O seu objetivo principal está em apoiar os envolvidos, proporcionando um meio para gravar e comunicar a argumentação e o raciocínio por trás do processo de projeto.

os cientistas estão realizando; (ii) a falta de um serviço que possibilite interação entre os cientistas; (iii) a dificuldade para reutilização de artefatos; e (iv) a perda de semântica ao encerrar a utilização de um artefato; entre outras.

Esta dissertação busca minimizar estas lacunas, e como proposta de solução para estes problemas, apresenta uma nova abordagem, denominada Collaborative PL-Science que é apresentada no capítulo 3 deste trabalho. Na seção seguinte são descritos os trabalhos existentes que se relacionam com a proposta de solução apresentada nesta dissertação.

2.6 TRABALHOS RELACIONADOS

Diferentes trabalhos encontrados na literatura estão inseridos no domínio de *e-Science*. Em sua maioria, busca-se explorar os principais desafios envolvidos no apoio à composição de experimentos científicos. Entretanto, alguns apresentam propostas mais próximas ao objetivo desta dissertação, ou seja, exploram o potencial da utilização da abordagem de Linha de Produtos de Software na área científica, visando auxiliar o trabalho dos cientistas e minimizar problemas neste domínio. Outros, embora não estejam diretamente relacionados ao domínio científico, abordam a utilização de elementos de colaboração para auxiliar as atividades dos participantes no domínio que estão inseridos.

MATTOSO et al. (2009) exploram os principais desafios envolvidos no apoio à composição de experimentos científicos, ressaltando os problemas encontrados nesse domínio. Neste trabalho, os autores destacam o apoio limitado que é oferecido aos cientistas durante a concepção e posterior instanciação do *workflow* científico. Para isso, utilizam técnicas de Engenharia de Software e Banco de Dados, como por exemplo, a utilização de uma Linha de Experimentos, que é inspirada no conceito de LPS em Engenharia de Software. Entretanto, seu propósito principal está na possibilidade de formalizar um *workflow* padrão para a realização de um determinado tipo de experimentação. O trabalho de MATTOSO et al. (2009) se relaciona com a abordagem Collaborative PL-Science no sentido de mostrar a relevância de explorar os problemas e os desafios para o avanço das pesquisas em e-Science. Entretanto, este trabalho não aborda a utilização de elementos de colaboração para apoiar este cenário científico. Diferentemente da proposta de MATTOSO et al. (2009), a abordagem Collaborative PL-Science está interessada em

oferecer auxílio aos cientistas durante a realização da sua atividade de concepção de *workflow* científico, de modo a gerar oportunidades de interação entre eles. Para isso, utiliza alguns elementos de colaboração, tais como informações de percepção e contexto e um mecanismo de suporte à comunicação conectados em uma LPSC. A Collaborative PL-Science não tem o interesse inicial na execução do *workflow* científico.

IVANOV et al. (2013) descrevem uma forma de organizar o apoio à decisão colaborativa com base na plataforma CLAVIRE, no domínio de *e-Science*. Neste trabalho, os autores asseveram sobre a importância de mecanismos para estreitar a interação entre especialistas e tomadores de decisão para resolverem os problemas interdisciplinares complexos em simulações de *workflows* científicos. Este trabalho segue relacionado com a abordagem Collaborative PL-Science no sentido de procurar promover interações entre os cientistas como forma de apoiá-los no seu ambiente de trabalho compartilhado. Entretanto, os autores não abordam a utilização de Linha de Produtos de Softwares e elementos de colaboração no cenário no qual estão trabalhando. Diferente ainda do trabalho de IVANOV et al. (2013), o interesse da Collaborative PL-Science está em oferecer apoio aos cientistas, gerando oportunidades de interação entre eles no momento de concepção de novos *workflows* científicos e não durante a sua execução.

Outros trabalhos já abordam especificamente a utilização de Linha de Produtos de Softwares (FILHO et al., 2012), (COSTA et al., 2013), e seguem relacionado com este trabalho.

FILHO et al. (2012) apresentam uma abordagem para enriquecimento semântico de LPS usando uma ontologia de topo que especifica conceitos genéricos e as relações em uma LPS. Os autores relacionam esta ontologia com o modelo de *features* através da realização de um mapeamento semiautomático com o propósito de obter um enriquecimento semântico na aplicação. Entretanto, este trabalho não está diretamente relacionado a um domínio específico, sendo apresentado de forma mais genérica. Neste trabalho os autores não utilizam nenhum histórico ou "*rationale*" para permitir que outros pesquisadores possam, no futuro, tomar conhecimento sobre como os artefatos utilizados no domínio foram empregados, dificultando a sua reutilização na abordagem. Com isso, ao se encerrar a utilização destes artefatos na LPS, todo o contexto no qual eles foram empregados é perdido, dificultando a apresentação e utilização de informações que possam orientar os participantes da abordagem durante a realização de suas atividades.

COSTA et al. (2013), conforme já detalhado, apresentam uma abordagem para conectar ontologia ao modelo de *features* do núcleo de uma LPS com o propósito de oferecer mais semântica para a concepção de *workflow* científico. Os autores procuram demonstrar que através da associação entre estes modelos, informações adicionais e relevantes para a área de Linha de Produtos de Software Científico podem ser fornecidas. Para isso, apresentam uma abordagem denominada PL-Science, a qual tem o propósito de apoiar a especificação e a condução de experimentos científicos em Bioinformática, através da associação de modelo de *features* e ontologia. Nesta pesquisa, investiga-se que utilizando os modelos de *features* e ontologias, muitas informações relevantes para o domínio continuam sendo perdidas ou mesmo não são fornecidas para os cientistas. Com isso, as ações realizadas na concepção de um *workflow* científico, por exemplo, são feitas baseadas apenas no conhecimento e experiência do cientista. A ausência de informações de percepção neste contexto, dificulta a criação de oportunidades de interação entre os pesquisadores, desfavorecendo e dificultando ainda, a colaboração efetiva entre esses cientistas. Tudo isso, sem falar na dificuldade para reutilização dos artefatos do núcleo da LPSC, já que COSTA et al. (2013) não consideram informações sobre como os artefatos foram utilizados. Por exemplo, um modelo de *features* pode ser combinado a diferentes ontologias e com isso permitir a criação de outros produtos. Oferecer este tipo de informação para os cientistas contribui também com o trabalho dos demais pesquisadores, mantendo-os conscientes sobre as ações que estão sendo realizadas pelo grupo.

A abordagem Collaborative PL-Science, proposta nesta dissertação, visa enriquecer os artefatos presentes no núcleo da LPS através da utilização de elementos de colaboração (informações de percepção, contexto e um mecanismo de suporte à comunicação). Com isso, espera-se oferecer mais semântica para o domínio científico e gerar oportunidades de interação entre os cientistas. A falta de um serviço para capturar e oferecer informações no contexto de *e-Science*, permite que a experiência do cientista seja perdida, pois o conhecimento a ela associado não foi persistido em uma memória de grupo, por exemplo.

A utilização de informações de percepção e contexto em banco de dados já foi tratada previamente na literatura de Sistemas Colaborativos (BORGES et al., 2000), (VIEIRA et al., 2004). Entretanto, esses trabalhos não abordam artefatos inerentes à LPS para o domínio de *e-Science*. BORGES et al. (2000) apresentam requisitos associados a Banco de Dados voltados para Sistemas

Colaborativos. Este trabalho auxiliou na identificação de requisitos para serem tratados na memória de grupo que compõem a abordagem Collaborative PL-Science. No trabalho de VIEIRA et al. (2004) é apresentado um mecanismo, denominado ARIANE, cuja a proposta é apoiar a percepção sobre as ações executadas em um Sistema Gerenciador de Banco de Dados (SGBD). Neste trabalho, as informações de percepção são capturadas pela aplicação durante a sua execução. Depois, elas são armazenadas em um repositório e exibidas aos usuários como forma de apoiar a realização das suas ações. As informações são capturadas de forma a responder as questões levantadas pelo Framework 5W+1H (GUTWIN & GREENBERG, 2002). Desta forma, consultas complexas podem ser realizadas no banco de dados de forma a oferecer informações mais relevantes aos participantes de atividades colaborativas. Na abordagem Collaborative PL-Science é apresentado um serviço para capturar e armazenar o histórico das interações entre os cientistas com os artefatos do núcleo da LPS, bem como as decisões tomadas durante a concepção de *workflows*. Um banco de dados relacional foi modelado para responder às questões do Framework 5W+1H. A partir das informações coletadas são apresentadas, aos cientistas, informações de percepção de forma que uma interação possa ser iniciada. A interação é capturada e persistida na memória de grupo. O Collaborative PL-Science permite também que consultas mais complexas possam ser realizadas na memória de grupo, por exemplo, conhecer o todo o processo que um cientista cumpriu para realizar o seu experimento. Para isso, utiliza uma ontologia que descreve o contexto da abordagem para possibilitar estas consultas.

ZHANG et al. (2014) apresentam técnicas relacionadas a aspectos de Colaboração para apoiar o gerenciamento de proveniência e reprodutibilidade em *workflows* científicos. Baseada em uma ontologia de colaboração científica, os autores propõem um modelo de colaboração orientada a serviços suportados por um conjunto de protocolos combináveis de colaboração. A proposta é que estes protocolos sejam aplicados para apoiar os cientistas durante a composição colaborativa de *workflows*. Para isso, os autores apresentam uma ferramenta para a concepção e desenvolvimento destes produtos, denominada CONFUCIUS.

Na abordagem proposta nesta dissertação, uma ontologia que descreve o contexto no qual a abordagem está inserida foi desenvolvida e é utilizada como parte da memória de grupo. O objetivo principal da ontologia é possibilitar que os cientistas possam realizar a descoberta de outros conhecimentos sobre o domínio, apoiando-os na realização da atividade de concepção de

workflows através de um Linha de Produtos de Softwares Científicos. Além disso, acredita-se que através da descoberta de novos conhecimentos a partir da ontologia da abordagem, oportunidades de colaboração e interação entre estes pesquisadores possam ser geradas, e apoiadas através de outros elementos de colaboração oferecidos pela Collaborative PL-Science, tais como, mecanismo de suporte à comunicação, apresentação de informações de percepção e contexto.

A Tabela 3 apresenta um resumo comparativo entre os trabalhos relacionados apresentados e a abordagem Collaborative PL-Science, considerando os seguintes aspectos: (i) se estão inseridos no domínio de e-Science ou se são tratados em outro domínio; (ii) se estes trabalhos abordam para o paradigma de LPS ou se utilizam alguma outra abordagem que trate o reúso de artefatos; (iii) se oferecem requisitos para a criação de uma memória de grupo; e por último, (iv) se abordam a utilização de elementos de colaboração em suas propostas.

Tabela 3. Comparação dos Trabalhos Relacionados

<i>Trabalhos Relacionados</i>	<i>e-Science</i>	<i>LPS</i>	<i>Memória de Grupo</i>	<i>Elementos de Colaboração</i>
Mattoso et al. (2009)	Sim	Não	Não	Não
Ivanov et al. (2013)	Sim	Não	Não	Não
Filho et al. (2012)	Não	Sim	Não	Não
Costa et al. (2013)	Sim	Sim	Não	Não
Vieira et al. (2004)	Não	Não	Sim	Sim
Borges et al. (2000)	Não	Não	Sim	Não
Zhang et al. (2014)	Sim	Não	Sim	Sim
Collaborative PL-Science	Sim	Sim	Sim	Sim

2.7 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO

Neste capítulo foram apresentados os principais conceitos relacionados ao desenvolvimento da abordagem Collaborative PL-Science proposta nesta dissertação. Foram apresentados conceitos, tais como Linha de Produtos de Softwares (LPS), destacando também os benefícios de utilizar esta

abordagem. Foram discutidos os principais tipos de percepção encontrados na literatura além das principais formas utilizadas para a representação de contexto, as quais apoiaram na seleção da representação de contexto através de ontologias. Aspectos de colaboração, bem como a forma pela qual eles se manifestam no contexto de *e-Science* também foram apresentados neste capítulo. Foi apresentada ainda, a abordagem PL-Science, proposta por COSTA et al. (2013), a partir da qual a Collaborative PL-Science é uma extensão. Por último, os trabalhos relacionados com a solução proposta para o problema levantado nesta dissertação, foram destacados.

No Capítulo seguinte será apresentada a abordagem Collaborative PL-Science, envolvendo a fase de análise, projeto e implementação da solução proposta.

3 A ABORDAGEM COLLABORATIVE PL-SCIENCE

Neste capítulo é apresentada a abordagem Collaborative PL-Science, que é uma evolução da PL-Science proposta por COSTA et al. (2013). A visão geral da abordagem proposta é apresentada na seção 3.1. Na seção 3.2, são apresentados os requisitos funcionais e não-funcionais especificados como parte da solução proposta nesta pesquisa. Na seção 3.3 a arquitetura da Collaborative PL-Science é detalhada, apresentando seus principais componentes: entre eles os elementos de colaboração utilizados e a memória de grupo. Na seção 3.4 são apresentados alguns aspectos de implementação da aplicação, bem como os testes funcionais que foram realizados. Por fim, são apresentadas as considerações finais deste capítulo.

3.1 VISÃO GERAL DA ABORDAGEM

Para ilustrar o cenário no qual a abordagem Collaborative PL-Science se aplica, é apresentada uma breve descrição do contexto que envolve os participantes da abordagem. O principal objetivo é apresentar as funcionalidades especificadas na abordagem, bem como detalhar sua utilização.

Apresentamos a seguir um cenário, envolvendo um *workflow* específico, relacionado ao domínio de bioinformática, mais especificamente ao subdomínio de sequenciamento genético, denominado *Pipeline PhredPhrap*. Neste contexto, os cientistas possuem, para especificar um dado *workflow*, apenas os artefatos persistidos no núcleo¹ da Linha de Produtos de Software Científico (LPSC) (modelos de *features*, ontologias, arquivos de mapeamento, entre outros). Assim, para a construção de um *workflow*, considerando o *Pipeline PhredPhrap*, por exemplo, um

¹ O núcleo de artefatos da abordagem é detalhado na subseção seguinte.

cientista recupera no núcleo de artefatos o modelo de *features* `SequenceAligningFeatureModel.xml`² para utilizá-lo em um dado experimento.

O *Pipeline PhredPhrap* é um *workflow* que compreende os seguintes passos: (i) ler os dados gerados por um sequenciador genético e identificar a sequência de DNA gerada, atribuindo valores de qualidade para cada posição nucleotídica identificada; (ii) realizar uma busca de regiões na sequência produzida que devem ser retiradas ou mascaradas; (iii) agrupar os resultados advindos do processo de leitura dos dados formando em sequências maiores (COSTA et al. 2013). A figura 4 ilustra os processos envolvidos neste workflow.

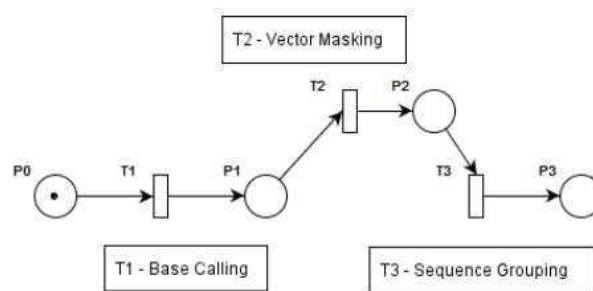


Figura 3.4: *Workflow* do produto *Pipeline PhredPhrap*.

Neste contexto, ou seja, na abordagem PL-Science, os *pontos de variação*³ que devem ser escolhidos no modelo de *features*, ficam a cargo da experiência que o cientista adquiriu durante a composição de outros produtos semelhantes ao *Pipeline PhredPhrap*. Portanto, o cientista desenvolve o *workflow* científico baseado apenas no seu conhecimento. Nenhum histórico, ou *rationale*, é gerado neste ambiente, ficando o conhecimento sobre as funcionalidades do artefato `SequenceAligningFeatureModel.xml` (modelo de *features*), as experiências do cientista e as tomadas de decisões por conta do usuário.

² Disponível em: <http://plscience.superdignus.com/SequenceAligningFeatureModelv1b.xml>

³ Pontos de Variação: são pontos nos quais as características dos artefatos da linha de produto de software podem sofrer variações (alterações).

Considerando estas dificuldades, o estudo apresentado nesta dissertação busca minimizar as restrições encontradas na abordagem PL-Science no que tange aos aspectos de colaboração. Para isso, este trabalho propõe a extensão da arquitetura de COSTA et al. (2013) considerando a utilização de elementos de colaboração e a criação de uma memória de grupo na abordagem. Tudo isso, através do desenvolvimento e conexão de alguns serviços responsáveis por apoiar os cientistas na realização das suas atividades de composição de *workflows* científicos. A figura 5 ilustra o escopo deste trabalho, destacando o foco do estudo.

