

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

Phillipe Israel Marques

**Apoiando a Composição de Serviços em um
Ecossistema de Software Científico**

Juiz de Fora

2017

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

Phillipe Israel Marques

**Apoiando a Composição de Serviços em um
Ecossistema de Software Científico**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Orientador: José Maria Nazar David

Juiz de Fora

2017

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

Marques, Phillipe Israel.

Apoiando a Composição de Serviços em um Ecossistema de Software Científico / Phillipe Israel Marques. -- 2017.

122 f. : il.

Orientador: José Maria Nazar David

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, ICE/Engenharia. Programa de Pós-Graduação em Ciência da Computação, 2017.

1. Ecossistema de Software. 2. e-Science. 3. Redes Sociais Científicas. 4. Composição de Serviços Web. 5. Reutilização. I. David, José Maria Nazar, orient. II. Título.

Phillipe Israel Marques

Apoiando a Composição de Serviços em um Ecossistema de Software Científico

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Aprovada em 23 de Agosto de 2017.

BANCA EXAMINADORA

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

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

Prof. D.Sc. Rodrigo Pereira dos Santos
Universidade Federal do Estado do Rio de Janeiro

AGRADECIMENTOS

A Deus por permitir alcançar este objetivo.

À minha família, em especial aos meus pais Amarildo Israel Rocha e Maria Aparecida Marques Israel por me apoiarem nos momentos de dificuldade.

Ao meu orientador, professor José Maria, pela paciência, dedicação e aprendizado propiciados. Agradeço também pelas diversas críticas que permitiram aumentar a qualidade do trabalho e por sempre ter acreditado no meu potencial.

Aos professores Fernanda Campos, Regina Braga, Marco Antônio e Victor Ströele pelos ensinamentos, críticas e sugestões que permitiram realizar este trabalho.

Ao professor Rodrigo Santos por aceitar participar da banca e por todas as contribuições que enriqueceram muito este trabalho.

Aos demais professores do Departamento de Ciência da Computação da UFJF, por todo o conhecimento adquirido durante os períodos de graduação e mestrado.

Aos amigos Claudio Lelis, Leandro Simões e Weiner Oliveira, pelo convívio, pelo conhecimento compartilhado durante as disciplinas e por suas contribuições durante todo o período de mestrado.

Aos amigos do NEnC, André Abdalla, Camila Paiva, Frâncila Neiva, Heleno Campos, Heitor Magaldi, Hugo Guércio, Iuri Carvalho, Jade Ferreira, Lenita Ambrósio, Leonardo Pereira, Márcio Tadeu, Marcio Arakaki, Maria Luiza Falci, Pedro Ivo, Tássio Sirqueira. Agradeço pelos excelentes momentos neste laboratório e pelas contribuições para a realização deste trabalho.

Aos demais amigos do PGCC pela convivência e contribuições durante o mestrado.

*“You are a child of the universe,
no less than the
trees and the stars;
you have a right to be here.”*

Max Ehrmann

RESUMO

A área de *e-Science* envolve a realização de experimentos científicos complexos, normalmente apoiados por *workflows*. Esses experimentos geralmente utilizam dados e recursos distribuídos, e podem ser apoiados por uma plataforma de ecossistema de software científico. Neste contexto, é necessário permitir que diferentes serviços web possam ser compostos, reutilizados, além de interoperarem na plataforma para tratar das complexidades dos experimentos. Entretanto, compor serviços em plataformas de ecossistemas é uma atividade complexa, considerando, sobretudo, os requisitos funcionais e não funcionais desses serviços. Diante disso, o objetivo deste trabalho é apresentar um mecanismo que busca apoiar a composição de serviços no contexto de um ecossistema de software científico. Para tanto, esse mecanismo é associado ao processo de criação de serviços da plataforma de ecossistema de software científico. Oferece elementos de visualização para representar os relacionamentos de dependência funcional e interoperabilidade entre os serviços. Além disso, utiliza a análise de redes sociais científicas para identificar potenciais colaboradores. Os pesquisadores identificados poderão interagir com o auxílio das visualizações existentes, no espaço de trabalho compartilhado, para avaliar as composições. Essa plataforma, denominada E-SECO, apoia as diferentes fases do ciclo de vida de um experimento científico. A partir desse mecanismo, cientistas interagem e analisam as relações entre serviços nas composições realizadas considerando, sobretudo, as métricas de dependência funcional e a interoperabilidade entre os serviços existentes em diferentes instâncias da plataforma. Visando avaliar o mecanismo para apoiar a composição de serviços, foram realizados estudos de caso na plataforma E-SECO.

Palavras-chave: Composição de Serviços Web. Reutilização. Interoperabilidade. Proveniência de Dados. Ecossistema de Software. *e-Science*. Redes Sociais Científicas.

ABSTRACT

The area of e-Science encompasses performing complex scientific experiments, usually supported by workflows. These experiments generally use distributed data and resources, and can be supported by a scientific software ecosystem platform. In this context, it is necessary to allow different web services to be composed, reused, and interoperate in the platform to deal with the complexities of the experiments. However, performing services composition on ecosystem platform is a complex activity which requires computational support, considering, above all, the functional and non-functional requirements of these services. Therefore, the goal of this work is to present a mechanism that aims to support services composition in scientific software ecosystem context. To this end, this mechanism is associated to the service construction process of the scientific software ecosystem platform. It also provides visualization elements to represent functional dependency and interoperability relationships between the services. In addition, it uses scientific social networks analysis to identify potential collaborators. The identified researchers may interact through the visualizations, in the shared workspace, to evaluate the compositions. This platform, named E-SECO, supports different phases of the scientific experiment life cycles. From this mechanism, scientists interact and analyze the relationships between services in compositions which were performed considering, above all, the functional dependency metrics and interoperability issues between existing services in different instances of the platform. In order to evaluate the mechanism to support services composition, case studies were carried out on the E-SECO platform.