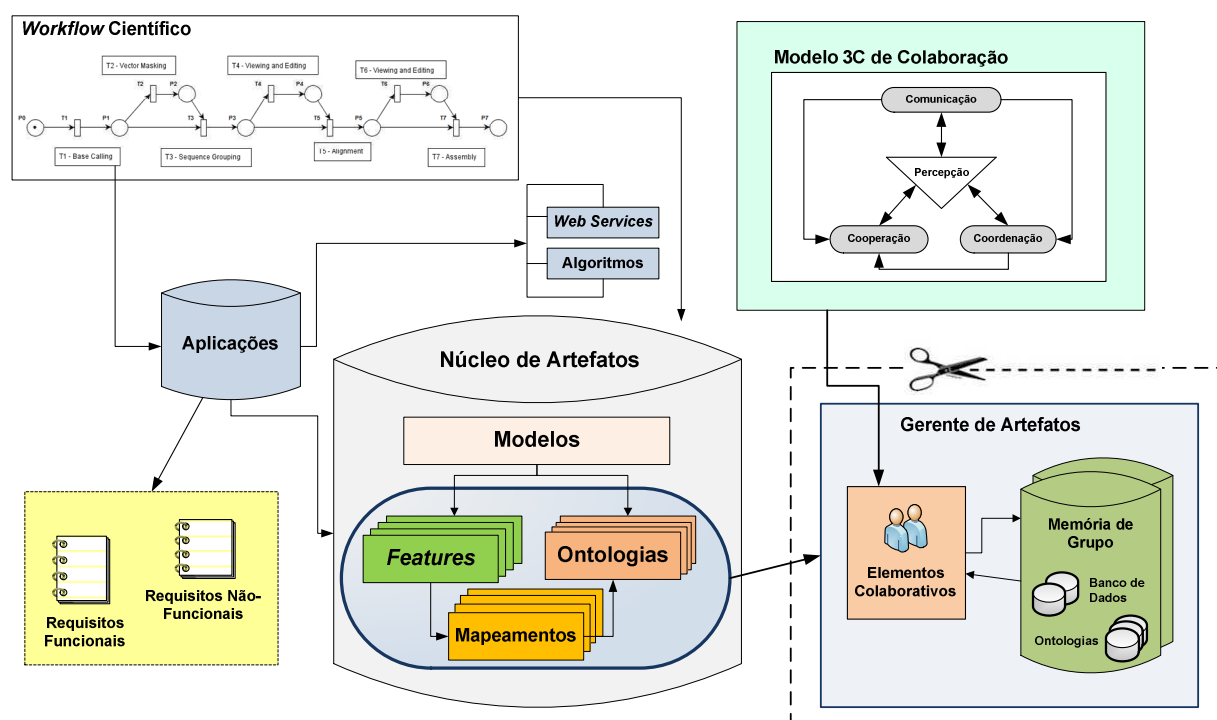


Figura 3.5: Visão Geral da Collaborative PL-Science - adaptada de (PEREIRA et al., 2013)

A partir da representação de um *workflow*, várias aplicações semelhantes ou distintas podem ser geradas. Estas podem ser compostas por diversos algoritmos, e formam o produto final da LPSC. Os produtos gerados são construídos a partir de modelos, os quais são armazenados no núcleo da LPSC. Dentre estes modelos, temos

os modelos conceituais, com destaque, no contexto deste trabalho, para os modelos de *features*, as ontologias e arquivos de mapeamentos (COSTA et al., 2013) (FILHO et al., 2012). Cada arquivo de mapeamento está associado a um modelo de *features* e a uma ontologia (Figura 5). A ontologia armazenada no núcleo de artefatos é utilizada para descrever o conhecimento sobre o domínio. Para ilustrar a aplicabilidade deste estudo, uma ontologia de alinhamento de sequência está sendo utilizada como artefato no núcleo para auxiliar o desenvolvimento de novos *workflows*. Este estudo busca enriquecer estes modelos, através da utilização de elementos de colaboração com o objetivo de oferecer mais semântica para o domínio. Acredita-se que, através do entendimento de como um *workflow* foi criado, é possível que os cientistas e/ou desenvolvedores tenham consciência dos requisitos funcionais (RF) e/ou requisitos não funcionais (RNF) que compõem a aplicação desenvolvida. Por exemplo, as *features* selecionadas para o desenvolvimento do produto *Pipeline PhredPhrap* são parte dos seus requisitos funcionais. Caso um cientista priorize o RNF desempenho neste *workflow*, ele pode optar por utilizar outro algoritmo em relação àquele que foi utilizado por outro cientista que também desenvolveu este produto.

Além disso, todo o rastro deixado pelos cientistas durante a realização das atividades colabora para que outros desenvolvedores possam criar suas aplicações coordenando uma base de conhecimento comum. Como resultado, novas oportunidades são geradas para que informações implícitas sobre o domínio, ou mesmo sobre o produto sejam ressaltadas. Mais ainda, resultados previamente obtidos podem ser reutilizados possibilitando a redução do custo do experimento.

Considerando o *workflow Pipeline PhredPhrap* e o funcionamento da abordagem proposta, podemos destacar que, durante a escolha dos algoritmos ou *web services* que irão compor a aplicação, a abordagem informa ao cientista os algoritmos que podem ser escolhidos. Ao visualizá-los, um dado cientista A, como conhecedor do domínio, pode entender por experiência própria que existem outros algoritmos que também podem ser utilizados para a realização de uma dada tarefa no *workflow*. Por

exemplo: ao invés de utilizar o *web service* runPhrap⁴ para compor o *workflow*, ele percebe a existência de outro algoritmo ou *web service* que pode apresentar um melhor resultado. Atualmente as informações sobre algoritmos e *web services* sugeridos para os cientistas na concepção de um produto estão armazenadas como indivíduos na ontologia que está sendo utilizada para a concepção do *workflow*. Neste contexto, uma interação com outros cientistas, que trabalham, ou trabalharam, compondo o produto *Pipeline PhredPharp*, pode ser iniciada. Através desta interação eles podem discutir se a inclusão deste novo indivíduo na ontologia que auxilia a composição do *workflow*, é ou não necessária.

3.1.1. Elementos de Colaboração

Os elementos identificados e apresentados pela Collaborative PL-Science são descritos abaixo, destacando as principais dimensões e relações envolvidas no **Modelo 3C** de colaboração, conforme representado por FUKS et al. (2007). A seguir os elementos de colaboração são detalhados no domínio científico:

i) Comunicação: O suporte à comunicação proposto pela Collaborative PL-Science ocorre através de um mecanismo semelhante a um *chat*. Uma das vantagens deste suporte é a persistência das mensagens. O suporte à comunicação nesta abordagem é um meio para os cientistas trocarem experiências ou esclarecerem possíveis dúvidas sobre um ponto específico. Esta atividade contribui com o trabalho dos demais participantes mantendo-os conscientes sobre as ações realizadas pelo grupo. A ausência deste mecanismo na abordagem PL-Science faz com que os cientistas trabalhem e tomem decisões amparadas em seu próprio conhecimento sobre o domínio. Essa ausência desfavorece que interações espontâneas entre os pesquisadores sejam iniciadas, e dificulta a colaboração entre eles, conforme pode ser evidenciado nos exemplos citados anteriormente.

⁴ Disponível em: <http://www.biocatalogue.org/services/2268>.

ii) Cooperação: Nessa proposta, os passos deixados pelos cientistas ao realizarem uma tarefa possibilita que outro cientista siga passos semelhantes, ou mesmo tome uma decisão a partir de um passo já realizado por outro participante. Essas operações são capturadas e persistidas pelo Gerente de Artefatos, durante o momento em que um cientista trabalha na aplicação. Por exemplo, a partir dos rastros deixados por um cientista A que já construiu o produto *Pipeline PhredPhrap*, um cientista B pode conhecer o processo pelo qual o produto foi desenvolvido, bem como as decisões tomadas, algoritmos que foram utilizados, notas realizadas pelo cientista A, entre outros. Em caso de uma possível dúvida ou falta de experiência por parte do cientista B na composição deste produto, ele pode interagir com o cientista A.

iii) Coordenação: Nessa proposta, o objetivo é construir *workflows* científicos, embora possa existir interesse na composição de um produto específico. A lista dos cientistas que já desenvolveram um dado produto, por exemplo, é aberta a todos os participantes. Através de elementos de visualização, notificações são exibidas para o cientista para que ele coordene as suas atividades apoiado nos passos realizados por outros membros do grupo.

iv) Percepção: Quando a ação executada pelo cientista é parecida com outra ação já realizada, por exemplo, o cientista C inicia o desenvolvimento do *workflow* Pipeline PhredPhrap, atividade essa similar a um desenvolvimento anterior. Ao identificar esta ação, a aplicação pode oferecer informações de percepção e contexto ao cientista através de um quadro de informações, como, quais foram os cientistas que já desenvolveram este produto, quais os algoritmos e/ou *web services* que foram escolhidos, quais as anotações e decisões tomadas pelo cientista, entre outras. Estas informações têm o objetivo de contextualizar os participantes que atuam ou atuaram em um experimento científico parecido, apoiando-os na realização de seus experimentos. Essas informações de percepção são geradas a partir da captura dos passos deixados pelos cientistas na realização das suas atividades, respondendo às questões do Framework 5W+1H (GUTWIN & GREENBERG, 2002). Os tipos de informação que foram definidas inicialmente para compor este trabalho são: a percepção social, a percepção no espaço de trabalho e a percepção de atividade

(VIEIRA, TEDESCO & SALGADO, 2011). Estes tipos foram escolhidos por oferecerem a possibilidade de apresentar informações que possam apoiar a interação entre os cientistas, minimizando a sobrecarga de informações neste ambiente. Além disso, a partir dos tipos de percepção escolhidos, foram definidas as informações que serão apresentadas no ambiente compartilhado de trabalho. São elas: (i) presença do cientista no espaço de trabalho; (ii) disponibilidade para indicar se o momento é adequado para iniciar uma comunicação síncrona; (iii) quem foi o cientista e quando ele executou uma determinada ação no espaço de trabalho; (iv) participantes que já trabalharam em determinada atividade; (v) decisões tomadas e comunicações realizadas durante a realização da atividade; (vi) contexto dos artefatos (onde foi utilizado, por exemplo); (vii) notas realizadas sobre o artefato e/ou *features* do modelo utilizado.

3.1.2. Memória de Grupo

Outro elemento que pode ser encontrado na Collaborative PL-Science é a Memória de Grupo. Trata-se de um banco de dados relacional e de uma ontologia de aplicação modelados para persistirem os históricos de atividades dos cientistas, capturados durante a realização das suas atividades de composição de *workflows* científicos. O banco de dados relacional desempenha ainda o papel de prover as informações de percepção para os cientistas através de elementos de visualização. A ontologia é um recurso pelo qual os cientistas podem realizar consultas semânticas sobre o ambiente compartilhado de trabalho no qual está inserido. A especificação de uma ontologia neste contexto é relevante por se tratar de um domínio com conhecimentos tácitos e explícitos por parte dos pesquisadores envolvidos. É também a oportunidade de representar o contexto no qual os participantes da abordagem fazem parte.

Por fornecer um vocabulário comum para a representação do conhecimento, a utilização desta ontologia evita que os cientistas façam interpretações ambíguas do contexto. Além disso, através do vocabulário, das regras, e propriedades especificadas, é possível estruturar e representar o conhecimento sobre o ambiente compartilhado

de trabalho de maneira mais formal, permitindo a descoberta de conhecimento implícito derivado das relações e regras ontológicas. Portanto, através deste conhecimento, outras oportunidades de interação entre estes pesquisadores podem ser geradas. Com isso, a memória de grupo pode auxiliar os cientistas, por exemplo, na decisão se esse seria mesmo o recurso apropriado para compor o *workflow*, como, por exemplo, considerando o exemplo dado anteriormente, na análise da decisão do cientista "A" em escolher o *web service* runPhrap (<http://www.biocatalogue.org/services/2268>) em detrimento de outro, poupando tempo do cientista "B" na concepção e execução do novo *workflow* científico. Assim, através deste recurso, a abordagem Collaborative PL-Science gera ainda uma oportunidade de interação entre os cientistas tornando mais clara a compreensão do contexto. É também importante ressaltar que o uso de uma ontologia nesta abordagem, diminui o esforço necessário para obter conhecimento sobre o domínio, já que, através da sua utilização, a especificação de consultas complexas é facilitada.

A ontologia da Collaborative PL-Science tem o papel de descrever os conceitos envolvidos na abordagem e pode ser considerada uma ontologia de aplicação, representando conhecimento do domínio de uma Linha de Produtos de Software Científico (LPSC) e do domínio de colaboração. É um mecanismo importante para auxiliar os cientistas na aquisição de conhecimento ao longo da realização das atividades. O objetivo de se utilizar uma ontologia nesta proposta é a oportunidade de representar o contexto no qual os indivíduos, participantes da abordagem, estão envolvidos. Abaixo são apresentadas algumas vantagens que podem justificar a utilidade de uma ontologia para este domínio:

- Oportunidade de compartilhamento do conhecimento entre agentes humanos e de software, uma vez que a ontologia criada representa uma compreensão compartilhada sobre um conjunto de termos e descrições sobre o domínio da abordagem Collaborative PL-Science, e descreve de forma precisa as propriedades deste ambiente e os vários conceitos utilizados. Esta oportunidade permite, por exemplo, minimizar a falta de semântica na busca e recuperação de um artefato no núcleo da LPSC, agregando conhecimento durante a realização de uma atividade.

Outro ponto é perda de informações relevantes para o domínio, devido a não representação / explicitação do conhecimento por parte dos cientistas, que pode ser minimizada através da especificação de novas regras e novos termos sobre o domínio.

- Oportunidade de integração e reuso do conhecimento compartilhado pela ontologia Collaborative PL-Science. Esse tópico permite que diferentes ontologias, armazenadas em diferentes locais e criadas por diferentes autores, possam ser associadas, integradas ou mesmo conectadas com esta ontologia e reutilizadas para compor uma ontologia em larga escala sem a necessidade de começar do zero. Outra vantagem ainda, está na oportunidade de se reutilizar, de forma menos custosa, o conhecimento associado à concepção de um *workflow* científico, por exemplo.

3.2 ESPECIFICAÇÃO DE REQUISITOS

A partir da visão geral apresentada e da análise da proposta anterior desenvolvida por COSTA et. al. (2013), algumas questões foram levantadas, considerando a necessidade de colaboração efetiva entre os cientistas no desenvolvimento de novas aplicações através da LPSC.

A partir destas questões e considerando uma evolução necessária da abordagem PL-Science para melhor suporte à colaboração entre os cientistas, foi possível definir as funcionalidades necessárias para a solução apresentada nesta dissertação. Com isso, requisitos funcionais e não-funcionais puderam ser especificados com o intuito de identificar claramente a evolução necessária na abordagem PL-Science.

Os requisitos funcionais tratados na solução proposta são apresentados a seguir:

- (i) Oferecer aos cientistas, usuários da aplicação, um mecanismo que possa prover informações sobre o contexto que envolve as atividades que

estão realizando, como por exemplo: quem realizou o experimento, quando, como foi realizado, decisões tomadas, entre outras;

- (ii) Oferecer funcionalidades que possibilitem que os cientistas possam interagir entre eles, trocando experiências e habilidades que podem ser úteis na concepção de *workflows científicos*;
- (iii) Oferecer mecanismos para minimizar a dificuldade para reutilização dos artefatos do núcleo da LPSC, considerando a ausência de informações sobre como o artefato foi utilizado, em quais produtos, por quais cientistas, entre outras; e ainda,
- (iv) Buscar minimizar a perda de semântica ao encerrar a utilização de um artefato, ou seja, oferecer um mecanismo que possa retratar o contexto no qual o artefato foi empregado por um cientista, permitindo consultas a este recurso para orientá-los durante a realização de suas atividades na aplicação.

Para a especificação da ontologia da abordagem Collaborative PL-Science, algumas questões foram formuladas para descrever o escopo e definir o contexto. Algumas das questões respondidas pela ontologia são:

- Quais são os recursos materiais e intelectuais necessários para compor um determinado *workflow*?
- Quais são os produtos ou artefatos resultantes da execução de uma determinada atividade?
- Quais produtos científicos têm sido construídos pelos cientistas?
- Quais características estão presentes em um *workflow*? Como foi construído?
- Quais interações foram geradas durante a realização de uma atividade? E quem eram os cientistas envolvidos?
- Que decisões foram tomadas ao desenvolver um *workflow*? (*Passos de construção deste produto + Notas + Mensagens*).

- Onde estão disponíveis os artefatos utilizados para a construção de um produto?
- Que atividades foram realizadas sobre o Artefato?
- Qual o objetivo do Produto?
- Quando a atividade foi realizada? Por Quem?
- Quem são os cientistas de uma determinada comunidade?
- Quem são os cientistas que já utilizaram um determinado artefato?

3.2.1 Requisitos Não Funcionais

Alguns requisitos não funcionais também foram estabelecidos para a abordagem Collaborative PL-Science, de modo a delimitar o escopo da proposta, uma vez que o problema, colaboração em e-Science é bastante amplo e complexo. Portanto, com o propósito de estabelecer atributos de qualidade para a abordagem proposta, os seguintes requisitos não funcionais foram especificados:

- *Escalabilidade:* a arquitetura da Collaborative PL-Science foi desenvolvida de maneira a possibilitar a inclusão de outros serviços, como, por exemplo serviços de colaboração, serviços de suporte à colaboração, serviços de ontologias, entre outros, com o objetivo de permitir uma melhor qualidade no apoio computacional oferecido as cientistas durante a realização da atividade de concepção de *workflows* científicos.
- *Portabilidade:* a arquitetura proposta tem como um dos requisitos a possibilidade de substituir os serviços oferecidos atualmente por outros. A utilização de novos serviços pode agregar outras funcionalidades para a abordagem, ou mesmo permitir a migração de um serviço antigo para outro que apresente um melhor resultado, por exemplo. Além disto, a proposta contribui para que outras formas de representação do contexto possam ser avaliadas futuramente.

- *Robustez*: os artefatos persistidos no núcleo de artefatos devem ser arquivos com extensão XML ou OWL - para artefatos que sejam ontologias. Caso um artefato não atenda a essa regra o mesmo será ignorado pela aplicação, ou seja, nesta versão inicial da arquitetura não serão suportados arquivos com outras extensões.
- *Reusabilidade*: a arquitetura da LPSC, os serviços desenvolvidos e os elementos de contexto (expressos através de ontologia) foram modelados de forma que possam ser reutilizados em futuras versões da arquitetura.
- *Interoperabilidade*: a solução proposta nesta dissertação busca a disponibilização de serviços independentemente da plataforma e tecnologia, permitindo a integração com outros serviços, aplicativos e sistemas legados.
- *Padronização*: A abordagem Collaborative PL-Science foi modelada e desenvolvida baseada no uso de padrões de projetos [LARMAN, 2008] e em melhores práticas de desenvolvimento de software [PRESSMAN, 2010].

3.3 ARQUITETURA DA SOLUÇÃO

Conforme já foi mencionado, a Collaborative PL-Science é uma evolução da PL-Science (COSTA et al., 2013). Sendo assim, os elementos presentes na abordagem PL-Science também estão representados na arquitetura da abordagem proposta nesta dissertação (Figura 6).

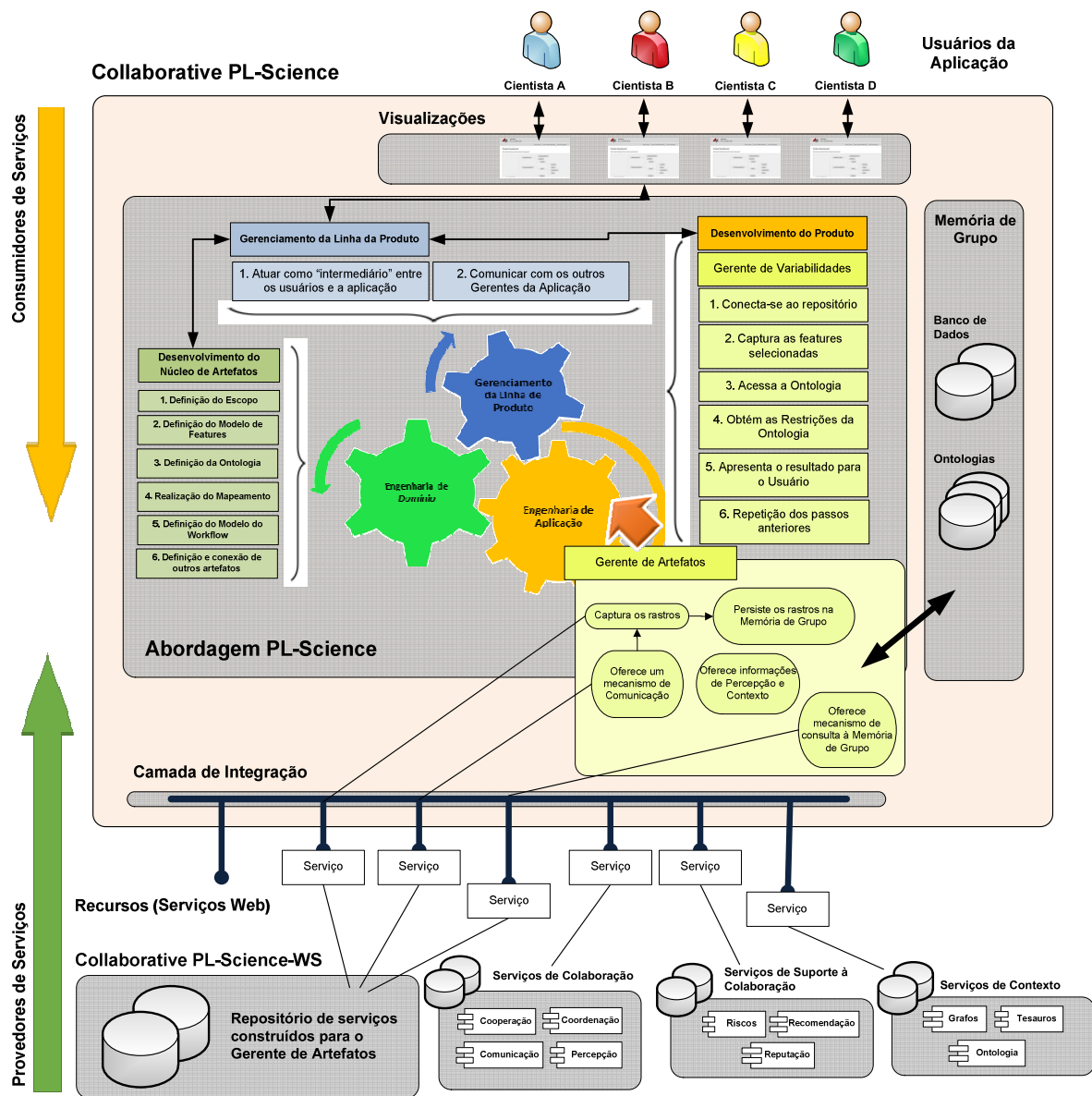


Figura 3.6: Visão da Arquitetura da Abordagem Collaborative PL-Science.

Com o propósito de atender aos requisitos especificados para a Collaborative PL-Science, a arquitetura da aplicação foi estruturada considerando o padrão arquitetural SOA – Service Oriented Architecture [HURWITZ et al. 2009], ou no português Arquitetura Orientada a Serviço, conjuntamente com uma arquitetura em camadas. Na primeira camada pode-se visualizar os usuários da abordagem, ou seja,

os cientistas. Eles podem estar geograficamente distribuídos e estão identificados por pseudônimos: Cientista A, Cientista B, Cientista C e Cientista D. Na segunda camada da arquitetura está detalhada a interface web da aplicação. Chamada de Visualizações, esta camada pode permitir que no futuro seja conectado qualquer outro elemento de visualização, como por exemplo, ferramentas de mineração de dados, de maneira a facilitar o trabalho dos cientistas. É através desta camada que os usuários podem interagir com a aplicação. Um exemplo desta interface é apresentado na Figura 7, destacando o núcleo de artefatos da aplicação [1] e o quadro de informações de percepção [2], que nesta tela apresenta as últimas atividades desenvolvidas pelos cientistas que estão conectados na aplicação.

[1]

Core Assets:			
Id	Type	Name	Purpose
1	FeatureModel	SequenceAligningFeatureModel	Modelo de Features sobre Alinhamento de Sequencia
2	MappingFile	SequenceAligningMapping	Mapear as features nas classes da Ontologia
3	Ontology	SequenceAligningOntology	Ontologia sobre alinhamento de Sequencia
4	FeatureModel	BioinformaticFeatureModel	Criar workflows para o domínio de Bioinformática
5	MappingFile	BioinformaticMappingFile	Mapear as features do modelo nas classes da Ontologia

[2]

Last Tasks Performed

TimeLine

Magnus Rbero UFJF
PRODUCT DEVELOPMENT
31/1/2014 - 20:23

Magnus Rbero UFJF
PRODUCT DEVELOPMENT
31/1/2014 - 14:10

Ontological Research

Federal University of Juiz de Fora - MG / Brazil - Institute of Mathematical Sciences
Department of Computer Science
Phone: +55(32)2102-3311 - E-mail: pls.cience.developers@gmail.com

UFJF NEnC

Figura 3.7: Interface de visualização do Núcleo de Artefatos.

A terceira camada apresenta o núcleo da abordagem PL-Science. Nela, estão presentes os módulos propostos por COSTA et al. (2013), bem como os processos envolvidos na atividade de concepção de *workflows* científicos. Ainda nesta camada estão disponíveis as rotinas propostas pela Collaborative PL-Science (seção 3.3.1).

Além destas três camadas, temos a camada da memória de grupo da aplicação, englobando o banco de dados relacional e a ontologia de contexto da abordagem. Os projetos físico e lógico do banco de dados relacional são apresentados e discutidos na seção 3.3.2.1. A estrutura da ontologia, bem como mais detalhes sobre a sua contribuição para este domínio são detalhados na seção 3.3.2.2.

Com o intuito de atender aos requisitos não funcionais especificados para a proposta apresentada, a última camada permite que novos serviços possam ser acoplados na abordagem sendo, portanto, a camada de integração de serviços, permitindo a conexão dos processos que são realizados pelos cientistas através do núcleo da abordagem. Para isso, foi prevista a existência de um *Enterprise Service Bus* (ESB), conhecido como barramento de serviços, para cuidar de aspectos tais como governança, políticas de segurança, entre outros. É importante destacar que não faz parte do escopo deste trabalho a implementação do ESB.

3.3.1. Processos adicionados pela Collaborative PL-Science

A seguir (Figura 8) são detalhadas as principais atividades relacionadas ao processo de oportunidades de interação e colaboração gerada entre os pesquisadores pela abordagem Collaborative PL-Science.

Passo 1: Captura o histórico das atividades realizadas pelos cientistas na aplicação. A criação deste histórico se inicia no momento em que um cientista realiza a atividade de especificação de um novo produto.

Passo 2: Consiste no armazenamento dos dados que foram capturados na memória de grupo. Estes dados são armazenados primeiramente no banco de dados relacional e depois uma carga de dados é realizada na ontologia, que também faz parte da memória de grupo.

Passo 3: Através de consultas pré-definidas, que são executadas no banco de dados relacional que compõe a memória de grupo, informações de percepção e contexto são apresentadas para os usuários da aplicação. Isso permite que os usuários

sejam contextualizados sobre o que acontece ao seu redor no ambiente compartilhado de trabalho no qual está inserido.

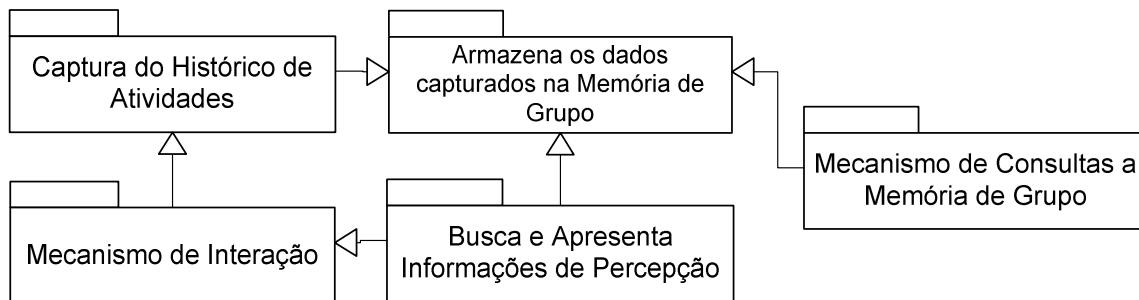


Figura 3.8: Processos envolvidos no Gerente de Artefatos.

Passo 4: Um mecanismo é oferecido aos cientistas como forma de gerar oportunidades de interação entre eles. A partir deste mecanismo, os cientistas podem interagir sobre a atividade em questão, trocando experiências, esclarecendo dúvidas, ou mesmo sugerindo a correção de artefatos, inclusão de novos algoritmos para a realização de uma tarefa, entre outros. Toda a interação gerada também é armazenada na memória de grupo, possibilitando que em uma consulta futura, um cientista possa conhecer as notas realizadas, os artefatos utilizados durante aquela tarefa, bem como todos os cientistas que já interagiram em relação à atividade que está ou estava sendo realizada.

Passo 5: Como última atividade, está o mecanismo que permite a realização de consultas à memória de grupo. Através deste mecanismo, consultas complexas podem ser realizadas na ontologia que é parte integrante da memória de grupo. Como resultado, os cientistas podem ter acesso a informações mais detalhadas sobre o conhecimento que compõe o domínio no qual estão inseridos.

3.3.2 Memória de Grupo

3.3.2.1 Banco de Dados

Nesta seção são apresentados os projetos lógico e físico do banco de dados. No projeto lógico são apresentadas as tabelas, os relacionamentos, regras, entre outros. A Figura 9 ilustra o esquema do banco de dados. Por questão de espaço e para melhor visualização da figura alguns campos da tabela foram omitidos, sendo apresentados no quadro 1.

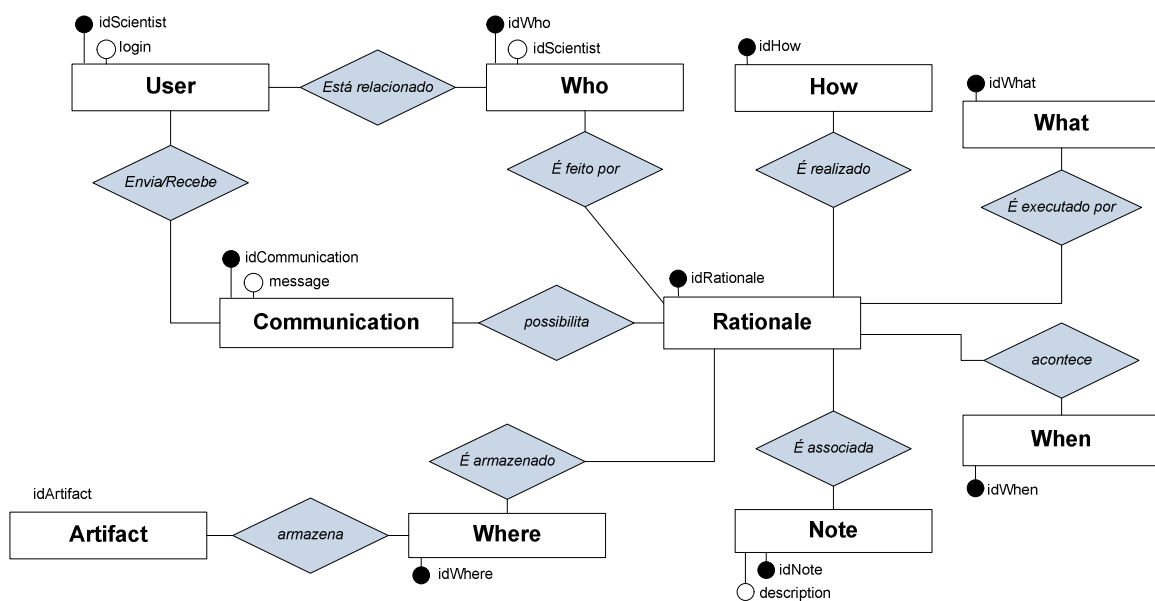


Figura 3.9: Projeto Lógico do Banco de Dados.

Quadro 1: Modelo do Projeto Lógico do Banco de Dados.

<p>User (<u>idScientist</u>, login, password, NameScientist, CPFScientist, emailScientist, institutionScientist, status)</p> <p>Rationale (<u>idRationale</u>, idWho, idHow, idWhen, idWhat, idWhere, idNewRationale) idWho <i>referencia</i> idWho <i>tabela</i> Who idHow <i>referencia</i> idHow <i>tabela</i> How idWhen <i>referencia</i> idWhen <i>tabela</i> When idWhat <i>referencia</i> idHow <i>tabela</i> What idWhere <i>referencia</i> idHow <i>tabela</i> Where</p> <p>Who (<u>idWho</u>, idScientist) idScientist <i>referencia</i> idScientist <i>tabela</i> User</p> <p>How (<u>idHow</u>, description)</p> <p>When (<u>idWhen</u>, day, month, year, time, minute)</p> <p>What (<u>idWhat</u>, description, productName, purposeProduct)</p> <p>Where1 (<u>idWhere</u>)</p> <p>Note (<u>idNote</u>, description, subfeatures, idRationale) idRationale <i>referencia</i> idRationale <i>tabela</i> Rationale</p> <p>Communication (<u>idCommunication</u>, message, idRationale, subfeature, idScientistReceptor, idScientistIssuing, date, time, lidaSN) idRationale <i>referencia</i> idRationale <i>tabela</i> Rationale idScientistReceptor <i>referencia</i> idScientist <i>tabela</i> User idScientistIssuing <i>referencia</i> idScientist <i>tabela</i> User</p> <p>Artifact (<u>idArtifcat</u>, name, purpose, local, status, type)</p>

A Figura 10 apresenta o projeto físico do banco de dados relacional, desenvolvido como parte da solução proposta nesta dissertação.

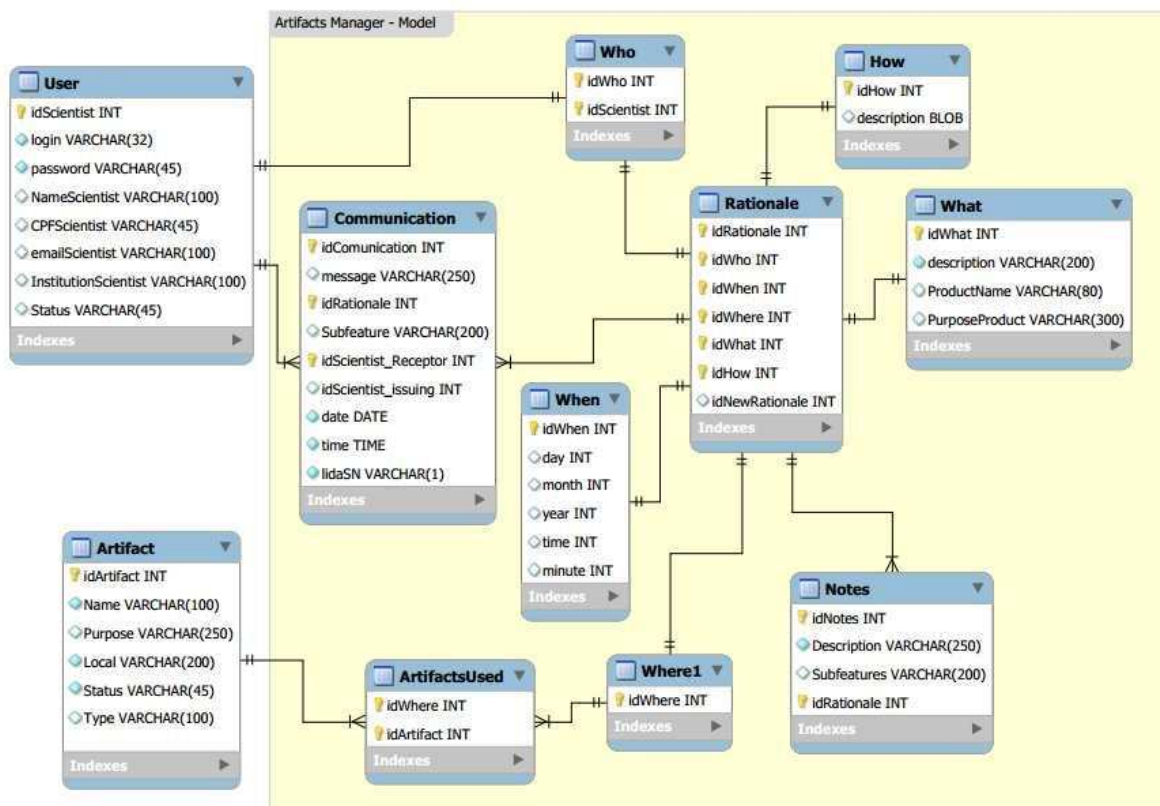


Figura 3.10: Projeto Físico do Banco de Dados.

A modelagem e a utilização deste banco de dados relacional foram importantes para que a solução alcançasse os seus objetivos, já que ele foi desenvolvido para armazenar os dados necessários para gerar as informações de percepção para contextualizar os cientistas que trabalham na LPSC.

3.3.2.2 Ontologia Collaborative PL-Science

A partir das questões especificadas para a ontologia Collaborative PL-Science, a mesma foi desenvolvida com foco no apoio ao desenvolvimento de *workflows* científicos. Sendo assim, os termos apresentados na ontologia, são relacionados a atividades de desenvolvimento de workflows. Desta forma, são através das atividades que novos artefatos podem ser inseridos no núcleo da LPSC, editados pelos cientistas,

ou mesmo escolhidos para serem utilizados na concepção de novos *workflows* científicos, por exemplo (Figura 11).

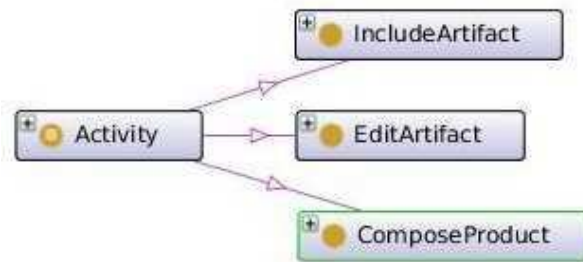


Figura 3.11: Tipos de atividades modeladas na ontologia.

As atividades necessitam de recursos para serem desempenhadas. Podemos dizer que neste caso um dos recursos necessários e utilizados em uma atividade é o artefato. Por artefato, entendemos como sendo os modelos que compõe o núcleo da LPSC. Além dos modelos clássicos de uma LPS, nesta dissertação o enfoque está voltado para os modelos de *features*, arquivos de mapeamentos e ontologias, conforme já foi mencionado. Para isso, os artefatos armazenados são enquadrados em algum dos tipos mencionados acima, facilitando a sua classificação e a sua manipulação dentro da abordagem proposta (Figura 12).

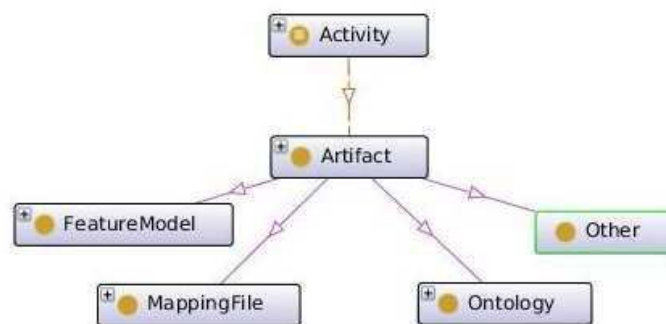


Figura 3.12: Tipos de Artefatos modelados na ontologia.