Keywords: Web Services Composition. Services Reuse. Interoperability. Provenance-Data. Software Ecosystem. e-Science. Scientific Social Networks.

LISTA DE FIGURAS

1.1	Estrutura da Dissertação	21
2.1	Taxonomia para ECOS (BOSCH, 2009)	24
2.2	Ciclo de vida de um experimento científico (BELLOUM et al., 2011)	26
2.3	Ciclo de vida de um experimento em um ecossistema de software científico	27
2.4	Estruturas do Núcleo do PROV	29
2.5	Relações do PROV-DM	30
2.6	Relações causais da ontologia PROV-OEXT (SIRQUEIRA et al., 2016)	31
2.7	Inferências do PROV-OEXT (SIRQUEIRA et al., 2016)	32
2.8	Ontologia de serviço (MARTIN et al., 2004a)	35
2.9	Elementos de Colaboração (FUKS et al., 2003)	36
3.1	Processo de Desenvolvimento de Serviços em um Ecossistema de Software Científico (estendido de (ARAKAKI et al., 2016))	45
3.2	Etapa de Busca e Recuperação de Serviços na Plataforma E-SECO	46
3.3	Arquitetura Simplificada da Plataforma E-SECO com a adição do módulo de composição de serviços (estendido de (SIRQUEIRA et al., 2016))	49
3.4	Processo de Composição de Serviços na Plataforma E-SECO	53
3.5	Grafo gerado pelo SCView com relações de interoperabilidade e dependências entre serviços	56
3.6	Filtro utilizado para identificar interações entre pesquisadores	57
3.7	Métricas de centralidade global (WASSERMAN; FAUST, 1994)	58
3.8	Normalização min-max (HAN et al., 2011)	59
3.9	Análise de Métricas para um Serviço Específico	60
3.10	Identificação de Colaborações em uma Rede Social Científica	61
3.11	Métrica local para análise de colaboração entre pesquisadores (WASSERMAN; FAUST, 1994)	61
3.12	Diagrama de Componentes Relacionados ao SCView	63
3.13	Diagrama de Classes (Serviço SCView)	64
3.14	Esquema Relacional para Análise de Redes Sociais	65

3.15	Diagrama de Atividades para a Busca por Serviços	66
3.16	Diagrama de Atividades para a Busca por Serviços	67
3.17	Tela de Busca por Serviços em Diferentes Repositórios	68
3.18	Parâmetros Utilizados para a Busca em Repositórios na Plataforma E-SECO .	69
3.19	Serviços Encontrados a Partir da Busca	70
3.20	Metadados obtidos para o Serviço Selecionado	70
3.21	Dados de Proveniência obtidos para a Tarefa “geneID”	71
3.22	Anotação Semântica Automática de Métodos do Serviço Selecionado	72
3.23	Busca Por Serviços Relevantes	73
3.24	Serviços Relevantes Encontrados	73
3.25	Visualização de Rede Social com Relações de Colaboração	73
4.1	Dimensões relacionadas à colaboração no apoio à composição de serviços . . .	78
4.2	Análise Comparativa do Apoio Oferecido Pela Colaboração	88
4.3	Avaliação do Grafo Gerado no Estudo de Caso Regular	89

LISTA DE TABELAS

2.1	Trabalhos Relacionados à Estruturação de Dados em ECOS	39
2.2	Trabalhos Relacionados à Composição de Serviços	41
2.3	Trabalhos Relacionados ao Desenvolvimento e Colaboração em um ECOS . . .	43
4.1	Colaboração entre Participantes Durante o Estudo de Caso Piloto	84
4.2	Colaboração entre Participantes Durante o Estudo de Caso Regular	89

LISTA DE ABREVIATURAS E SIGLAS

ARS Análise de Rede Social

ECOS Ecosystema de Software

ECOSC Ecosystema de Software Científico

E-SECO *e-Science Ecosystem*

ESBC Engenharia de Software Baseada em Componentes

LPS Linha de Produto de Software

LPSC Linha de Produto de Software Científico

PRIME *PRagmatic Interoperability to MEaningful collaboration*

SOA *Service-Oriented Architecture*

UFJF Universidade Federal de Juiz de Fora

SUMÁRIO

1	INTRODUÇÃO	14
1.1	CONTEXTUALIZAÇÃO	14
1.2	MOTIVAÇÃO	16
1.3	OBJETIVO	19
1.4	ORGANIZAÇÃO	20
2	FUNDAMENTAÇÃO TEÓRICA	22
2.1	ECOSSISTEMA DE SOFTWARE	22
2.2	E-SCIENCE	25
2.2.1	Proveniência de Dados	28
2.3	ARQUITETURA ORIENTADA A SERVIÇO	33
2.3.1	Interoperabilidade	33
2.3.2	Composição de Serviços Web Semânticos	34
2.4	COLABORAÇÃO	36
2.5	TRABALHOS RELACIONADOS	37
2.5.1	Estruturação dos dados para busca em ECOS	37
2.5.2	Composição de Serviços	39
2.5.3	Desenvolvimento e Colaboração em ECOS	41
2.6	CONSIDERAÇÕES FINAIS DO CAPÍTULO	43
3	COMPOSIÇÃO DE SERVIÇOS EM UM ECOSSISTEMA DE SOFTWARE CIENTÍFICO	44
3.1	PROCESSO DE DESENVOLVIMENTO DE SERVIÇOS NA PLATAFORMA E-SECO	44
3.2	ATIVIDADE DE BUSCA E RECUPERAÇÃO DE SERVIÇOS	46
3.3	COMPOSIÇÃO DE SERVIÇOS	52
3.3.1	Análise de Dependências entre Serviços	57
3.3.2	Análise de Redes Sociais Científicas	60
3.3.3	Projeto Detalhado	63
3.3.4	A Proposta de Solução em Ação	65