Outro tipo para definir os artefatos também pode ser utilizado, como por exemplo, o tipo “outros”, que podem ser modelos UML, por exemplo. Os artefatos utilizados pela abordagem Collaborative PL-Science podem estar armazenados em repositórios distintos ou em outros servidores. Para isso, o local em que estes artefatos estão disponíveis são modelados também nesta ontologia. Como exemplo, temos os artefatos que são utilizados no cenário de uso desta pesquisa, apresentados na Tabela 4.

Tabela 4. Localização dos artefatos utilizados pela Collaborative PL-Science

<i>Nome do Artefato</i>	<i>Localização:</i>
SequenceAligningFeatureModel	http://plscience.superdignus.com/SequenceAligningFeatureModelv1b.xml
SequenceAligningMapping	http://plscience.superdignus.com/SequenceAligningMappingv1b.xml
SequenceAligningOntology	http://plscience.superdignus.com/SequenceAligningOntologyv1b.owl

Um cientista é um participante importante neste domínio e o responsável por manipular os artefatos do núcleo e conseqüentemente realizar a especificação dos produtos científicos. Os cientistas estão associados a uma comunidade, que pode ser um centro de pesquisas, como por exemplo, FIOCRUZ e a EMBRAPA, ou uma Universidade, por exemplo, UFJF (Figura 13).

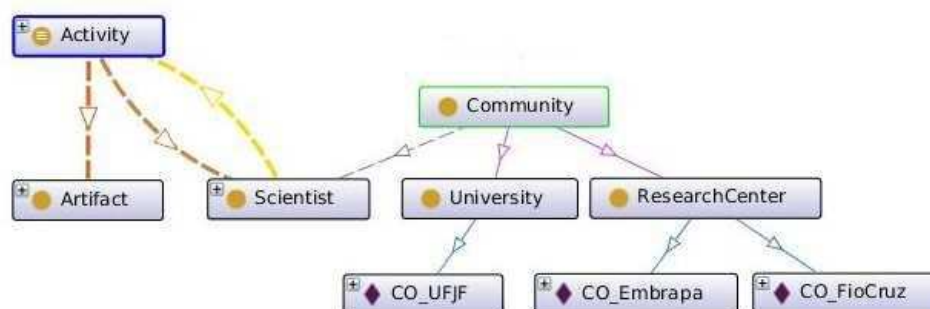


Figura 3.13: Cientistas e Comunidades modelados na ontologia.

Durante a realização de uma atividade na abordagem Collaborative PL-Science, os cientistas podem interagir, seja para trocar habilidades, competências ou esclarecerem possíveis dúvidas. Esta interação ocorre através da troca de mensagens entre os pesquisadores. As mensagens trocadas estão relacionadas com a atividade que estava sendo desenvolvida naquele momento (composição de um produto, edição de um artefato, ou mesmo a inclusão de um novo artefato no núcleo da LPSC). Logo as mensagens são escritas por cientistas e também endereçadas a outro cientista. Para isso, os cientistas estão representados da seguinte maneira nesta ontologia: *Desenvolvedores* – pesquisadores que manipulam os artefatos para a concepção de um produto, *emissores* – cientistas que enviam novas mensagens e *receptores* – os cientistas que recebem as mensagens enviadas.

Notas também podem ser feitas pelos pesquisadores durante a realização das atividades. Estas notas são opiniões expressas por estes participantes ou mesmo anotações sobre decisões importantes que tomaram durante a especificação de um produto, por exemplo. Outro exemplo para justificar a utilização de notas pelos cientistas, pode ser a necessidade de argumentar ou mesmo manter registrado a edição de um artefato do núcleo. As notas carregam também a identificação do cientista que a escreve, e o artefato relacionado (Figura 14).

Por último está o produto gerado a partir de uma atividade (composição de produto). É construído no mínimo por 3 artefatos (modelo de *features*, arquivo de mapeamento e uma ontologia). Um dado importante sobre o produto é o seu histórico de construção. Este histórico carrega os passos dados pelo cientista durante a realização da atividade que o construiu. A partir destes passos outros participantes podem tomar consciência sobre como um produto foi construído e se apoiar neste histórico para a construção de novos produtos, minimizando tempo e o custo na concepção de novos *workflows* científicos, por exemplo (Figura 15).

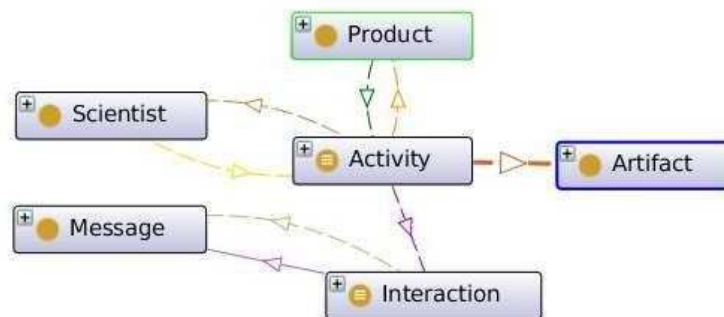


Figura 3.14: Interações modeladas na ontologia.

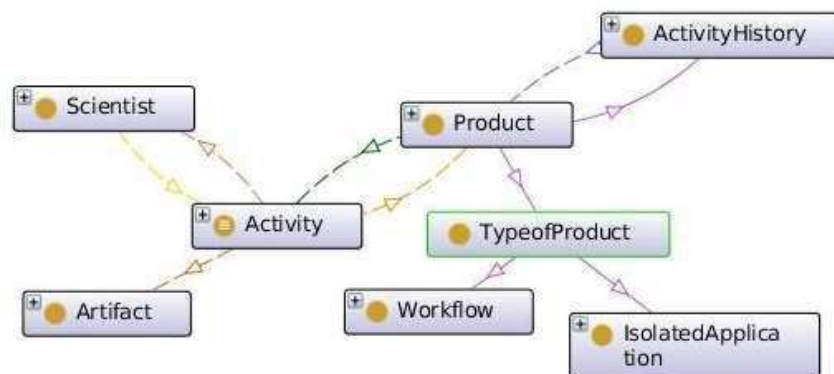


Figura 3.15: Produtos modelados na ontologia.

A Figura 16, apresenta as classes modeladas para esta ontologia (indicador 1). Além disso traz também os indivíduos que a compõem (indicador 2). Utilizando a máquina de inferência *Pellet*, podem ser vistas as inferências apresentadas para o produto *Pipeline PhredPhrap*, o qual foi utilizado como exemplo para ilustrar as inferências feitas na ontologia (indicador 3).

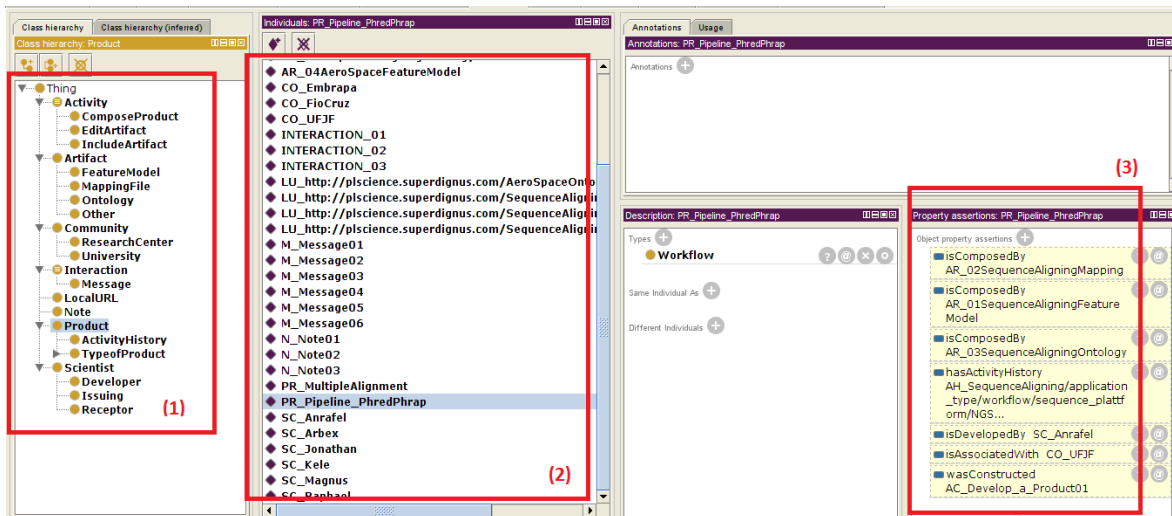


Figura 3.16: Collaborative PL-Science Ontology - Ferramenta Protegé.

As regras e restrições utilizadas pela máquina de inferência foram especificadas através de *Property Chains*, um recurso disponível em OWL 2.0, que permite a especificação de restrições ontológicas mais complexas, sem a necessidade de se utilizar SWRL. Por exemplo, para que se pudesse conhecer os participantes que interagiram em uma atividade foram utilizadas as seguintes propriedades: uma interação *gera* Mensagem - Mensagem é escrita por Cientista - Mensagem *tem remetente* Cientista, logo interação *tem como participante* cientista. A partir desta *Property Chain* é possível conhecer a lista de participantes (cientistas que interagiram) de uma atividade. A Figura 17 apresenta as propriedades criadas para esta cadeia, e a Figura 18 apresenta o resultado inferido pela máquina de inferência *Pellet*.

produto é construído por um cientista. Por fim, a terceira *Property Chain* mostra que um artefato é manipulado por um cientista no momento da realização de uma atividade.

As propriedades inversas destas *Properties Chains* também foram modeladas na ontologia. Embora não tenham sido apresentadas neste trabalho, as propriedades inversas também são importantes neste cenário, principalmente por possibilitar que se conheça de forma rápida e formal o caminho inverso de um processo.

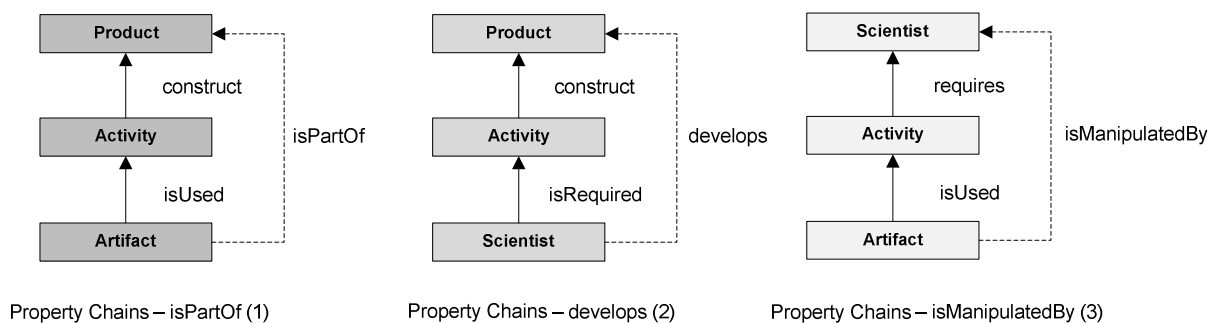


Figura 3.19: Exemplos de *Properties Chains* criadas.

Outra vantagem associada à utilização de *Properties Chains* está na composição das regras que serão utilizadas. Por exemplo, para a modelagem desta ontologia pode-se reutilizar muitas das propriedades escritas, criando assim novas cadeias para expressar as restrições deste domínio a partir de propriedades existentes, gerando novas *Properties Chains*.

3.4. IMPLEMENTAÇÃO

3.4.1 Desenvolvimento da Memória de Grupo

O desenvolvimento da memória de grupo buscou tratar principalmente os seguintes requisitos funcionais especificados para este trabalho: (i) Tratar a ausência de

informações sobre o contexto que envolve as atividades que os cientistas estão realizando; e (ii) oferecer um mecanismo que possa retratar o contexto no qual o artefato foi empregado por um cientista, permitindo consultas a este recurso para orientá-los durante as suas atividades.

Os principais requisitos não funcionais atendidos através do desenvolvimento da memória de grupo foram: escalabilidade, portabilidade e interoperabilidade.

Nesta dissertação, o banco de dados relacional foi implementado utilizando o MySQL, na versão 5.7 e a ontologia da abordagem foi implementada utilizando a linguagem OWL (*Ontology Web Language*) na versão 2. Esta versão foi escolhida por permitir trabalhar com *Property Chains* (Propriedades de Cadeia). A ferramenta utilizada para construção desta ontologia foi o Framework Protégé.

3.4.2 Aplicação Collaborative PL-Science

Para o desenvolvimento da aplicação Collaborative PL-Science levou-se em consideração a necessidade de atender aos requisitos funcionais e não funcionais especificados para a solução proposta. Para isso, a maior parte dos processos envolvidos neste cenário foi desenvolvida como serviços web. Detalhes sobre as tecnologias utilizadas e como os requisitos especificados foram atendidos podem ser encontrados em <http://plscience.superdignus.com>, bem como mais informações sobre a aplicação desenvolvida.

3.4.3 Testes Funcionais

Nesta etapa foram realizados alguns testes na solução proposta nesta dissertação. Foram testados os seguintes recursos: banco de dados relacional, ontologia de contexto da solução e a aplicação Collaborative PL-Science.

No banco de dados relacional, foram incluídos alguns dados de maneira a popular as tabelas. Após isso, as consultas que foram pré-definidas para gerar as

informações de percepção, foram executadas no banco de dados com o propósito de verificar as informações retornadas.

Para a realização de testes na ontologia de contexto da aplicação, algumas instâncias foram criadas utilizando a ferramenta Protégé. A partir da utilização do motor de inferências, também provido por esta ferramenta, foi verificada a consistência da ontologia diante das regras que foram definidas para as suas classes.

Por fim, foram realizados os testes unitários das funcionalidades implementadas na aplicação Collaborative PL-Science, já integrando o banco de dados relacional e a ontologia de contexto com a aplicação.

À medida que falhas e melhorias foram sendo observadas, ajustes foram realizados de forma a preparar a proposta para a realização da avaliação da solução apresentada no capítulo seguinte.

3.5 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO

Este capítulo detalhou o desenvolvimento da abordagem Collaborative PL-Science. A visão geral da abordagem Collaborative PL-Science permitiu revisitar as lacunas que foram deixadas pela PL-Science (COSTA et al., 2013), e colaborar para o entendimento de como a solução proposta nesta dissertação pode atuar para minimizar estes problemas. A partir desta visão geral, requisitos funcionais e não-funcionais foram especificados de forma a definir o escopo da abordagem e ainda apresentar os atributos de qualidade necessários para este cenário.

A arquitetura da abordagem Collaborative PL-Science também foi apresentada neste capítulo. Foram evidenciados ainda, os pontos que evoluíram em relação a PL-Science, tais como: (i) a utilização do paradigma SOA, o que permitiu a inclusão de serviços na abordagem, possibilitando que a solução proposta nesta dissertação possa se expandir de tal forma que possa responder de forma efetiva e rápida aos desafios enfrentados no domínio científico; (ii) o suporte a colaboração oferecido aos cientistas que trabalham na LPSC; entre outros.

Os projetos físico e lógico do banco de dados relacional que compõe a memória de grupo da abordagem também foram apresentados. Outro mecanismo importante na solução proposta também foi detalhado, a ontologia Collaborative PL-Science. Por fim, aspectos de implementação e testes funcionais realizados na aplicação foram apresentados. Mais detalhes sobre a abordagem podem ser encontrados em: <http://plscience.superdignus.com>.

4 AVALIAÇÃO DA PROPOSTA

No decorrer deste capítulo são descritas as avaliações realizadas na abordagem Collaborative PL-Science. Para isso, a definição e o planejamento do método de pesquisa para a avaliação da hipótese formulada nesta dissertação são apresentados a seguir. Os estudos definidos foram executados no contexto de concepção de *workflows* científicos aplicados em experimentos realizados na EMBRAPA¹ e na FIOCRUZ², mais precisamente, para sequenciamento/alinhamento genético.

Uma vez que estes *workflows* já foram abordados na PL-Science (COSTA 2013) com o propósito de demonstrar a aplicabilidade da solução, o objetivo deste capítulo é apresentar e discutir as avaliações realizadas na Collaborative PL-Science. Este contexto foi escolhido devido ao grupo de pesquisa no qual estamos inseridos já possuir trabalhos desenvolvidos neste domínio. Entretanto, vale ressaltar que este estudo também pode ser aplicado em outros contextos, não se limitando apenas ao contexto de Bioinformática.

Nas próximas seções, são detalhados os estudos realizados.

4.1. DEFINIÇÃO E PLANEJAMENTO DO MÉTODO DE PESQUISA

No contexto deste trabalho o objetivo principal foi verificar a viabilidade da abordagem Collaborative PL-Science. Conforme mencionado nos capítulos anteriores, a abordagem Collaborative PL-Science utiliza alguns elementos de colaboração (informações de percepção, contexto e um mecanismo de suporte à comunicação) integrados ao núcleo de artefatos de uma Linha de Produtos de Software Científico. A avaliação da solução apresentada nesta dissertação tem o propósito de verificar os

¹ Empresa Brasileira de Pesquisa Agropecuária

² Fundação Oswaldo Cruz

benefícios oferecidos pela abordagem no apoio aos cientistas durante a concepção de *workflows* científicos e na geração de oportunidades de interação entre eles.

Para isso, baseada no método GQM (BASILI 1994), é apresentada a estrutura a seguir, que define o estudo realizado nesta dissertação.

“Analisar a utilização de informações de percepção e contexto no domínio de Bioinformática

para o propósito de caracterizar o apoio oferecido aos cientistas e as oportunidades de interações geradas entre eles

com relação à eficácia do serviço oferecido pela abordagem Collaborative PL-Science

do ponto de vista do observador (aplicador do estudo)

no contexto de concepção de workflows científicos para alinhamento/sequenciamento genético, construídos através de uma Linha de Produtos de Software Científico.”

Nesta dissertação, a definição da palavra “*caracterizar*” pode ser entendida como o ato de descrever as características observadas em relação à utilização da abordagem proposta. Outro termo que é necessário definir para o contexto deste trabalho é “*eficácia*”, que pode ser entendido como os benefícios propiciados aos cientistas através da utilização da abordagem Collaborative PL-Science.

A hipótese formulada para este estudo tem o intuito de oferecer respostas ao problema apresentado nessa dissertação, sendo ela: *se oferecermos aos cientistas um serviço que apresente alguns elementos de colaboração (informações de percepção, contexto e um mecanismo de suporte à comunicação) integrados ao núcleo de artefatos de uma Linha de Produtos de Software Científico (LPSC), podemos criar oportunidades para promover e apoiar a interação entre eles, contextualizando-os sobre o ambiente compartilhado de trabalho no qual estão inseridos.*

Entre os métodos de pesquisa adotados para verificar a hipótese deste trabalho está a realização de provas de conceitos (Proof of Concept- PoC). Para isso, foram

utilizados dados obtidos através de experimentações prévias executadas na EMBRAPA e na FIOCRUZ (COSTA 2013). Outro método de pesquisa utilizado foi o estudo de caso. Após a realização das provas de conceitos, um estudo de caso foi realizado no laboratório de pesquisa (NEnC – Núcleo de Pesquisa em Engenharia do Conhecimento - UFJF) com o propósito de responder as questões de pesquisas que foram levantadas.

Apesar de não apresentar o formalismo de um estudo experimental, a realização de provas de conceitos neste domínio, contribuiu para a avaliação preliminar da hipótese formulada. Além disso, possibilitou também a verificação da viabilidade da solução proposta nesta dissertação.

Estas avaliações foram aplicadas em um problema real de concepção de *workflows* científicos. Quanto à generalidade, a avaliação foi realizada em um domínio específico, o da bioinformática genômica, na concepção de *workflows* científicos para alinhamento/sequenciamento genético. Quanto à dimensão observada, buscou-se a eficácia do apoio colaborativo oferecido aos usuários ao utilizarem a abordagem Collaborative PL-Science.

4.2 PROVAS DE CONCEITOS

4.2.1 Cenário e Caracterização do Grupo

O cenário das provas de conceitos apresentadas nesta dissertação está inserido no contexto de Bioinformática Genômica, e veio ao encontro de alguns projetos realizados na EMBRAPA e na FIOCRUZ. A EMBRAPA, por exemplo, possui algumas pesquisas voltadas para a área de Bioinformática Genômica, com o foco para bovinos de leite e que estão relacionadas a características genéticas relacionadas à resistência a parasitas, em específico, ao carrapato bovino.

Este cenário foi obtido a partir de dados de experimentações prévias realizadas por COSTA (2013), permitindo assim, o desenvolvimento das quatro provas de conceitos apresentadas nesta dissertação.

Os cientistas participantes destas provas de conceitos foram identificados pelos pseudônimos: Cientista A e Cientista B, ambos pertencentes ao grupo de pesquisa da EMBRAPA, e os cientistas Cientista C e Cientista D do grupo de pesquisa da FIOCRUZ.

Para a realização deste estudo, instâncias da aplicação Collaborative PL-Science³ foram disponibilizadas para estes pesquisadores, os quais puderam trabalhar na composição de *workflows* científicos e explorar os recursos propostos pela aplicação.

A duração prevista para a concepção destes produtos foi de aproximadamente 10 minutos para cada *workflow*, visto que o propósito destas provas de conceitos compreendeu apenas a concepção dos novos produtos e não a sua execução em um Sistema Gerenciador de *Workflows* Científicos.

Para apoiar a concepção destes produtos, os artefatos necessários para realizar a atividade também foram disponibilizados na aplicação, podendo ser acessados através dos endereços apresentados na Tabela 5:

Tabela 5. Artefatos utilizados na Avaliação da Solução

Nome do Artefato:	Endereço:
SequenceAligningFeatureModelv1b	http://plscience.superdignus.com/SequenceAligningFeatureModelv1b.xml
SequenceAligningMappingv1b	http://plscience.superdignus.com/SequenceAligningMappingv1b.xml
SequenceAligningOntologyv1b	http://plscience.superdignus.com/SequenceAligningOntologyv1b.owl

4.2.2 Primeira PoC: *Workflow* para Sequenciamento/Alinhamento Genético

O *workflow* científico desenvolvido através da primeira prova de conceitos tem o propósito de obter os dados de uma sequência genética, através de um sequenciador,

³ Disponível em: <http://plscience.superdignus.com/>

até a comparação desta sequência obtida com as bases de sequências já existentes, através da realização da tarefa de alinhamento (Figura 20).

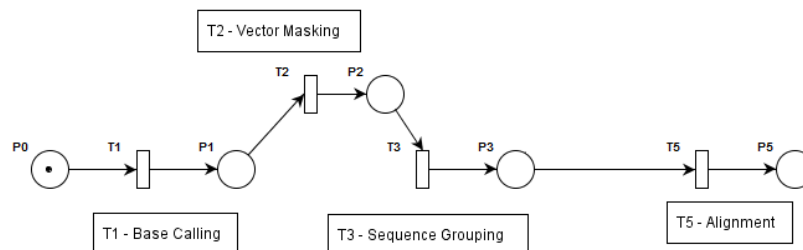


Figura 4.20: *Workflow* para Sequenciamento e Alinhamento Genético (COSTA, 2013).

A realização desta prova de conceitos tem o objetivo de avaliar a viabilidade de uso da Collaborative PL-Science sob o aspecto das funcionalidades oferecidas pela aplicação. Para isso, descreveu os processos realizados pelo Cientista A durante a concepção de seu produto através da aplicação Collaborative PL-Science.

Ao iniciar a concepção do *workflow*, o Cientista A selecionou os artefatos necessários a partir do núcleo da LPSC. Ao selecioná-los foi informado que alguns produtos já tinham sido desenvolvidos a partir dos mesmos artefatos. Isso foi possível graças às informações de percepção apresentadas através do mecanismo de notificações oferecido pela aplicação.

Esse mecanismo apresenta informações de percepção geradas a partir do histórico de atividades de outros cientistas, capturado pela aplicação e persistido na memória de grupo para compor o conhecimento sobre o domínio.

A partir das informações apresentadas, o Cientista A pode conhecer produtos que já haviam sido desenvolvidos através da Collaborative PL-Science, e que compartilhavam os mesmos artefatos que ele havia selecionado. Entretanto, ele pode tomar conhecimento de que nenhum *workflow* similar ao que ele estava desenvolvendo, havia sido criado. Caso existisse algum produto similar, essa informação poderia poupar-lhe tempo durante a concepção de um novo *workflow*,

porém, não foi o que ocorreu. Sendo assim, o Cientista A optou por desenvolver um novo produto.

Como passo seguinte, foi solicitada pela aplicação a escolha das *features* necessárias para compor o *workflow*. A Figura 21 apresenta a interface de desenvolvimento de novos produtos, destacando a seleção de *features* (A). Através dela é possível observar algumas *features* selecionadas para compor o novo produto (B), entre elas estão: Application_type, que neste caso foi escolhida a *feature* workflow, a Sequence_Platform, Sanger, entre outras.

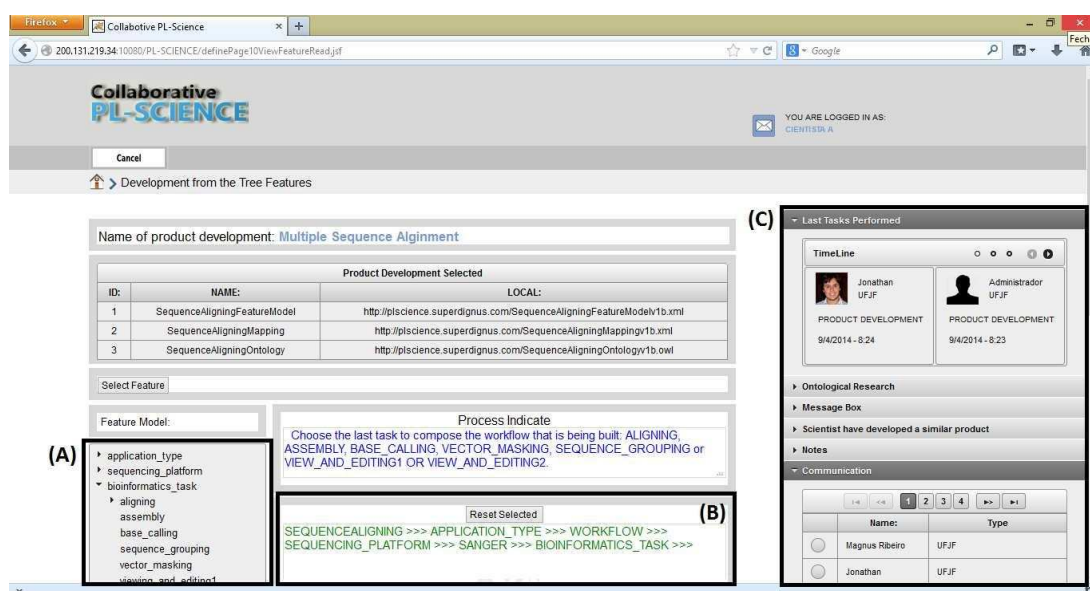


Figura 4.21: Interface de desenvolvimento de novos *workflows* científicos.

Na Figura 21, é possível observar um quadro de informação apresentado para o Cientista A no canto direito da tela (C). Neste quadro são apresentadas informações como: decisões que foram tomadas e/ou anotações realizadas por outros cientistas que já desenvolveram produtos similares ao produto que estava sendo desenvolvido pelo Cientista A, além de outras informações de percepção. A partir das informações apresentadas neste quadro, o cientista A pode conhecer as atividades que estavam sendo realizadas por outros participantes. Entretanto, nenhuma informação sobre

produtos similares ao que ele estava desenvolvendo existiam ainda na memória de grupo. Apesar disso, através do mecanismo de suporte à comunicação oferecido pela aplicação, e ainda, através do conhecimento das atividades que outros cientistas estavam desempenhando na Collaborative PL-Science, o Cientista A pode aproveitar a oportunidade para interagir com outros cientistas, para esclarecer possíveis dúvidas.

Nesta prova de conceitos, por exemplo, o Cientista A estava em dúvidas quanto à escolha do *web service* apropriado para realizar a tarefa de agrupamento de sequências (T5) que diz respeito a última tarefa do *workflow* que ele estava desenvolvendo. Para isso, aproveitou a oportunidade e o mecanismo de suporte à comunicação oferecido pela aplicação para interagir com o Cientista B, para que pudessem decidir juntos se o Cientista A escolheria *web services* referentes à classe PHRAP ou à classe CAP3 para realizar esta tarefa, conforme foram apresentados para ele através de inferências na ontologia de domínio SequenceAlginOntologyv1b.owl.

Através desta interação, os cientistas A e B puderam esclarecer eventuais dúvidas. Além disso, puderam ainda, trocar experiências sobre os algoritmos e/ou *web services* que já tinham sido utilizados pelo Cientista B durante a concepção de outros *workflows* científicos que também contavam com a tarefa de agrupamento de sequências (T5). Nesta prova de conceitos, o Cientista B pode informar ao Cientista A, por exemplo, sobre a existência de outro *web service* (INB-dev:genome.imim.es:runPhrap)⁴ que também poderia ser utilizado para a realização da atividade de agrupamento de sequências, e que até o momento não estava presente entre os *web services* sugeridos pela ontologia de domínio.

Conhecendo a existência deste outro artefato, o Cientista A pode sugerir para outros cientistas a inclusão deste “novo *web service*” no núcleo da Collaborative PL-Science. Além disso, pode também, realizar uma anotação sobre a atividade desenvolvida através da aplicação, permitindo que outros cientistas pudessem em um futuro próximo se apoiar nas decisões e comentários disponibilizados por ele durante a concepção do *workflow* científico.

⁴ <http://www.biocatalogue.org/services/2268>

Como último passo do processo, a aplicação Collaborative PL-Science gerou um arquivo especificado em XML com o detalhamento das tarefas realizadas por este produto. No final deste processo de desenvolvimento, o cientista A realizou o *download* deste arquivo na aplicação.

4.2.2.1 Evidências Observadas na Primeira PoC

No cenário apresentado na primeira prova de conceitos ainda não existiam informações na memória de grupo sobre produtos similares ao *workflow* desenvolvido pelo Cientista A. Entretanto, outros serviços e tipos de informações de percepção foram oferecidos pela aplicação para o cientista, e puderam apoiá-lo, e orientá-lo, no momento de concepção de um novo produto. Como exemplo, podemos citar as informações apresentadas através da “*linha do tempo*” do quadro de Informações. Através dela o Cientista A pode conhecer os pesquisadores que estavam conectados na aplicação e as tarefas que eles estavam realizando naquele momento. Com isso, ele pode decidir se queria ou não interagir com algum dos pesquisadores através do mecanismo de suporte à comunicação, por exemplo.

Pode-se observar ainda que as informações de percepção apresentadas através do quadro de informações da aplicação geraram oportunidades para que o Cientista A pudesse se manter atualizado sobre o que estava acontecendo no seu ambiente de trabalho. Além disso, possibilitou que através do serviço de suporte à comunicação, os pesquisadores pudessem ter um canal direto com outros membros do grupo, minimizando assim a sensação de isolamento e de ausência de recursos no espaço de trabalho compartilhado.

Entretanto, para que pudéssemos obter novas evidências sobre a importância da memória de grupo para este cenário, uma vez que a memória de grupo não existia nesta primeira prova de conceitos, sob o aspecto de apresentação de informações de percepção, geradas a partir deste mesmo núcleo comum de conhecimento, uma segunda prova de conceitos foi realizada, e é apresentada a seguir.

4.2.3 Segunda PoC: *Workflow Pipeline PhredPhrap*

Nesta segunda prova de conceitos, um novo produto foi desenvolvido. Tratou-se de um *workflow* denominado *Pipeline PhredPhrap*, que compreende os seguintes passos: (i) ler os dados gerados por um sequenciador genético e identificar a sequência de DNA gerada, atribuindo valores de qualidade para cada posição nucleotídica identificada; (ii) realizar uma busca de regiões na sequência produzida que devem ser retiradas ou mascaradas; (iii) agrupar os resultados advindos do processo de leitura dos dados formando em sequências maiores (ARBEX, 2009). O modelo do segundo *workflow* é apresentado pela Figura 22.

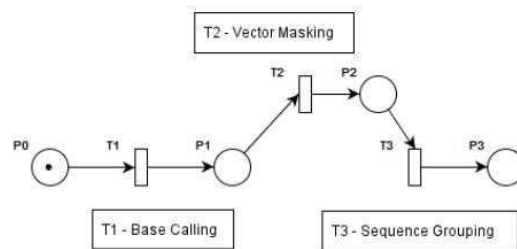


Figura 4.22: Modelo do *workflow* Pipeline PhredPhrap (ARBEX, 2009).

O objetivo desta segunda prova de conceitos foi avaliar a viabilidade da presença da memória de grupo, como recurso utilizado para manter registros sobre as atividades desenvolvidas pelos cientistas, e ainda, oferecer informações de percepção para contextualizá-los sobre o ambiente de trabalho compartilhado.

Para iniciar a atividade, o cientista B, membro do grupo de pesquisas da EMBRAPA, se conectou na Collaborative PL-Science e iniciou a concepção de um novo *workflow* científico, denominado *Pipeline PhredPhrap*. Após escolher a opção de desenvolvimento de produtos e informar os artefatos que seriam utilizados, ele pode perceber que outro cientista já havia desenvolvido um produto similar ao que estava sendo criado. Com isso, pode valer-se destas informações para realizar a atividade, explorando a oportunidade de reuso de conhecimento oferecida pela abordagem.

Assim, após conhecer como produtos similares foram construídos, o Cientista B iniciou a concepção do *workflow*. O primeiro passo compreendeu a escolha das *features* necessárias para compor o produto *Pipeline PhredPhrap*.

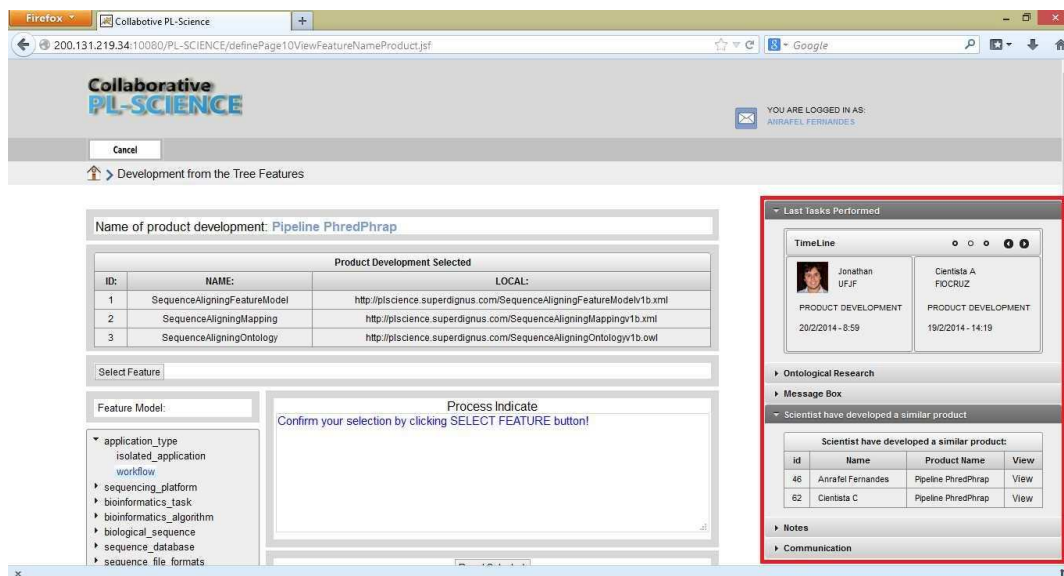


Figura 4.23: Desenvolvimento do Produto *Pipeline PhredPhrap*.

Amparado pelas informações de percepção oferecidas para o usuário através de um quadro apresentado na tela de concepção de *workflows* científicos (Figura 23), o Cientista B pode escolher os *pontos de variação*⁵ no modelo de *features*. Isto aconteceu não apenas baseado em sua experiência, mas no conhecimento que obteve a partir dos históricos das atividades de outros cientistas que desenvolveram *workflows* semelhantes. Com isso, as *features* escolhidas pelo Cientista B para compor o produto foram: (i) tipo de aplicação construída: *feature workflow*; (ii) plataforma de

⁵ Pontos de Variação: são pontos nos quais as características dos artefatos da linha de produto de software podem sofrer variações (alterações).

sequenciamento: *feature* Sanger; e (iii) tarefas para compor o produto: *features* *sequence_grouping*, *vector_masking* e *base_calling*.

Considerando o funcionamento da Collaborative PL-Science, o passo seguinte foi a escolha dos algoritmos e/ou *web services* necessários para compor o produto. Nesta prova de conceitos, por exemplo, o Cientista B precisava de mais desempenho no *workflow* que estava desenvolvendo. Com isso, a partir das informações de percepção apresentadas pela aplicação, ele pode optar por utilizar outro *web service* em relação àquele que foi utilizado por cientistas que já desenvolveram produtos similares.

Ao visualizar os algoritmos e *web services* utilizados por outros pesquisadores, o Cientista B, como conhecedor do domínio, pode entender por experiência própria que existiam outros *web services* que também poderiam ser utilizados para a realização da tarefa *sequence_grouping*. Assim, ao invés de utilizar o recurso *runPhrap*, disponível em: <http://www.biocatalogue.org/services/2268>, para realizar esta tarefa, ele optou pela utilização de outro artefato denominado *runPhrapService* (<http://www.biocatalogue.org/services/1567>) que também pode realizar esta tarefa e ainda, com a possibilidade de apresentar um melhor resultado neste produto.

Isso permitiu que o Cientista B conhecesse alguns dos requisitos não funcionais dos *workflows* científicos que já haviam sido construídos. Mais ainda, resultados previamente obtidos por outros pesquisadores puderam ser reutilizados por ele reduzindo o custo do experimento. Além disso, o mecanismo de suporte à comunicação oferecido pela aplicação pode ser utilizado pelo Cientista B, que aproveitou a oportunidade para interagir com outros pesquisadores do grupo para discutir a inclusão do *web service* utilizado por ele na ontologia de domínio.