3.4	CONSIDERAÇÕES FINAIS DO CAPÍTULO	74
4	AVALIAÇÃO DA PROPOSTA	75
4.1	DEFINIÇÃO DOS ESTUDOS DE CASO	75
4.2	PLANEJAMENTO DOS ESTUDOS DE CASO	76
4.2.1	Definição dos participantes	76
4.3	ESTUDO PILOTO	79
4.3.1	Resultados Obtidos do Estudo Piloto	80
4.3.1.1	Etapa 1	80
4.3.1.2	Etapa 2 - Estudo de Viabilidade.....	83
4.4	ESTUDO DE CASO	84
4.4.1	Resultados Obtidos do Estudo de Caso	85
4.4.1.1	Etapa 1	85
4.4.1.2	Etapa 2 - Estudo de Viabilidade.....	87
4.5	AMEAÇAS À VALIDADE	90
4.5.1	Validade Interna.....	90
4.5.2	Validade Externa.....	91
4.5.3	Validade de Constructo.....	91
4.5.4	Validade de Conclusão.....	91
4.6	CONSIDERAÇÕES FINAIS DO CAPÍTULO	91
5	CONCLUSÕES	92
5.1	BREVE RESUMO	92
5.2	CONTRIBUIÇÕES	92
5.3	LIMITAÇÕES	94
5.4	TRABALHOS FUTUROS	94
	REFERÊNCIAS	96
	APÊNDICES	106
A.1	PROCEDIMENTOS	107
A.2	TRATAMENTO DE RISCOS	107
A.3	BENEFÍCIOS E CUSTOS	108
A.4	CONFIDENCIALIDADE DA PESQUISA	108

A.5	PARTICIPAÇÃO	108
A.6	DECLARAÇÃO DE CONSENTIMENTO	108

1 INTRODUÇÃO

Este capítulo tem como objetivo apresentar a motivação para este trabalho, assim como o contexto relacionado ao campo de estudo. O objetivo desta dissertação e a estrutura deste documento também são apresentados.

1.1 CONTEXTUALIZAÇÃO

A área de *e-Science* demanda o suporte à realização de atividades distribuídas e multidisciplinares. Esta área tem como objetivo auxiliar cientistas na realização de experimentos científicos utilizando equipes geograficamente distribuídas (OLSON, 2009). O ciclo de vida desses experimentos, normalmente compreende atividades, como: (i) investigação do problema; (ii) prototipação do experimento; (iii) execução do experimento; e (iv) publicação dos resultados (BELLOUM et al., 2011).

O suporte à colaboração é necessário no contexto de *e-Science*, por permitir a ordenação de atividades na execução de tarefas que compõem a experimentação. Além disso, oportunidades de comunicação podem ser úteis por permitir que cientistas interajam com potenciais parceiros em suas pesquisas. Neste contexto, elementos que criam oportunidades de cooperação na área científica podem ser úteis também para o apoio à decisão. Consideremos, por exemplo, um contexto no qual cientistas de diferentes institutos de pesquisa estão geograficamente distribuídos. Se estes cientistas possuírem uma relação de coautoria ou trabalharem em áreas similares, uma oportunidade de colaboração pode surgir com o objetivo, comum, de conduzirem um experimento. Isto permite a eles reutilizarem os mesmos serviços científicos, ou serviços semelhantes.

Manikas (2016) define um Ecossistema de Software (ECOS) como: “A interação entre software e ator em relação a uma infraestrutura tecnológica comum, que resulta em um conjunto de contribuições e influencia direta ou indiretamente o ecossistema”. Considerando a necessidade de auxiliar cientistas em todas as etapas do ciclo de vida, a plataforma E-SECO (E-Science Software Ecosystem) foi proposta (FREITAS et al., 2015). Esta plataforma baseia-se na abordagem de um Ecossistema de Software Científico (ECOSC). No contexto de *e-Science*, Freitas et al. (2015) caracterizam um ECOSC por suas relações com fornecedores de software científico, institutos de pesquisa, órgãos de fomento, instituições

financiadoras e partes interessadas nos resultados de pesquisa.

No contexto de um ECOSC, reutilizar ou estender serviços já desenvolvidos é importante, para evitar o retrabalho. Quando muitos componentes são adicionados, a arquitetura do ecossistema deve ser capaz de acomodar o crescimento do sistema. Por este fato, a extensibilidade é também um atributo de qualidade que contribui para o sucesso do ecossistema (AMORIM et al., 2013).

Para auxiliar as atividades dos cientistas, E-SECO oferece serviços desenvolvidos de acordo com os conceitos de uma Linha de Produto de Software (LPS). LPS são definidas como um conjunto intensivo de sistemas de software que compartilham e gerenciam características comuns, satisfazendo necessidades de um segmento particular de mercado ou missão, desenvolvido a partir de um núcleo de artefatos comum (NORTHROP, 2002). Travassos et al. (2002) destacam que uma LPS envolve um conjunto de aplicações similares dentro de um domínio, permitindo identificar aspectos comuns e diferenças entre artefatos de software no decorrer do processo de desenvolvimento da LPS. No contexto científico, a utilização de LPS permite auxiliar cientistas na gerência de *workflows* (FREITAS et al., 2015). Além disso, considerando que o núcleo de artefatos possui itens reutilizáveis, uma LPS pode auxiliar desenvolvedores da plataforma E-SECO na utilização dos serviços disponíveis, analisando as variabilidades entre eles (NORTHROP, 2002).

O conceito de Arquitetura Orientada a Serviço (SOA) também está presente na plataforma E-SECO. SOA é uma das abordagens para se resolver os problemas de interoperabilidade (ANVAARI et al., 2013). A interoperabilidade pode ser vista como a capacidade de sistemas distintos trabalharem juntos visando alcançar um objetivo comum, com intervenções manuais mínimas (IDE; PUSTEJOVSKY, 2010) (USMAN et al., 2011). A plataforma E-SECO, visa apoiar, entre outros requisitos, a capacidade de composição e interoperabilidade entre serviços. Neiva et al. (2015) destacam que interoperabilidade oferecem os seguintes mecanismos, no contexto de um ECOSC: (i) suporte à descoberta de serviços, (ii) suporte à seleção de serviços, (iii) suporte à composição de serviços, (iv) suporte ao registro de novos serviços e (v) suporte à geração de *workflows* científicos. O suporte à interoperabilidade permite apoiar a composição na plataforma. Por meio da capacidade de composição, é possível reutilizar serviços desenvolvidos anteriormente. A composição de serviços é vista como uma forma de reuso (ERL, 2008).

1.2 MOTIVAÇÃO

O compartilhamento de dados tem se tornado essencial para a ciência, permitindo que dados de uma determinada pesquisa sejam disponibilizados para outros cientistas. Grandes experimentos podem envolver a reutilização de recursos desenvolvidos por outros cientistas. Desta forma, é possível aumentar a produtividade em experimentos. Em *e-Science* existe um interesse crescente em relação ao compartilhamento e reutilização de dados científicos (TENOPIR et al., 2015). Neste sentido, informações relacionadas a outros experimentos podem ser de grande importância para verificar o que está disponível no ecossistema. Isto pode auxiliar cientistas na tomada de decisões sobre reutilizar um serviço, analisando detalhes de como este serviço foi utilizado anteriormente e o contexto de utilização. *Workflows* compartilhados permitem a construção de blocos que podem ser reutilizados em novos experimentos. Porém, sozinhos eles não são suficientes para a condução do experimento. Eles requerem a adição de novos recursos, como dados de proveniência, por exemplo (BELHAJJAME et al., 2015). Dados de proveniência podem ser úteis para promover a confiabilidade e a reutilização. A plataforma E-SECO possui um meta-repositório que inclui dados de proveniência relacionados a serviços e *workflows* utilizados em experimentos. O uso de dados de proveniência permite que dados históricos de experimentos científicos sejam reutilizados. Estes dados são relevantes por proverem importantes informações contextuais para uma análise mais abrangente dos resultados científicos (LIM et al., 2010). Assim, é possível saber, por exemplo, em qual *workflow* um serviço específico foi utilizado e o experimento em que o *workflow* foi utilizado. A partir dos serviços identificados, é possível realizar composições, buscando alcançar reusabilidade.

Extensibilidade, interoperabilidade, reusabilidade e capacidade de composição de serviços são atributos diretamente relacionadas (ERL, 2008). Na plataforma E-SECO, serviços que foram utilizados em experimentos anteriores podem ser reutilizados. Para analisar as características destes serviços, podemos utilizar mecanismos que visam obter serviços capazes de se comunicar, a partir de mecanismos que analisam interoperabilidade entre eles. A extensibilidade se destaca neste sentido, pois as composições permitem que o sistema cresça com a adição de novos componentes. Um suporte à interoperabilidade entre serviços e a possibilidade de identificar dependências entre serviços, em uma composição, pode auxiliar na avaliação de composições. A interoperabilidade pode ser definida

como a capacidade de diferentes sistemas usarem os serviços uns dos outros efetivamente (POKRAEV, 2009).

Web services permitem prover um ambiente para sistemas, que são interoperáveis e com baixo acoplamento (MCGOVERN et al., 2003). A composição de serviços em SOA se concentra em agregar serviços interoperáveis, permitindo que objetivos de colaboração sejam satisfeitos (PESSOA et al., 2008). A arquitetura de um ECOS formaliza regras de interoperabilidade, permitindo que diferentes equipes trabalhem juntas, de maneira independente (BOSCH; BOSCH-SIJTSEMA, 2010). McGovern et al. (2003) destacam que a utilização de serviços preexistentes e testados melhora a qualidade de um sistema. Considerando as APIs (*Application Program Interfaces*) utilizadas na plataforma E-SECO, é possível obter serviços preexistentes que podem ser reutilizados e compostos. Isto auxilia a manter a extensibilidade da plataforma, contribuindo para o sucesso do ECOS. Amorim et al. (2013) analisam a extensibilidade, relacionando-a aos impactos gerados em APIs definidas na arquitetura, bem como as contribuições de APIs para o sucesso de um ECOS. Além disso, analisam que uma API não deve ser de difícil utilização, por ser importante para permitir a novos desenvolvedores adicionarem novas características na plataforma.

A plataforma E-SECO possui um processo de desenvolvimento que permite construir serviços e adicioná-los à plataforma (ARAKAKI et al., 2016). O processo de desenvolvimento inclui as seguintes atividades: (i) planejamento inicial, (ii) busca e recuperação de serviços, (iii) criação de serviços, e (iv) composição de serviços. Esta última permite que serviços que foram utilizados em experimentos anteriores possam ser reutilizados e compostos. Para tanto, é necessário prover suporte à colaboração entre diferentes grupos de pesquisa. Neste contexto, a identificação de relações entre diferentes cientistas, geograficamente distribuídos, é importante para auxiliar a avaliação destas novas composições. Acreditamos que uma solução que permita analisar e visualizar redes sociais científicas seja capaz de prover informações para auxiliar o desenvolvimento de serviços, especialmente, a atividade de composição de serviços. Para realizar a composição em um ECOS é necessário, inicialmente, buscar e recuperar serviços em diferentes repositórios. O processo de desenvolvimento aborda a busca e recuperação de serviços. No entanto, ele não considera a distribuição dos recursos entre diferentes repositórios integrados à plataforma. Além disso, a busca e recuperação é realizada sem um mecanismo formalmente definido. A necessidade de melhorar a atividade de busca e recuperação é resultante de um estudo

prévio utilizando o processo de desenvolvimento de serviços (ARAKAKI et al., 2016).