Após finalizar a concepção de produto, o histórico das atividades realizadas pelo Cientista B, poderá também, num futuro próximo, apoiar os passos de outros cientistas na aplicação Collaborative PL-Science.

4.2.3.1 Evidências observadas na Segunda PoC

Além das evidências observadas através da primeira prova de conceitos, destacou-se na segunda PoC a importância de manter um núcleo de conhecimento comum sobre o domínio, sempre atualizado e disponível para que os cientistas possam explorá-lo a qualquer momento. Isso possibilita que os pesquisadores possam auxiliar o trabalho, uns dos outros, mesmo que indiretamente. Pois ao permitir que seja capturado o histórico das atividades que um cientista desenvolveu no ambiente de trabalho compartilhado, ele acaba colaborando, mesmo que indiretamente com o trabalho de outros membros do grupo de pesquisa ou ainda de instituições parceiras.

A partir do conhecimento sobre como as oportunidades de colaboração e interação foram geradas pela aplicação neste cenário, uma terceira prova de conceitos foi realizada para obter novas evidências sobre solução proposta em relação à utilização do mecanismo de suporte à comunicação. Esta prova de conceitos é apresentada a seguir.

4.2.4 Terceira PoC: *Workflow* para Alinhamento Múltiplo de Sequências

Na terceira prova de conceitos, o objetivo foi avaliar a viabilidade do serviço de suporte à comunicação oferecido pela Collaborative PL-Science, sob o aspecto de sincronismo ou assincronismo das interações realizadas através do mecanismo. Para isso, um cientista do grupo de pesquisa da FIOCRUZ, identificado como Cientista C foi o usuário responsável pelo desenvolvimento do *workflow* científico. Tratou-se de um produto para realizar o alinhamento múltiplo de sequências genéticas. A atividade associada a esta aplicação constitui em alinhar três ou mais sequências biológicas (sendo estas sequências de proteína ou sequências nucleotídicas) de comprimentos similares. A utilização de uma aplicação que realize esse tipo de alinhamento possibilita a identificação de regiões de similaridades entre as sequências que podem indicar relações funcionais, estruturais e/ou evolutivas entre as sequências estudadas (EMBL-EBI, 2012). A Figura 24 ilustra a única tarefa envolvida neste produto.

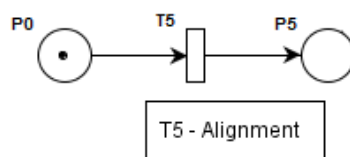


Figura 4.24: Aplicação para Alinhamento Múltiplo de Sequências (COSTA, 2013).

A tarefa apresentada na Figura 24 é similar à última tarefa do *workflow* construído pelo Cientista A, pesquisador da EMBRAPA, apresentada na primeira prova de conceitos. Assim, ao iniciar a concepção do novo produto, após informar os artefatos que seriam utilizados, a aplicação pode indicar para este cientista os produtos que já tinham sido construídos a partir da utilização dos mesmos artefatos informados pelo usuário neste momento. Com isso, o Cientista C pode perceber que outro usuário da aplicação, denominado Cientista A, pertencente ao grupo de pesquisa da EMBRAPA, já havia construído um produto para realizar sequenciamento/alinhamento genético.

Neste momento o Cientista C optou por conhecer mais sobre o produto construído, ou seja, clicou sobre o *link* apresentado no quadro de informações e pode conhecer os passos que o Cientista A realizou para compor o produto. Através deste conhecimento, ele pode visualizar que a última tarefa do *workflow* desenvolvido era semelhante à tarefa do produto que ele precisava construir.

Como passo seguinte, o Cientista C iniciou a concepção do *workflow*. Neste passo, ele escolheu as *features* necessárias para compor o experimento, sendo elas: *isolated_application*, para o tipo de aplicação, a feature NGS para a plataforma de sequência e por último, pode escolher as *features* que correspondiam à tarefa que seria realizada pelo produto, foram elas: *Aligning*, *Global_aligning* e *Multiple_global_aligning*.

Após a escolha das *features*, a aplicação apresentou para o Cientista C os algoritmos e *web services* específicos que foram encontrados a partir das restrições especificadas na ontologia de domínio, e que poderiam ser escolhidos por ele para realizar a tarefa especificada no produto. Sendo assim, algoritmos e *web services*

referentes às classes BLAST e SmithWaterman da ontologia de Alinhamento de Sequências, foram fornecidos ao Cientista C como possibilidade de escolha para a concepção do *workflow* científico.

Neste momento, para que pudesse tomar decisão, o Cientista C optou por visualizar quais foram os algoritmos e/ou *web services* que foram utilizados pelo Cientista A em seu produto, já que a última tarefa do produto do Cientista A era semelhante à tarefa do produto que ele estava construindo.

Assim, através das informações apresentadas no quadro da aplicação, oportunidades de interação entre o Cientista C e outros cientistas foram geradas. Neste caso, conhecendo qual foi o algoritmo escolhido pelo Cientista A, o Cientista C optou por interagir com este usuário para entender, por exemplo, o porquê da escolha do algoritmo psi-blast para realizar busca de banco de dados de proteínas a fim de encontrar regiões de semelhança entre as sequências biológicas, já que ele não tinha conhecimento sobre este algoritmo. Para isso, o Cientista C, utilizou o mecanismo de suporte à comunicação para encaminhar a dúvida para o Cientista A.

Entretanto, o Cientista A não estava conectado na aplicação, o que dificultou a realização de uma interação síncrona naquele momento. Apesar disso, a mensagem que foi encaminhada para o Cientista A não foi perdida. No momento em que este usuário se conectou na Collaborative PL-Science, ele pode responder e manter registrado o conhecimento na memória de grupo e, com isso, apoiar, em um futuro próximo, outros cientistas durante a concepção de novos *workflows*.

O Cientista C, por sua vez, não pode aguardar a resposta do Cientista A para esclarecer a sua dúvida, pois até o momento da finalização da atividade o Cientista A ainda não havia respondido a mensagem. Apesar disso, o Cientista C pode valer-se das informações que estavam armazenadas na memória de grupo, para conhecer mais detalhes sobre o *workflow* que o Cientista A havia desenvolvido.

4.2.4.1 Evidências Observadas na Terceira PoC

Através desta prova de conceitos buscou-se explorar as oportunidades de interação que poderiam ser geradas através da Collaborative PL-Science. Assim como nas duas primeiras provas de conceitos, nesta terceira também foi possível observar algumas contribuições, tais como: i) a oportunidade de promover a interação entre os cientistas a partir de informações de percepção e contexto que foram oferecidas através dos históricos de atividades capturados quando os cientistas realizavam as suas atividades; e ii) o suporte para que os cientistas pudessem tomar decisões e desenvolver experimentos a partir de artefatos já analisados, comentados e persistidos por outros pesquisadores.

Embora a Collaborative PL-Science ofereça aos usuários um mecanismo de suporte à comunicação, explorar a presença simultânea dos cientistas na aplicação pode ser importante para que novas interações possam ser geradas e, ao mesmo tempo, para permitir que este conhecimento possa ser registrado na memória de grupo. Interações espontâneas entre os cientistas podem ser motivadas explorando a presença destes pesquisadores no espaço de trabalho compartilhado, de maneira que essas interações possam apoiar o desenvolvimento de suas atividades. Como resultado, a sensação de trabalho individual e isolamento no ambiente de trabalho pode ser reduzida.

Apesar disso, foi possível obter evidências sobre a possibilidade de interação e colaboração entre cientistas de grupos de pesquisas geograficamente distribuídos (EMBRAPA e FIOCRUZ) durante composição de um novo *workflow* científico. Caso o Cientista A, pertencente ao grupo de pesquisa da EMBRAPA, estivesse conectado na aplicação naquele momento, ele, como conhecedor do domínio, poderia sugerir outros algoritmos e/ou *web services* para o Cientista C, ou ainda, trocaram experiências quanto à utilização de um algoritmo em detrimento de outro. Mesmo assim, no momento em que o Cientista A pode responder a dúvida do Cientista C, o conhecimento foi armazenado na memória de grupo da aplicação apoiando outros cientistas em um futuro próximo. Além disso, a partir desta resposta, Cientista C pode,

por exemplo, realizar a alteração do artefato *SequenceAlignOntologyv1b.owl*, gerando uma versão atualizada, que pode incluir outros algoritmos sugeridos pelo Cientista A, o que também apoiará o trabalho de outros cientistas na aplicação.

Além disso, através de consultas realizadas na memória de grupo, os cientistas podem buscar outras informações sobre o domínio que estão atuando, e conseqüentemente explorar melhor o reuso de artefatos. Sendo assim, uma quarta prova de conceitos foi realizada com objetivo de avaliar o mecanismo de consultas ontológicas à memória de grupo sob o aspecto de representação do contexto no qual a aplicação Collaborative PL-Science está inserida. Através desta quarta prova de conceitos é apresentado como este recurso auxilia as atividades desempenhadas pelos pesquisadores, apoiando-os na descoberta de outras informações sobre o ambiente de trabalho compartilhado.

4.2.5 Quarta PoC: Realização de uma Consulta à Ontologia da Aplicação

Na quarta prova de conceitos, uma consulta à ontologia que descreve o contexto da abordagem Collaborative PL-Science foi realizada. O propósito desta consulta foi buscar informações sobre o produto *Pipeline PhredPhrap*. Para isso, um cientista identificado como Cientista D, do grupo de pesquisa da FIOCRUZ, se conectou à aplicação e fez uso do recurso denominado "*Ontological Research*" disponibilizado pela aplicação.

Como passo seguinte, o Cientista D escreveu a consulta utilizando a linguagem SPARQL e a submeteu à aplicação.

A partir das inferências realizadas nesta ontologia e das *Properties Chains* descritas para ela, o Cientista D teve como retorno as seguintes informações:

- a) o tipo do produto pesquisado, neste caso um *workflow*;
- b) os artefatos utilizados para o desenvolvimento do produto: *SequenceAligningFeatureModel*, *SequenceAligningMapping* e o *SequenceAligningOntology*;
- c) o histórico de atividades envolvidas na composição deste produto;

d) o cientista responsável pelo desenvolvimento do produto selecionado. Neste exemplo, o cientista é o SC_Anrafel;

e) a instituição a qual o cientista, que desenvolveu o produto, está associado. Neste caso, a instituição é a CO_UFJF; e

f) a atividade que gerou o produto desenvolvido (AC_Develop_a_Product_01).

O resultado desta consulta pode ser observado através da Figura 25. O *Framework Protegé* foi escolhido para apresentar os resultados com o propósito de deixar mais compreensíveis as inferências realizadas pela ontologia da Collaborative PL-Science.

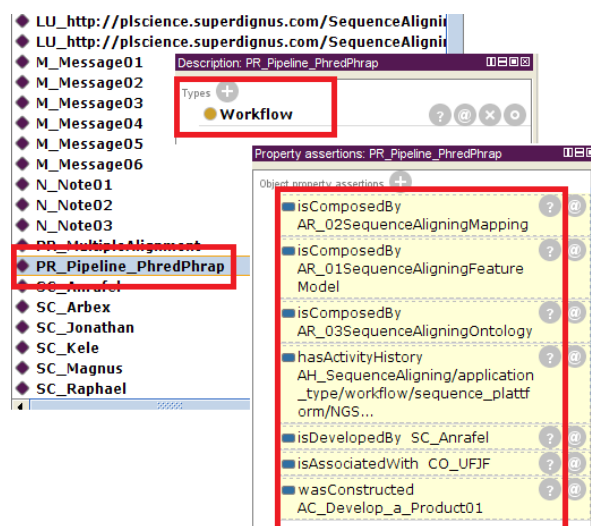


Figura 4.25: Informações sobre o produto *Pipeline PhredPhrap*.

Após tomar conhecimento sobre estas informações, o Cientista D pode conhecer mais sobre o produto *Pipeline PhredPhrap*, e assim realizar outras consultas na ontologia.

Assim, o Cientista D buscou outras informações sobre a opção que corresponde à atividade que deu origem ao produto pesquisado: AC_Develop_a_Product_01. Assim,

escreveu uma nova consulta em SPARQL e a submeteu à aplicação novamente. Sobre esta consulta foi possível conhecer as seguintes informações (Figura 26):

- Os cientistas que interagiram durante a sua realização. Neste exemplo temos: SC_Magnus, SC_Jonathan e SC_Anrafel;
- As mensagens que foram trocadas entre os cientistas no momento da concepção do produto pesquisado, neste caso o *workflow Pipeline PhredPhrap* (M_Message02, M_Message01 e M_Message06);
- Os artefatos que foram utilizados e o histórico de atividades, ambos já foram apresentados ao buscar informações sobre o produto;
- A lista de interação que foi iniciada (INTERACTION_01); e
- O cientista responsável pela sua execução, e que neste caso levou à concepção do produto pesquisado (SC_Anrafel).

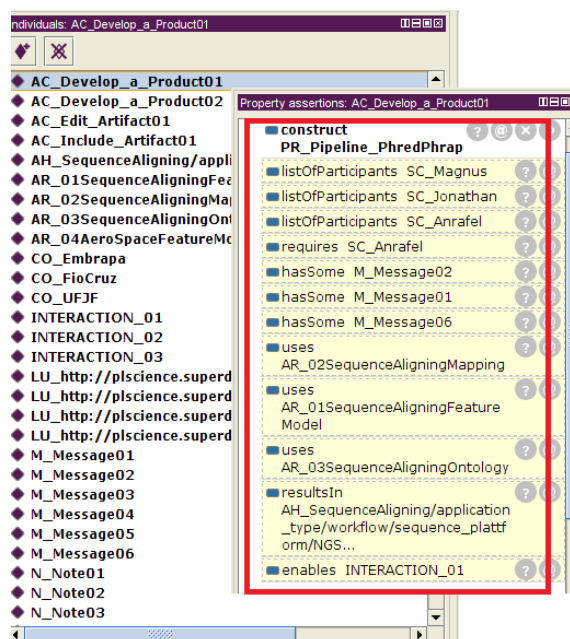


Figura 4.26: Informações sobre a atividade que gerou o *Pipeline PhredPhrap*.

O propósito de apresentar esta prova de conceitos foi o de evidenciar as vantagens das *Properties Chains* criadas na ontologia de Collaborative PL-Science. Com

isso, evidenciou-se também a viabilidade das consultas realizadas na ontologia da solução proposta. Com pouco esforço foi possível conhecer como os recursos se relacionavam ao contexto da solução proposta.

4.2.5.1 Evidências observadas na Quarta PoC

Nesta prova de conceitos foi apresentado o recurso de consulta à memória de grupo, mais especificamente, consultas à ontologia da aplicação. Com isso, possibilitou-se a descoberta de outros conhecimentos sobre o domínio a partir do histórico de atividades que foi coletado pela Collaborative PL-Science, durante a realização das atividades dos cientistas. Sendo assim, a principal evidência observada através desta prova de conceitos foi a oportunidade oferecida aos cientistas para explorar melhor o reuso de artefatos e de conhecimento nos quais outros usuários já inseriram comentários ou deixaram registradas suas experiências em sua utilização. Além disso, existe ainda a possibilidade de conhecer não apenas o processo de desenvolvimento de um produto da Linha de Produtos de Software Científico, mas as decisões tomadas, cientistas envolvidos, artefatos utilizados, interações feitas, e os grupos de pesquisas que trabalham ou trabalharam juntos, entre outras.

A realização das quatro provas de conceitos apresentadas nesta dissertação, nos permitiu avaliar a viabilidade dos principais recursos oferecidos pela Collaborative PL-Science, e através delas observar algumas contribuições da aplicação para o domínio científico. A aproximação entre os cientistas, bem como a forma de atuar de cada um deles, é um exemplo de contribuição da aplicação, por conduzir a uma visão mais ativa e participativa deste participante no seu ambiente de trabalho. Entretanto, observou-se a necessidade de realizar uma investigação adicional através de um estudo experimental (conduzido em um problema real e com participantes conhecedores do domínio de Bioinformática, no qual este trabalho está inserido), para que fosse possível obter evidências sobre informações relacionadas ao uso da aplicação, principalmente em aspectos relacionados à interface. Sendo assim, um estudo de caso foi planejado e conduzido neste contexto, sendo apresentado a seguir.

4.3 ESTUDO DE CASO

4.3.1 Planejamento do Estudo de Caso

A partir da análise das provas de conceitos, foram formuladas questões para a realização do estudo de caso. As questões principais levantadas neste estudo se relacionam à investigação das oportunidades de interação geradas entre os participantes durante a realização da atividade de concepção de *workflow* científico através da Collaborative PL-Science.

Para isso, duas questões estiveram sob o foco deste estudo:

Questão 1: Como a utilização de informações de Percepção e Contexto pela aplicação Collaborative PL-Science apoia os usuários na concepção de *workflows* científicos?

Questão 2: Como a utilização da aplicação Collaborative PL-Science promove a colaboração entre os usuários?

YIN (2009) recomenda a utilização do método de estudos de casos quando as questões da pesquisa são do tipo “Como?” e “Por quê?”. Este método é recomendado ainda, quando a pesquisa necessita de uma investigação empírica dos fenômenos dentro de um contexto real. Este método de pesquisa foi selecionado para esta dissertação devidos os seguintes fatores: (i) as questões deste estudo são do tipo “Como?”; e (ii) a concepção dos *workflows* científicos corresponde a um aspecto da vida real.

Logo, a metodologia utilizada na avaliação foi a realização de um estudo de caso, onde foram observadas questões relacionadas à utilização da aplicação com os seus respectivos elementos de colaboração em direção aos objetivos dos participantes.

A dimensão observada neste estudo foi realizada através da seguinte variável dependente: *a eficácia do apoio colaborativo oferecido aos cientistas ao utilizarem a aplicação Collaborative PL-Science no contexto de concepção de workflows científicos.*

4.3.2 Objetivo do Estudo

O objetivo do estudo realizado foi avaliar se as funcionalidades oferecidas pela Collaborative PL-Science poderiam estimular a colaboração e/ou interação entre os usuários. Além disso, os estudos realizados visaram à discussão e à avaliação sobre como as informações de Percepção e Contexto, recursos também oferecidos pela aplicação, poderiam apoiar estes mesmos usuários na concepção de seus *workflows* científicos.

4.3.3 Caracterização do Grupo

Foram convidados dois participantes para atuarem neste estudo, ambos integrantes do NEnC - UFJF e conhecedores do domínio de Bioinformática, contexto no qual o estudo foi realizado. Iniciada a realização da atividade proposta, a duração foi de aproximadamente 20 minutos. Os participantes já conheciam a metodologia de concepção de *workflows* através da Collaborative PL-Science. Sendo assim, não houve a necessidade de um treinamento prévio.

4.3.4 Cenário do Estudo

Duas instâncias da aplicação Collaborative PL-Science foram disponibilizadas para os participantes. Através delas, eles iniciaram o desenvolvimento da tarefa proposta. Ambos se encontravam no mesmo espaço físico de trabalho, ou seja, no NEnC-UFJF.