Por utilizar uma abordagem de Engenharia de Software Baseada em Componentes (ESBC) com enfoque em um ECOSC, o processo de desenvolvimento e reuso de componentes se torna complexo. A ESBC almeja desenvolver software a partir de componentes existentes e construir novos componentes que podem ser reutilizados (CRNKOVIC et al., 2011). Por este motivo, recuperar componentes utilizando um mecanismo formal permite a obtenção de um maior número de serviços candidatos para a composição. Isto ocorre porque a plataforma E-SECO não considera apenas repositórios internos. Repositórios externos integrados também necessitam ser considerados. *Web Services* podem ser obtidos de várias fontes que contribuem com a plataforma, tais como BioCatalogue¹ e MyExperiment². Diante disto, observa-se uma complexidade para reutilizar serviços neste contexto. Dessa forma, apoiar a composição de serviços pode aumentar a reusabilidade na plataforma, considerando os aspectos relacionados a um ECOSC.

Identificar e entender corretamente os ativos de software (componentes reutilizáveis) é importante para auxiliar na decisão de reutilização (MARSHALL et al., 2003). Além disso, Schots (2014) destaca a importância de desenvolvedores explorarem repositórios de reuso de maneira a identificar e entender ativos de software, buscando analisar suas dependências e comparar ativos similares. A visualização de software se mostra importante neste sentido, pois é capaz de auxiliar *stakeholders* (desenvolvedores e pesquisadores) envolvidos no processo de composição de serviços na tomada de decisões em relação às composições realizadas. Na plataforma E-SECO, a visualização pode ser utilizada de maneira a permitir que *stakeholders* compreendam como os serviços estão dispostos em composições. Além disto, a visualização permite analisar relações de interoperabilidade, dependências e importância de serviços específicos na rede de composições geradas. Buscando prover uma colaboração efetiva na plataforma, a visualização também pode ser utilizada para identificar redes de pesquisadores e, através da comunicação entre esses cientistas, apoiar a tomada de decisões sobre as composições de serviços realizadas na plataforma E-SECO.

Este trabalho é realizado no contexto de uma plataforma de ECOS, na qual existe uma necessidade de tratar vários aspectos técnicos que influenciam a composição de serviços. Além disto, aspectos sociais também necessitam ser tratados. A partir da análise de diferentes aspectos que influenciam o processo de desenvolvimento de serviços em um

¹<https://www.biocatalogue.org>

²<http://www.myexperiment.org>

ECOSC, observa-se uma complexidade para compor serviços neste contexto. Busca-se assim uma solução integrada a uma plataforma de ECOSC capaz de apoiar a composição de serviços no desenvolvimento de serviços, considerando a distribuição de recursos e a necessidade de uma colaboração efetiva entre pesquisadores para avaliar os serviços compostos na plataforma. Busca-se assim evoluir o processo de desenvolvimento de serviços existente na plataforma.

1.3 OBJETIVO

Este trabalho tem como objetivo evoluir o processo de desenvolvimento de serviços na plataforma E-SECO com a adição do módulo de suporte à composição de serviços. Desta forma, será possível auxiliar a reutilização, a interoperabilidade e a extensibilidade de serviços científicos. No contexto de um ECOSC, é necessário considerar inicialmente um processo de busca que permite obter serviços de diferentes repositórios. A evolução deste processo ajuda a realizar uma busca eficiente de serviços que permite novas composições. Para este propósito, este trabalho apresenta um mecanismo que permite buscar e recuperar serviços de diferentes repositórios, incluindo a anotação semântica dos serviços. Este mecanismo é associado ao serviço SCView (MARQUES et al., 2017), que permite visualizar composições realizadas e analisar redes sociais científicas. Desta forma, é possível avaliar as composições realizadas a partir do suporte à interação entre pesquisadores. A identificação de pesquisadores é realizada utilizando redes de cientistas formadas a partir de relações de coautoria. A solução apresentada neste trabalho busca apoiar a composição de serviços em um ECOSC, considerando os seguintes objetivos específicos:

1. Apoiar a busca de serviços em diferentes repositórios externos à plataforma (busca externa) e em outras instâncias da plataforma E-SECO (busca interna);
2. Permitir anotação semântica de serviços, buscando apoiar a composição de serviços;
3. Utilizar dados de proveniência, obtidos de experimentos científicos, associados aos serviços buscados, para a reutilização;
4. Utilizar o suporte à interoperabilidade sintática, semântica e pragmática entre os serviços registrados, para apoiar a composição;

5. Utilizar a análise semântica dos parâmetros relacionados à interface dos serviços, buscando obter serviços relevantes a partir dos graus de alinhamento entre parâmetros identificados na literatura;
6. Apoiar a visualização gráfica da composição de serviços baseada em grafos, permitindo analisar a interoperabilidade e as dependências entre serviços, bem como a importância de um serviço na rede, para a tomada de decisões em relação à utilização de serviços;
7. Integrar e adicionar novos recursos a um serviço que permite analisar uma rede social científica na plataforma, permitindo analisar relações de colaboração e identificação de pesquisadores capazes de avaliar as composições realizadas.

1.4 ORGANIZAÇÃO

Este trabalho está dividido em seis capítulos, conforme apresentado na Figura 1.1. O Capítulo 2 apresenta os aspectos teóricos da pesquisa, destacando os principais conceitos relacionados ao trabalho, tais como Ecossistemas de Software, *e-Science*, Arquitetura Orientada a Serviço e Colaboração. Posteriormente, realiza uma discussão sobre os trabalhos relacionados que serviram como base para a abordagem proposta. O Capítulo 3 apresenta o mecanismo proposto para apoiar a composição de serviços, incluindo detalhes sobre sua implementação e exemplos de utilização. O Capítulo 4 apresenta a avaliação do mecanismo apresentado, destacando o planejamento, a execução e conclusões obtidas em relação à proposta apresentada. O Capítulo 5 destaca as contribuições da pesquisa, as suas limitações e possíveis trabalhos futuros.

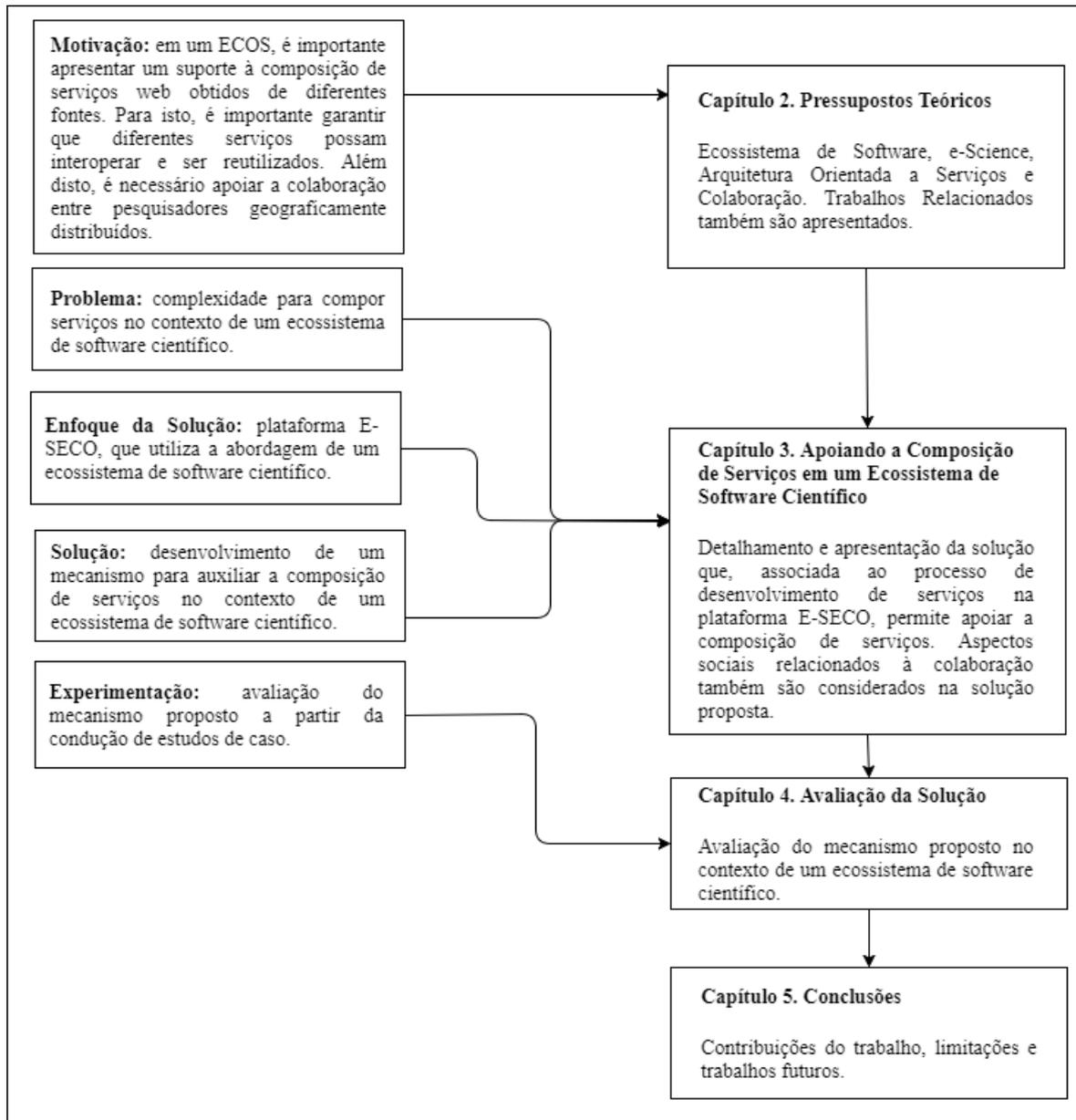


Figura 1.1: Estrutura da Dissertação

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo apresentar os principais conceitos abordados por este trabalho. Os conceitos relacionados às seguintes áreas: Ecossistemas de Software, e-Science, Arquitetura Orientada a Serviço e Colaboração são apresentados. Trabalhos Relacionados à solução proposta neste trabalho também são apresentados.

2.1 ECOSSISTEMA DE SOFTWARE

A noção de ecossistema tem origem na Ecologia, mas ao longo do tempo, passou a englobar ecossistemas humanos, tais como comerciais e sociais (BOSCH, 2009). Embora o termo Ecossistema de Software (do inglês *Software Ecosystem* ou SECO) seja derivado do arquétipo ecossistema natural, poucos trabalhos associam características entre ecossistemas naturais e ecossistemas de software (JANSEN et al., 2013).

Visando obter reuso de software em um ambiente intraorganizacional, Linhas de Produtos de Software (LPS) se tornaram importantes. No entanto, um novo cenário surgiu quando as organizações passaram a ter uma necessidade de disponibilizar sua plataforma em um contexto externo aos limites da organização (BOSCH, 2009). A interseção entre LPS e ECOS pode ser observada na literatura (ANVAARI et al., 2013) (MANIKAS, 2016). Uma LPS pode evoluir para um ECOS, com a LPS sendo a plataforma comum do ECOS (AOYAMA, 2014) (MANIKAS, 2016). A partir de desenvolvimentos e tendências associadas a uma engenharia baseada em LPS, originou-se uma reflexão sobre ECOS (JANSEN et al., 2013).