A tarefa designada para cada *subject* dizia respeito à composição de dois *workflows* científicos, sendo eles: (i) o produto *Pipeline PhredPhrap*; e (ii) *workflow* para Sequenciamento e Alinhamento genético, ambos já apresentados através das provas de conceitos.

As tarefas envolvidas nestes produtos já eram conhecidas pelos participantes, devido ao grupo de pesquisa já ter realizado projetos neste mesmo domínio. Os

artefatos utilizados foram os mesmos disponibilizados para a realização das provas de conceitos e disponíveis na Tabela 5 desta dissertação, estando disponíveis no núcleo de artefatos da LPSC.

Cada participante iniciou a concepção dos produtos e seguiu as etapas conforme foram propostas para eles. Para a realização deste estudo de caso, as informações presentes da memória de grupo foram mantidas, logo existiam informações de produtos similares que já haviam sido construídos na aplicação. Através de observações diretas realizadas durante todo o processo em que participantes desenvolviam as atividades propostas, foi possível perceber que ambos observaram e ficaram atentos às informações de percepção que eram apresentadas pela aplicação. Em alguns momentos buscaram informações sobre os produtos que já haviam sido construídos para desenvolver o *workflow* solicitado. Ao término, quatro *workflows* científicos foram desenvolvidos através da aplicação, sendo dois de cada participante. O tempo de concepção de cada *workflow* foi de aproximadamente 10 minutos.

Ao final do desenvolvimento das atividades, algumas sugestões e melhorias a serem implementadas foram sugeridas como trabalhos futuros na aplicação.

4.3.5 Fontes de Evidências

Neste estudo de caso, a variável definida anteriormente foi observada através de duas fontes de evidências. A primeira delas foi através de observações diretas. Esta fonte de coleta de dados baseou-se em atividades informais de observação realizadas no NEnC – UFJF, com o objetivo de obter evidências diferentes daquelas levantadas através das provas de conceitos.

A realização deste estudo no laboratório de pesquisa da UFJF foi importante para fornecer informações relacionadas ao uso da aplicação, principalmente em aspectos relacionados à interface, utilização do mecanismo de suporte à comunicação e do quadro de informações, responsável por oferecer as informações de percepção e contexto para os usuários. Através desta fonte de evidências foi possível observar a dificuldade e limitações de uso da Collaborative PL-Science.

A segunda fonte de evidência foi observada através de uma entrevista informal, realizada logo após o término da atividade proposta. Através desta entrevista, os dois participantes puderam expressar sugestões e melhorias a serem implementadas na aplicação.

Com isso, foi possível a realização de algumas modificações nas funcionalidades aplicação. Um exemplo foi a modificação da maneira de apresentação das informações de percepção, permitindo que os usuários pudessem visualizar mais detalhes sobre as atividades desenvolvidas por outros usuários, paralelamente à realização da própria atividade. Sugestões de *layout* e organização dos componentes da interface da aplicação também foram sugeridas.

Estas duas fontes de evidências foram acompanhadas sob o ponto de vista externo, realizada pelo autor desta dissertação. Nenhuma opinião sobre a utilização da aplicação ou sugestão foi emitida pelo autor.

4.3.6 Resultados do Estudo

Os *workflows* científicos desenvolvidos através deste estudo de caso foram construídos apoiados, principalmente, no histórico das atividades das provas de conceitos que estavam armazenados na memória de grupo. Esta constatação foi feita a partir da análise das mensagens trocadas entre os participantes do estudo, e também através de observação direta. Percebeu-se assim, que não havia uma relação direta entre a interação iniciada pelos usuários (conteúdo das mensagens) e os produtos construídos, ou seja, as mensagens foram trocadas não com o objetivo de esclarecerem dúvidas ou sugerir algo, mas apenas como forma de utilizar o mecanismo de suporte à comunicação. Um dos fatores que pode ter contribuído com isso deve-se ao conhecimento que estes participantes possuíam sobre os produtos. Além disso, pode-se atribuir este fato à apresentação das informações de percepção geradas a partir dos históricos de atividades que já estavam armazenados na memória de grupo. Isso pode ter auxiliado os usuários na realização da atividade.

Assim, acredita-se que a principal contribuição deste estudo de caso foi a oportunidade de avaliar, em um ambiente real, a usabilidade da aplicação. Pois, outro ponto observado foi que os usuários, ao iniciarem o desenvolvimento da atividade solicitada, dedicaram mais tempo para usar as funcionalidades da aplicação do que para desenvolver o produto. Acredita-se que o mais importante para eles naquele momento, era conhecer melhor as funcionalidades oferecidas pela Collaborative PL-Science.

Algumas ameaças à validade deste estudo também devem ser ressaltadas. Tanto as provas de conceitos quanto o estudo de caso realizado, foram desenvolvidos baseados em um problema real, a concepção de *workflows* científicos para sequenciamento/alinhamento genético.

Levando em consideração que, para a concepção destes *workflows*, o conhecimento e a qualificação necessários são específicos de um domínio, este estudo se propõe a ser válido dentro do contexto que foi especificado nesta dissertação. Sendo assim, não é possível generalizar os resultados para outros contextos e problemas enfrentados no domínio de *e-Science*. Com isso, faz-se necessário elaborar e realizar novos estudos para ampliar a validade das considerações feitas durante a realização desse trabalho.

Outro aspecto, observado através do Estudo de Caso, se relaciona à usabilidade da aplicação. Algumas considerações precisam ser feitas, por exemplo, os usuários do estudo de caso podem não ter interagido, ou não ter compreendido as informações oferecidas, considerando-se a maneira que elas estavam sendo apresentadas. Mesmo a interface escolhida para a apresentação das informações de percepção pode não ter facilitado o acesso aos recursos oferecidos pela aplicação, confundindo ou sobrecarregando os usuários com as informações apresentadas.

4.4 ANÁLISES DOS ESTUDOS

Conforme já mencionado, o propósito principal da realização destes estudos foi a caracterização e a obtenção de indícios preliminares a respeito da utilização de

informações de percepção e contexto para auxiliar os cientistas na concepção de *workflows* científicos e promover a colaboração entre eles. Sendo assim, através da realização destes estudos, foi possível obter evidências sobre a viabilidade da solução proposta nesta dissertação. Além disso, foi possível obter indícios sobre o apoio que as informações de percepção e contexto podem oferecer para um dado domínio de aplicação, neste caso específico, o sequenciamento/alinhamento genético. Através destes elementos, o ambiente de trabalho compartilhado pode manter-se atualizado, a “correção” de artefatos pode ser sugerida, o custo e o tempo na concepção de novos produtos científicos podem ser minimizados, pois os usuários puderam realizar suas atividades apoiados em artefatos já analisados e comentados por outros pesquisadores. Com isso, foi possível constatar também que a hipótese levantada por este estudo é suscetível de ser explorada através de estudos experimentais adicionais.

Deve-se ressaltar também que, através da utilização da abordagem Collaborative PL-Science, conseguiu-se minimizar as lacunas deixadas pela PL-Science (COSTA et al. 2013). Por exemplo, a ausência de informações sobre o contexto que envolve as atividades que os cientistas estão realizando, foram supridas através da criação de uma memória de grupo, e a partir dela a apresentação de informações de percepção e contexto para os cientistas.

A ausência de um serviço para que os cientistas pudessem interagir entre si, também foi suprida através da inclusão de um mecanismo de suporte à comunicação oferecido pela Collaborative PL-Science. Com isso, agora os cientistas podem trocar experiências e habilidades que podem ser úteis na concepção de novos *workflows* científicos. A dificuldade para reutilização de artefatos do núcleo da LPSC, também foi tratada. Através do suporte oferecido pela solução proposta, os cientistas possuem agora informações sobre como o artefato foi utilizado, em quais produtos, por quais cientistas, o que pode auxiliá-los na realização das atividades.

Com relação aos pontos que precisam ser aprimorados na abordagem Collaborative PL-Science pode-se citar:

- Outros tipos de informações de percepção podem ser utilizados para contextualizar os usuários da aplicação sobre o ambiente de trabalho

compartilhado no qual estão inseridos, visto que uma tabela com os principais tipos informações de percepção já foi levantada nesta dissertação. Assim podem-se verificar outros tipos que podem ser necessários para auxiliar os usuários na realização das suas atividades.

- Outros mecanismos de suporte à colaboração podem ser conectados e avaliados na abordagem Collaborative PL-Science, como por exemplo, o tratamento da reputação em atividades colaborativas, recompensa, gerência de conflitos, entre outros;
- A interface desenvolvida para a aplicação também pode ser avaliada através destes outros estudos.

Além disso, é importante destacar que as considerações feitas sobre estudo nos permite apenas tratar os resultados obtidos como indícios preliminares de possíveis benefícios da utilização da abordagem Collaborative PL-Science no domínio científico.

4.5 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO

Este capítulo apresentou as avaliações realizadas envolvendo a abordagem Collaborative PL-Science no domínio de bioinformática, mais precisamente, na concepção de *workflows* para realizar o sequenciamento/alinhamento genético. A análise sobre estudos realizados (quatro provas de conceitos e um estudo de caso) apresentou as principais contribuições que puderam ser observadas neste cenário, bem como destacou os pontos que precisam ser aprimorados na aplicação e as ameaças à validade deste estudo.

Através destas avaliações e, de acordo com a hipótese levantada nesta dissertação, conseguiu-se obter indícios de que ao oferecer aos cientistas um serviço que apresente elementos de colaboração (informações de percepção, contexto e um mecanismo de suporte à comunicação) integrados ao núcleo de artefatos de uma Linha de Produtos de Software Científico, é possível criar oportunidades para

promover e apoiar a interação entre os participantes de um grupo. Com isso, pode-se ainda contextualizá-los sobre o ambiente compartilhado de trabalho no qual estão inseridos. Assim, foi possível constatar também que a proposta de solução apresentada nesta dissertação é suscetível de ser explorada através de estudos experimentais adicionais.

5 CONCLUSÕES

Experimentos científicos apoiados computacionalmente, conhecidos como experimentos *in silico*, têm sido bastante utilizados em projetos científicos, como por exemplo, o Projeto Genoma (PEVSNER, 2009). Entretanto, mais que oferecer apoio computacional para os cientistas no desenvolvimento das pesquisas, faz-se necessário oferecer mecanismos para que estes pesquisadores possam interagir entre si. Através dessas interações, equipes de cientistas geograficamente distribuídas podem, por exemplo, trocar experiências e colaborar em experimentos científicos.

Este trabalho apresentou uma abordagem, denominada Collaborative PL-Science, que integra elementos de colaboração e Linha de Produtos de Software no cenário de *e-Science*. Neste contexto, as informações capturadas durante a realização de atividades por pesquisadores e persistidas na memória de grupo, considerando o cenário de desenvolvimento de aplicações científicas, podem orientá-los na realização de experimentos. Com isso, a Collaborative PL-Science tem por objetivo apoiar a concepção de *workflows* científicos através de uma Linha de Produtos de Software Científico (LPSC), bem como gerar oportunidades de interação entre os cientistas durante a construção de novos *workflows*.

Neste capítulo, as contribuições da abordagem Collaborative PL-Science são apresentadas (seção 5.1), justificando a sua utilização no domínio de *e-Science*. Entretanto, diante da complexidade do cenário científico, algumas limitações também são identificadas, sendo descritas na seção 5.2. A seção 5.3 apresenta alguns trabalhos futuros que podem ser realizados com o objetivo de evoluir a solução proposta.

5.1. CONTRIBUIÇÕES

Como contribuições da abordagem Collaborative PL-Science pode-se destacar a construção de uma arquitetura para fornecer informações de percepção e contexto

aos cientistas e promover a interação entre eles. Além disso, é também oferecido suporte às decisões que foram tomadas ao se desenvolver um *workflow* científico. Através deste suporte, cientistas podem tomar decisões e desenvolver experimentos a partir de artefatos já analisados, comentados e persistidos por outros pesquisadores na LPSC. Mesmo durante a manipulação de artefatos do núcleo da LPSC, as informações persistidas na memória de grupo podem ser consultadas a qualquer momento por outros pesquisadores. Essas informações podem auxiliar, por exemplo, na transferência de habilidades e competências para outros cientistas.

A possibilidade de realizar novas descobertas sobre o domínio a partir dos dados coletados é outra contribuição, além de facilitar o reuso de artefatos de software, uma vez que os cientistas podem inserir comentários sobre suas experiências na utilização destes artefatos. A aproximação entre os cientistas, bem como o conhecimento sobre a forma de atuar de cada um deles também merece destaque, por conduzir a uma visão mais ativa e colaborativa deste participante no seu ambiente de trabalho.

Além disso, a extensão da arquitetura proposta provê agora a inclusão de novos serviços que podem ser conectados na abordagem. Assim, a utilização de uma Arquitetura Orientada a Serviços traz ganhos para a abordagem Collaborative PL-Science, principalmente por possibilitar sua evolução, suporte e manutenção. O reuso de serviços também é outra vantagem, o que possibilita o aumento da produtividade e o alinhamento com os desafios encontrados no domínio no qual a proposta está inserida. Assim, a solução proposta tem a possibilidade de responder de forma efetiva e rápida aos desafios enfrentados no domínio científico, onde geralmente as aplicações precisam ter flexibilidade para executar mudanças.

5.2. LIMITAÇÕES

No decorrer do desenvolvimento deste trabalho, pode-se destacar que uma das principais dificuldades foi encontrar apoio em especialistas no domínio, de forma a

facilitar o desenvolvimento da solução proposta e alinhá-la com a real necessidade do domínio científico.

No que diz respeito a limitações em relação a colaboração, a aplicação Collaborative PL-Science não trata neste primeiro momento a recomendação de artefatos, a recomendação de cientistas com os quais outros pesquisadores poderiam iniciar uma interação solicitando ajuda, por exemplo, o tratamento da reputação dos membros do grupo, entre outros.

Além destas limitações relacionadas à colaboração, atualmente a abordagem não prevê o tratamento da proveniência de dados, nem durante a composição do *workflow* científico através da Linha de Produtos de Software Científico e, nem no momento de execução dos experimentos na aplicação. Entretanto, abordar a execução de experimentos científicos já não fazia parte do escopo deste trabalho, mesmo assim, pode ser apresentado como uma limitação atual da abordagem, já que no futuro será necessário tratar este ponto para que a aplicação possa evoluir e apoiar os cientistas na realização desta atividade. Assim, vale ressaltar também a limitação de elementos de colaboração, tais como, ferramentas de apoio para auxiliar os usuários durante a execução dos *workflows* científicos construídos a partir da abordagem proposta.

Além disso, a Collaborative PL-Science não explora a conexão e/ou integração da ontologia modelada para a abordagem com outras ontologias já existentes, sendo esta também outra limitação da aplicação.

A complexidade de manipular grandes volumes de dados em ontologias também foi outra dificuldade encontrada neste cenário. Para isso, o framework Stardog (<http://stardog.com/>) foi avaliado nesta dissertação. Ele permite a criação de um banco de dados gráfico que utiliza RDF para dados, consultas SPARQL, e raciocínio OWL. Entretanto, a versão deste framework que foi utilizada não atendeu aos objetivos da proposta, devido a ontologia modelada para a abordagem do Collaborative PL-Science utilizar *Properties Chains*, e as consultas estruturadas para rodar na abordagem não funcionarem adequadamente neste framework.

Devido ao escopo desta pesquisa e às limitações tecnológicas, o driver construído para realizar a leitura de modelos de *features* na aplicação, foi

desenvolvido apenas para arquivos provenientes do plugin Feature IDE, utilizado no Eclipse. Sendo assim, atualmente a aplicação está limitada a realizar a leitura de modelos de *features* provenientes apenas da ferramenta Feature IDE.

Outra limitação da abordagem proposta, diz respeito ao mecanismo de suporte à comunicação oferecido pela Collaborative PL-Science. Atualmente este mecanismo não explora a presença simultânea dos pesquisadores no ambiente de trabalho compartilhado, o que dificulta de certa forma aproveitar a oportunidade da presença do cientista no espaço de trabalho para iniciar uma interação.

5.3. TRABALHOS FUTUROS

Como trabalhos futuros podemos citar:

- Através da realização das provas de conceitos, percebeu-se que outros tipos de informações de percepção podem ser adicionados na abordagem, bem como outros mecanismos de suporte à colaboração. Por exemplo, desenvolver um mecanismo que aproveite melhor a presença simultânea dos pesquisadores no ambiente compartilhado de trabalho para gerar novas oportunidades de interação entre eles. A interface desenvolvida para apresentar as informações de percepção e permitir a realização de consultas à ontologia da abordagem também pode ser aprimorada. Pode-se evoluir estas interfaces, a partir da compreensão sobre como está acontecendo a interação entre os cientistas ou ainda, através conhecimento da utilização que eles fazem dos recursos oferecidos.
- A Collaborative PL-Science também precisa evoluir no que se refere ao tratamento de grandes volumes de dados. Considerando as informações armazenadas no banco de dados relacional, o tratamento atual se mostra adequado, apesar de poder ser aprimorado. Pode-se, por exemplo, buscar

alternativas através da utilização de mecanismos de sumarização de dados e otimização de consultas.

- Sobre a evolução do contexto da abordagem Collaborative PL-Science, a ontologia utilizada na solução proposta para representar o contexto, pode ser conectada a outras ontologias. Se a diferença for terminológica, por exemplo, a ontologia da Collaborative PL-Science pode permitir que as regras sejam traduzidas, os termos alterados e assim funcione corretamente no novo ambiente compartilhado no qual esteja inserida. Outro exemplo pode ser através da criação de uma rede de ontologias. Uma rede de ontologias permite a conexão de outras ontologias necessárias para aplicação e para o domínio. A criação desta rede de ontologias pode permitir também o reuso de ontologias já existentes, como por exemplo, a ontologia RISYS, que aborda riscos de colaboração em projetos de softwares (LIMA & DAVID, 2013). Algoritmos também podem ser utilizados para tratar a interoperabilidade entre a ontologia proposta com outras ontologias (HINZ & PALAZZO, 2008).
- Outros trabalhos que envolvam a utilização de atividades colaborativas também podem ser tratados no domínio deste estudo. Com isso, pode-se tratar outros aspectos da colaboração no cenário científico. Por exemplo, o tratamento da reputação em atividades colaborativas e a recomendação de artefatos já utilizados, comentados e analisados por outros cientistas durante a realização das suas atividades.