A literatura apresenta diversas definições para ECOS. Manikas and Hansen (2013), por exemplo, realizaram um mapeamento sistemática da literatura, através da qual buscavam obter uma visão geral deste campo e identificar áreas para pesquisas futuras. A partir desta revisão apresentam definições encontradas para um ECOS. (MESSERSCHMITT et al., 2005) definem ECOS da seguinte forma: “um ecossistema de software se refere a uma coleção de produtos de software que possuem um nível de relacionamento simbiótico”. Jansen et al. (2009) apresentam um ECOS como “um conjunto de negócios funcionando como uma unidade e interagindo com um mercado compartilhado de software e serviços, juntamente com as relações entre eles”. Bosch and Bosch-Sijtsema (2010) apresentam

um ECOS a partir da seguinte definição: “um conjunto de soluções que possibilitam, suportam e automatizam as atividades e transações realizadas pelos atores em um dado ecossistema social ou de negócio juntamente com as organizações que disponibilizam essas soluções”. Já (LUNGU et al., 2010) apresentam um ECOS como uma “coleção de projetos de software que são desenvolvidos e evoluem juntos em um mesmo ambiente”. Jansen and Cusumano (2012) apresentam um ECOS, em um alto nível de abstração, como sendo organizações relacionadas através de um software ou conceito de software, sendo subconjuntos de ecossistemas de negócio. A partir das diferentes definições para ECOS encontradas na literatura, Manikas and Hansen (2013) identificaram três elementos utilizados frequentemente: “software comum”, “negócios” e “relacionamentos de conexões”. A partir da combinação destes termos, apresentam um ECOS em termos da interação e colaboração entre atores que utilizam uma plataforma tecnológica comum, com uma série de soluções de software ou serviços.

Exemplos de ECOS são o ECOS mySQL/PHP, ECOS Microsoft, ECOS iPhone (JANSEN et al., 2009). Recentemente, observa-se também um crescimento do número de aplicativos para dispositivos móveis, através dos quais desenvolvedores externos podem contribuir com o desenvolvimento de aplicativos para ECOS como Android, IOS e Windows Phone (FONTÃO et al., 2016).

Bosch (2009) apresenta uma taxonomia que caracteriza um ECOS através de duas dimensões, tais como categoria e plataforma, conforme apresentado na Figura 2.1. A primeira dimensão (categoria) descreve o nível de abstração do ECOS, enquanto a segunda (plataforma) analisa a evolução industrial em termos de plataformas de computação dominantes. A dimensão categoria divide-se em três níveis: (i) sistema operacional: primeiro em que os ecossistemas de software foram identificados e gerenciados de forma explícita, (ii) aplicações de software: geralmente inicia a partir de uma aplicação *online* de sucesso, sem o auxílio de um ecossistema ao seu redor e (iii) programação para usuário final: menor e menos importante, porém fornece um ambiente de configuração no qual o usuário final pode criar as próprias aplicações requeridas. A dimensão plataforma se divide em (i) *desktop*, (ii) *web* e (iii) *mobile*, de acordo com a evolução de plataformas.

Campbell and Ahmed (2010) apresentaram uma visão em três dimensões de um ECOS. Esta visão foi posteriormente estendida por Santos and Werner (2011). As três dimensões identificadas são classificadas como: (i) técnica (ou arquitetural), (ii) negócios e (iii) social.

Programação para usuário final	MS Excel, Mathematica, VHDL	Yahoo! Pipes, Microsoft PopFly, Google's mashup editor	Nenhum
Aplicação de Software	MS Office	SalesForce, eBay, Amazon, Ning	Nenhum
Sistema Operacional	MS Windows, Linux, Apple OS X	Google AppEngine, Yahoo developer, Coghead, Bungee Labs	Nokia S60, Palm, Android, iPhone
<i>Categoria Plataforma</i>	Desktop	Web	Mobile

Figura 2.1: Taxonomia para ECOS (BOSCH, 2009)

Tais dimensões são definidas da seguinte maneira:

- **Dimensão Técnica:** é focada na plataforma de um ECOS. Relaciona-se ao desenvolvimento da arquitetura de um ECOS, através da qual busca-se desenvolver um produto múltiplo com o envolvimento de terceiros e utilizando uma plataforma organizacional comum. Assim, gerenciar características comuns e variabilidades em um ECOS, definir a arquitetura de uma LPS e realizar o processo de Engenharia de Domínio (estabelecendo ciclos de vida da plataforma) caracterizam alguns dos fatores que influenciam a arquitetura de um ecossistema.
- **Dimensão de Negócios:** é focada no fluxo de conhecimento. Esta dimensão é conduzida pelos fatores de “visão” (ideia de um futuro pretendido, relaciona-se a objetivos que devem ser alcançados), “inovação” (relaciona-se a melhoramentos contínuos no processo, criatividade e pioneirismo) e “planejamento estratégico” (foco dos esforços da organização, relaciona-se à elaboração de objetivos estratégicos para atingir o nível desejado em uma área específica).
- **Dimensão Social:** é focada em *stakeholders*. Esta dimensão permite observar características relacionadas às interações entre os *stakeholders*. Assim, busca-se entender porque eles integram, estendem e modificam o conhecimento em um ECOS. Esta dimensão também analisa como as capacidades e o envolvimento das partes interessadas estão reconhecidas de maneira implícita ou explícita. O conhecimento nesta dimensão é gerado a partir da colaboração e outras oportunidades geradas por redes sociais que contribuem para os *stakeholders*.

Na perspectiva de ECOS, questões sociais passaram a ser mais abordadas no desenvolvimento de software, concentrando-se em ecossistemas centrados em comunidade, onde desenvolvedores e usuários compartilham interesses em comum, criando uma rede (SANTOS et al., 2016). Módulos baseados em colaboração apoiam interação e comunicação entre as partes interessadas na plataforma (RODRIGUES et al., 2011).

No caso da plataforma E-SECO, que também utiliza uma abordagem de ECOS no contexto de *e-Science*, Freitas et al. (2015) apontam a definição apresentada por Bosch and Bosch-Sijtsema (2010) como a mais adequada para caracterizar essa plataforma. Nesta definição um ECOS é caracterizado como “uma plataforma de software, um conjunto de desenvolvedores internos e externos e uma comunidade de especialistas de domínio a serviço de uma comunidade de usuários que compõem elementos de soluções relevantes para satisfazer suas necessidades”. Assim, a plataforma considera os desenvolvedores internos e externos que atuam na plataforma e criam recursos para atender à comunidade científica. Freitas et al. (2015) apresentam um ECOSC como um subconjunto de ECOS, caracterizado por suas relações com fornecedores de software científico, institutos de pesquisa, órgãos de fomento, instituições financiadoras e partes interessadas nos resultados de pesquisa.