REFERÊNCIAS

- ALARCÓN, R., GUERRERO, L. A., OCHOA, S. F., PINO, J. A., **Context in Collaborative Mobile Scenarios**. In: *Fifth International and Interdisciplinary Conference on Modeling and Using Context - CONTEXT'05, Workshop on Context and Groupware*, Paris, France, 2005.
- ARBEX, W. A. **Modelos computacionais para identificação de informação genômica associada à resistência ao carrapato bovino**. Tese de Doutorado. Doutorado em Engenharia de Sistemas e Computação. Universidade Federal do Rio de Janeiro, 2009.
- BAECKER, R. M., NASTOS, D., POSNER, I. R., ET AL. **The User-Centered Iterative Design of Collaborative Writing Software**. In: INTERCHI'93, pp. 399-405, Amsterdam, The Netherlands, p. 399-405, 1993.
- BASILI, V. **GQM Approach Has Evolved To Include Models**. IEEE Softw. 11,p. 1-8, 1994.
- BAZIRE, M., BRÉZILLON, P., **Understanding Context Before Using It**. In: 5th International and Interdisciplinary Conference, CONTEXT 2005, v. LNAI 3554, pp. 29-40, Springer Verlag, Paris, France, 2005.
- BOH, W-F., REN, Y., KIESLER, S., & BUSSJAEGER, R. **Expertise And Collaboration in the Geographically Dispersed Organization**. Organization Science, v. 18, p. 595-612, 2007.
- BORGES, M. R. S., PINO, J. A., SALGADO, A. C. **Requirements for Shared Memory in CSCW Applications**. In: Proceedings of the 10th Workshop on Information Technologies and Systems, Austrália, p. 211-216, 2000.

- BOSCH, J. **Software Product Families in Nokia.** In Software Product Lines: 9th International Conference - SPLC 2005, Rennes, France, p. 2-6, 2005.
- BULCÃO NETO, R. F., PIMENTEL, M. G. C. **Toward a Domain-Independent Semantic Model for Context-Aware Computing.** In: Proceedings of the 3rd IW3C2 Latin American Web Congress, IEEE Computer Society, p. 61-70, Buenos Aires, Argentina, 2005.
- CAMPOS, F., DAVID, J. M. N., BRAGA, R., NERY, T., & SANTOS, N. **Rede de Ontologias: apoio semântico a linha de produtos de objetos de aprendizagem.** In *Anais do Simpósio Brasileiro de Informática na Educação* Vol. 23, No. 1, 2012.
- CHEN, H., FININ, T., JOSHI, A. **Using OWL in a Pervasive Computing Broker.** In: Workshop on Ontologies in Agent Systems, AAMAS-2003, Melbourne, Austrália, 2003.
- CHIN, G., JR., & LANSING, C. S. **Capturing And Supporting Contexts For Scientific Data Sharing Via the Biological Sciences Collaboratory,** Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, New York: ACM Press, p. 409-418, 2004.
- CHRISTOPOULOU, E., GOUMOPOULOS, C., ZAHARAKIS, I., KAMEAS, A. **An Ontology-based Conceptual Model for Composing Context-Aware Applications.** In: Sixth International Conference on Ubiquitous Computing (UbiComp 2004), Workshop on Advanced Context Modelling, Reasoning and Management, Nottingham, England, 2004.
- CLEMENTS, P., NORTHROP, L. **Software Product Lines: Practices and Patterns,** Addison-Wesley, Boston, p. 608, 2002.

COHEN, S. **Predicting When Software Product Lines Pays**. Disponível em: <http://www.sei.cmu.edu/pub/documents/03.reports/pdf/03tn017.pdf>. Acesso: 20 de dezembro de 2013.

CORDEIRO, D., BRAGHETTO, K. R., GOLDMAN, A., KON, F. **Da Ciência à e-Ciência: Paradigmas da Descoberta do Conhecimento**. Revista USP, nº 97, p.71-80, 2013.

CORLEY, E., BOARDMAN, C., & BOZEMAN, B. **Design and the management of multi-institutional research collaborations: Theoretical implications from two case studies**. Research Policy, v. 35, p. 975-993, 2006.

COSTA, G. C. B. **Uma Abordagem Para Linha de Produtos de Software Científico Baseada em Ontologia e Workflow**. Dissertação de Mestrado – Universidade Federal de Juiz de Fora, Juiz de Fora / MG, 2013.

COSTA, G. C. B.; BRAGA, R. M.; DAVID, J. M. N.; CAMPOS, F. C. A.; ARBEX, W. **PL-Science: A Scientific Software Product Line**. In: International Conference on Computational Science, Barcelona. International Conference on Computational Science, p. 1-10, 2013.

CUMMINGS, J. N.; KIESLER, S. **Coordination costs and project outcomes in multi-university collaborations**. Research Policy, v. 36, n. 10, p. 1620-1634, 2007.

CZARNECKI, K., KIM, C. H. P., KALLEBERG, K. T. **Feature Models are Views on Ontologies**. In Proceedings of the 10th International on Software Product Line Conference, p. 41–51, Washington, DC, USA. IEEE Computer Society, p. 41-51, 2006.

DAVID, J. M. N. **Um Serviço de Percepção para uma Infra-estrutura de Desenvolvimento de Groupware** – Tese de Doutorado. Doutorado em Engenharia de Sistemas e Computação. Universidade Federal do Rio de Janeiro UFRJ/COPPE, Rio de Janeiro, 2004.

DAVIES, J.; STUDER, R.; WARREN, P. **Semantic Web Technologies: Trends and Research in Ontology-Based Systems**. John Wiley & Sons, 2006.

DEY, A. K. **Understanding and Using Context**. *Personal and Ubiquitous Computing*, v. 5, n. 1, p. 4-7, 2001.

DOURISH, P., BELLOTTI, V. **Awareness and Coordination in Shared Workspaces**. In: Proc. ACM Conference on Computer Supported Cooperative Work - CSCW'92, Toronto, Canada, p. 107-114, 1992.

DOURISH, P., BLY, S. **Portholes: supporting awareness in a distributed work group**. In: CHI'92 Conference Proceedings: ACM Conference on Human Factors in Computing Systems, Monterey, California, New York: ACM Press, p. 541-547, 1992.

DRURY, J., WILLIAMS, M. G. **A Framework for Role-Based Specification and Evaluation of Awareness Support in Synchronous Collaborative Applications**. In: *IEEE 11th Int. Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises - WETICE'02*, Pittsburg, Pennsylvania, USA, p. 12-17, 2002.

ELLIS, C. A.; GIBBS, S. J.; REIN, G. L. **Groupware: Some Issues And Experiences**. *Communications of the ACM*, New York, NY, USA, v. 34, n.1, p.38-58, 1991.

EMBL-EBI, 2012. Disponível em: <http://www.ebi.ac.uk>. Acesso em 15 jun. 2013.

FILHO, J.B.F., BARAIS, O., BAUDRY, B., VIANA, W., ANDRADE, R. M. C. **An Approach for Semantic Enrichment of Software Product Lines**. *Proceedings of the 16th International Software Product Line Conference - SPLC 2012*, Salvador, p. 188-195, 2012.

FUKS, H.; RAPOSO, A. B.; GEROSA, M. A. **Do Modelo de Colaboração 3C à Engenharia de Groupware.** In: *WEBMIDIA 2003 – Simpósio Brasileiro de Sistemas Multimídia e Web*, Trilha especial de Trabalho Colaborativo Assistido por Computador, 2003, Salvador/BA. Anais. p. 445-452, 2003.

FUKS, H.; RAPOSO, A. B.; GEROSA, M. A.; PIMENTEL, M.; LUCENA, C. J. P. **The 3C Collaboration Model.** In: Ned Kock. (Org.). *The Encyclopedia of E-Collaboration.* Hershey, p. 637-644, 2007.

GAUVIN, M., BOURRY-BRISSET, A. C., AUGER, A. **Context, Ontology and Portfolio: Key Concepts for a Situational Awareness Knowledge Portal.** In: *Proceedings of the 37th Hawaii International Conference on System Sciences - Track 4*, p. 40111b, 2004.

GAVA, T. B. S., MENEZES, C. S. **Uma ontologia de domínio para a aprendizagem cooperativa.** In: XIV Simpósio Brasileiro de Informática na Educação, Rio de Janeiro, Brasil, 2003.

GEROSA, M. A., FUKS, H., LUCENA, C. J. P. **Elementos de percepção como forma de facilitar a colaboração em cursos via Internet.** In: *Anais XII Simpósio Brasileiro de Informática na Educação - SBIE, Vitória - ES – Brasil*, p. 194-202, 2001.

GIMENES, I. M. S., TRAVASSOS, G. H. **O Enfoque de Linha de Produto para desenvolvimento de Software,** XXI Jornada de Atualização em Informática (XXII Congresso da SBC), 2002.

GUTWIN, C., STARK, G., GREENBERG, S. **Support for Workspace Awareness in Educational Groupware.** In: *Proceedings of Computer Supported Collaborative Learning Conference – CSCL'95, USA*, p. 1-8, 1995.

GUTWIN, C., GREENBERG, S. AND ROSEMAN, M. **Workspace Awareness in Real-Time Distributed Groupware: Framework, Widgets, and Evaluation.** In M. A. Sasse, J. Cunningham, and R. L. Winder (eds): Proceedings of the Eleventh Conference of the British Computer Society Human Computer Interaction Specialist Group-People and Computers XI (HCI'96), London, United Kingdom. London: Springer-Verlag, p. 281–298, 1996.

GUTWIN, C. GREENBERG, S. **A descriptive framework of workspace awareness for real-time groupware.** Journal of Computer Supported Cooperative Work (CSCW). In: *Springer*, v. 11, n.3-4, p. 411-446, 2002

GREENBERG, S., **Context as a Dynamic Construct,** Human Computer Interaction, v. 16, n. 2-4, p. 257-268, 2001.

GRUBER, T. R. **A Translation Approach To Portable Ontologies. Knowledge Acquisition,** p. 199-220, 1993.

GRUBER, Thomas R. **Toward principles for the design of ontologies used for knowledge sharing?** International journal of human-computer studies, v. 43, n. 5, p. 907-928, 1995.

HINZ, V. T., & PALAZZO, L. A. M. **Algoritmos para interoperabilidade entre ontologias** Dissertação de Mestrado. Pelotas: Universidade Católica de Pelotas, Programa de Pós-Graduação em Informática, 2008.

HURWITZ, J., BLOOR, R., KAUFMAN, M., HALPER, F. **Arquitetura Orientada a Serviços – SOA para Leigos.** Editora Alta Books, Rio de Janeiro, 2ª edição, 2009.

IVANOV, S. V.; KOVALCHUK, S. V.; BOUKHANOVSKY, A. V. **Workflow-Based Collaborative Decision Support for Flood Management Systems.** *Procedia Computer Science*, p. 2213-2222, 2013.

- JOHANSEN, M. F., FLEUREY, F., ACHER, M., COLLET, P., AND LAHIRE, P. **Exploring the Synergies Between Feature Models and Ontologies**. In International Workshop on Model-driven Approaches in Software Product Line Engineering, Jeju Island, South Korea. Lancaster University, p. 163–171, 2010.
- LARMAN, C. **Utilizando UML E Padrões: Uma Introdução à Análise e ao Projeto Orientados a Objetos e ao Desenvolvimento Interativo**, 3a edição, Bookman, 2008.
- LAURILLAU, Y., & NIGAY, L. **Clover architecture for groupware**. In: Proceedings of the 2002 ACM Conference On Computer Supported Cooperative Work, p. 236-245, ACM, 2002.
- LICCARDI, I., DAVIS, H. C., WHITE, S., AND SOUTHAMPTON, H. S. O. **CAWS: Visualizing awareness to improve the effectiveness of co-authoring activities**. Special issue of Collaborative Computing in IEEE Distributed Systems Online, 2008.
- LIECHTI, O. **Awareness and the WWW: an Overview**. In: Proc. Workshop on 'Awareness and the WWW' (CSCW '00), v. 21, ACM Press, New York, USA, p. 3-12, 2000.
- LIMA, M. P., DAVID, J.M. N. **RISYS Considerando a análise de riscos e os elementos de contexto em um ambiente colaborativo de gestão de projetos**. In: Simpósio Brasileiro de Sistemas de Informação, Salvador. SBSI 2011, 2011. p. 464-469, 2011.
- MATOS, E. E. **CelOWS: Um Framework Baseado Em Ontologias Com Serviços Web Para Modelagem Conceitual Em Biologia Sistêmica**. Dissertação de Mestrado. Mestrado em Modelagem Computacional – Universidade Federal de Juiz de Fora. 2008.

- MATTOSO, M., WERNER, C., TRAVASSOS, G., BRAGANHOLO, V., MURTA, L. **Desafios no Apoio à Composição de Experimentos Científicos em Larga Escala**, *Jornal Seminário Integrada de Software e Hardware – SEMISH*, v.9, 36p, 2009.
- MOREIRA, A., VIEIRA, V., DEL ARCO, J.C. **Sanar: A Collaborative Environment to Support Knowledge Sharing with Medical Artifacts**, *SBSC - Brazilian Symposium on Collaborative Systems*, São Paulo, p. 35-42, 2012.
- OLIVEIRA, F. F. **Uma Ontologia de Colaboração e suas Aplicações**. 2009. 127f. Dissertação (Mestrado em Informática). Departamento de Informática, Universidade Federal do Espírito Santo, Vitória, 2009.
- OLSON, G. M., ZIMMERMAN, A., & BOS, N. **Scientific Collaboration on the Internet**. The MIT Press, 2008.
- PEREIRA, A. F., DAVID, J. M. N., BRAGA, R., CAMPOS, F. **Uma Abordagem Para Integração de Elementos de Colaboração ao Núcleo de Artefatos de uma Linha de Produtos de Software Científico**. *Proceedings of Brazilian Symposium on Collaborative Systems*. v. 179, p. 16-23, 2013.
- PEVSNER, J. **Bioinformatics and Functional Genomics**. 2nd ed. Hoboken, N.J: Wiley-Blackwell, 2009.
- PIMENTEL, M.; FUKS, H. **Sistemas Colaborativos**. Rio de Janeiro: ELSEVIER, 2011.
- PRESSMAN, R. S. **Software Engineering: A Practitioner's Approach**, 7ed., McGraw Hill, 2010.
- PREUVENEERS, D., DEN BERGH, J. V., WAGELAAR, D., GEORGES, A., RIGOLE, P., CLERCKX, T., BERBERS, Y., CONINX, K., JONCKERS, V., DE BOSSCHERE, K. **Towards an Extensible Context Ontology for Ambient Intelligence**. *In: Ambient*

Intelligence: Second European Symposium, EUSAI 2004, LNCS 3295, Eindhoven, The Netherlands, p. 148-159, 2004.

PRINZ, W. **“NESSIE: An Awareness Environment for Cooperative Settings.** *In: Proc. of the Sixth European Conference on Computer Supported Cooperative Work (ECSCW'99), Copenhagen, p. 391-410, 1999.*

RANGANATHAN, A., MCGRATH, R. E., CAMPBELL, R. H., MICKUNAS, D. **Ontologies in a Pervasive Computing Environment.** *In: Workshop on Ontologies and Distributed Systems (part of the 18'th International Joint Conference on Artificial Intelligence (IJCAI 2003), Acapulco, México, 2003.*

REMMEL, H., PAECH, B., ENGWER, C., BASTIAN, P. **Supporting The Testing Of Scientific Frameworks With Software Product Line Engineering: A Proposed Approach.** SECSE '11 Proceedings of The 4th International Workshop On Software Engineering For Computational Science And Engineering, 2011.

RITTENBRUCH, M., **Atmosphere: A Framework for Context-Awareness,** *International Journal of Human-Computer Interaction, v. 14, n. 2, p. 159-180, 2002.*

SANTANEN, E.; KOLFSCHOTEN, G.; GOLLA, K. **The Collaboration Engineering Maturity Model.** *In: Hawaii International Conference on System Sciences (HICSS'06), USA, Washington, DC, IEEE, 2006. v. 1, p. 1-10, 2006.*

SCHMIDT, K. **Some Notes on Mutual Awareness.** Paper presented at COTCOS Awareness SIG Workshop, Paris, France, 1998.

SEGAL J., MORRIS, C. **Developing Scientific Software.** *Journal of Software, IEEE. In: IEEE Computer Society, v. 25, n. 4, p. 18-20, 2008.*

SOFTWARE ENGINEERING INSTITUTE (SEI) - **Framework for Software Product Line Practice**. Disponível em: <http://www.sei.cmu.edu>. Acesso: 03 de fevereiro de 2014.

STEGER, M., TISCHER, C., BOSS, B., ET AL. **Introducing PLA at Bosch Gasoline Systems: Experiences And Practices**. In: Software Product Line Conference (SPLC), 2004.

STEINMACHER, I., CHAVES, A. P., GEROSA, M. A. **Awareness Support in Distributed Software Development: A Systematic Review and Mapping of the Literature**, Computer Supported Cooperative Work, 2012.

VERTEGAAL, R., VELICHKOVSKY, B., VAN DER VEER, G. **Catching the Eye: Management of Joint Attention in Cooperative Work**. SIGCHI Bulletin, p. 87-92, 1997.

VIEIRA, V., MANGAN, M. A. S., WERNER, C. M.L., MATTOSO, M. L. Q. **Ariane: An Awareness Mechanism for Shared Databases**. In: Proceedings of the X International Workshop on Groupware, CRIWG2004, San Carlos, Costa Rica, p. 92-104, 2004.

VIEIRA, V., TEDESCO, P., SALGADO, A. C. **Representação de Contextos em Ambientes Colaborativos Usando Ontologia**. In: Anais do XVI Simpósio Brasileiro de Informática na Educação, Workshop Brasileiro de Tecnologias para Colaboração, 2005.

VIEIRA, S. **Gerenciamento de Contexto em Sistemas Colaborativos**. Monografia de Qualificação e Proposta de Tese de Doutorado, 2006.

VIEIRA, V., TEDESCO, P., SALGADO, A. C. **Percepção e Contexto**. In Mariano Pimentel; Hugo Fuks. (Org.). *Sistemas Colaborativos*. 1ed. São Paulo: Elsevier, 157-172, 2011.

WANG, X. H., GU, T., ZHANG, D. Q., PUNG, H. K. **Ontology Based Context Modeling and Reasoning using OWL**. In: Workshop on Context Modeling and Reasoning at II IEEE International Conference on Pervasive Computing and Communication, Orlando, Florida, 2004.

WRAY, K. B. **The Epistemic Significance of Collaborative Research**, Philosophy of Science, p. 150-168, 2002.

WULF, W. **The National Collaboratory**. In *Towards a National Collaboratory*. Unpublished report of a National Science Foundation Invitational Workshop, Rockefeller University, New York, 1989.

YIN, R. K. **Case Study Research: Design and Methods**. 4 ed. Thousand Oaks, CA: Sage, 2009.

YU, W., SMITH, S. **Reusability of Feature Software: A Program Family Approach**. Proceeding Of The 2009 ICSE Workshop On Software Engineering for Computational Science And Engineering, IEEE Computer Society, 2009.

ZHANG, W., JARZAB, S. **Reuse Without Compromising Performance: Industrial Experience From RPG Software Product Line For Mobile Devices**. In Software Product Lines: 9th International Conference (SPLC 2005), 2005.

ZHANG, J., KUC, D., LU, S. **CONFUCIUS: A Tool Supporting Collaborative Scientific Workflow Composition**, IEEE TRANSACTIONS ON SERVICES COMPUTING, Vol. 7, No. 1, 2014.