2.2 E-SCIENCE

A realização de experimentos científicos necessita da colaboração entre cientistas geograficamente distribuídos, envolvendo também a utilização de dados distribuídos e recursos computacionais (BELLOUM et al., 2011). Redes de pesquisadores colaborativas e distribuídas são vistas como o caminho do futuro para a pesquisa científica. Neste cenário, o compartilhamento de dados se mostra essencial para apoiar o avanço científico (TENOPIR et al., 2015). Para os pesquisadores o compartilhamento de dados irá depender do contexto. Suporte institucional, infraestrutura tecnológica e interações com outros pesquisadores afetam o desejo e capacidade de compartilhamento de dados (TENOPIR et al., 2011) (KIM; STANTON, 2012).

Por permitir realizar atividades distribuídas, a área de *e-Science* auxilia cientistas na realização de experimentos científicos utilizando equipes geograficamente distribuídas (OLSON, 2009). Esta área relaciona-se ao uso de recursos computacionais para apoiar a realização de experimentos científicos (HEY; TREFETHEN, 2003). Atualmente, a área

de *e-Science* busca garantir que dados sejam acessíveis, interoperáveis e reutilizáveis a longo prazo (WILKINSON et al., 2016). Tenopir et al. (2011) destacam a importância de compartilhar e reutilizar dados como boa prática para a pesquisa científica, que inclui fases que envolvem gerar, coletar, analisar e compartilhar dados. Belloum et al. (2011) apresentam o ciclo de vida de um experimento científico. Este ciclo é ilustrado na Figura 2.2. A experimentação se inicia a partir da investigação do problema, onde é definido o escopo da pesquisa. Posteriormente, ocorre a prototipação do experimento, onde se desenvolve os componentes e *workflows* necessários. As duas últimas etapas consistem em executar o experimento e, por último, publicar os resultados obtidos. Todas as etapas utilizam repositórios compartilhados, necessitando de suporte à colaboração e interações.

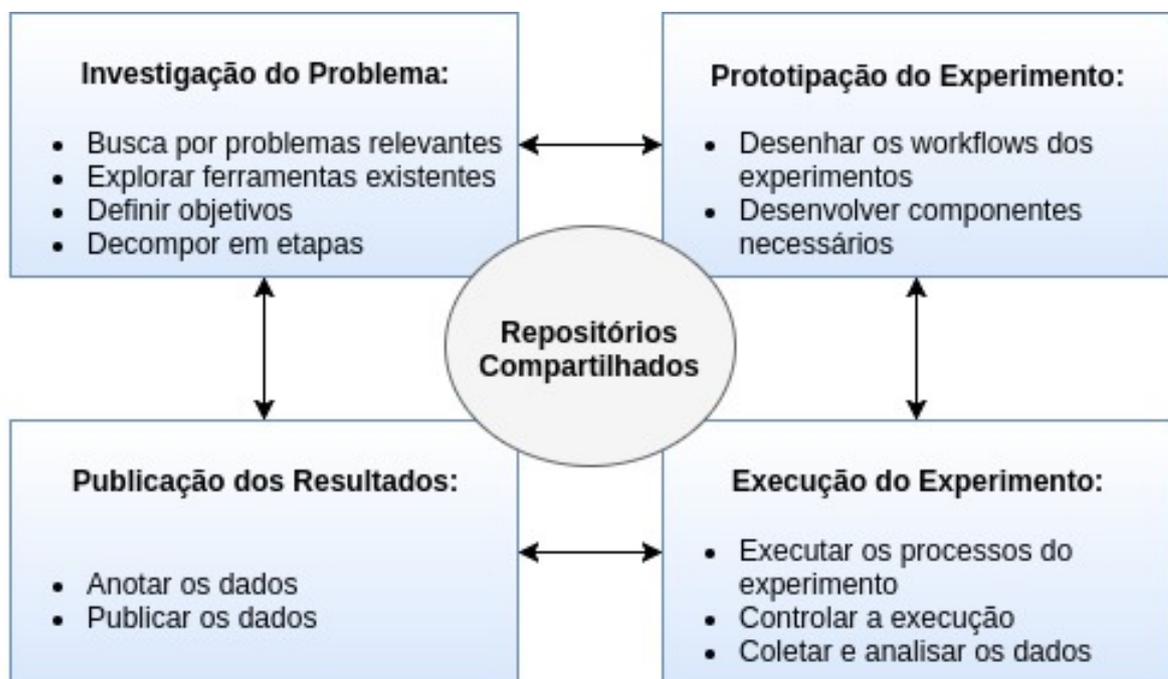


Figura 2.2: Ciclo de vida de um experimento científico (BELLOUM et al., 2011)

A partir dos conceitos abordados no ciclo de vida do experimento proposto por Belloum et al. (2011), Freitas et al. (2015) estenderam o ciclo de vida para o contexto de ecossistema de software. Neste contexto, a proposta de revisões sistemáticas da literatura é utilizada, a qual visa auxiliar a etapa de investigação do problema. Além disso, conceitos de linha de produto de software científico (LPSC) são apresentados na fase de prototipação de um experimento. O suporte à análise de proveniência dos dados também é apresentado, visando compartilhar dados sobre experimentos com outros pesquisadores.

A Figura 2.3 apresenta o ciclo de vida de um experimento científico no contexto de

um científico ECOSC, com as ferramentas associadas a cada etapa. Este ciclo de vida destaca a possibilidade de realização de uma revisão sistemática da literatura para apoiar a investigação do problema. Essa revisão pode ser realizada utilizando a ferramenta Parsifal¹. Ainda, a Figura 2.3 permite observar como os repositórios de *workflows* científicos podem apoiar todas as etapas do ciclo de vida. Sistemas de Gerenciamento de *Workflows* Científicos (SGWfCs) que podem ser utilizados para a execução de *workflows* também são apresentados. Após a execução de *workflows* dados de proveniência sobre o *workflow* são capturados e persistidos no repositório da plataforma, permitindo análises futuras.

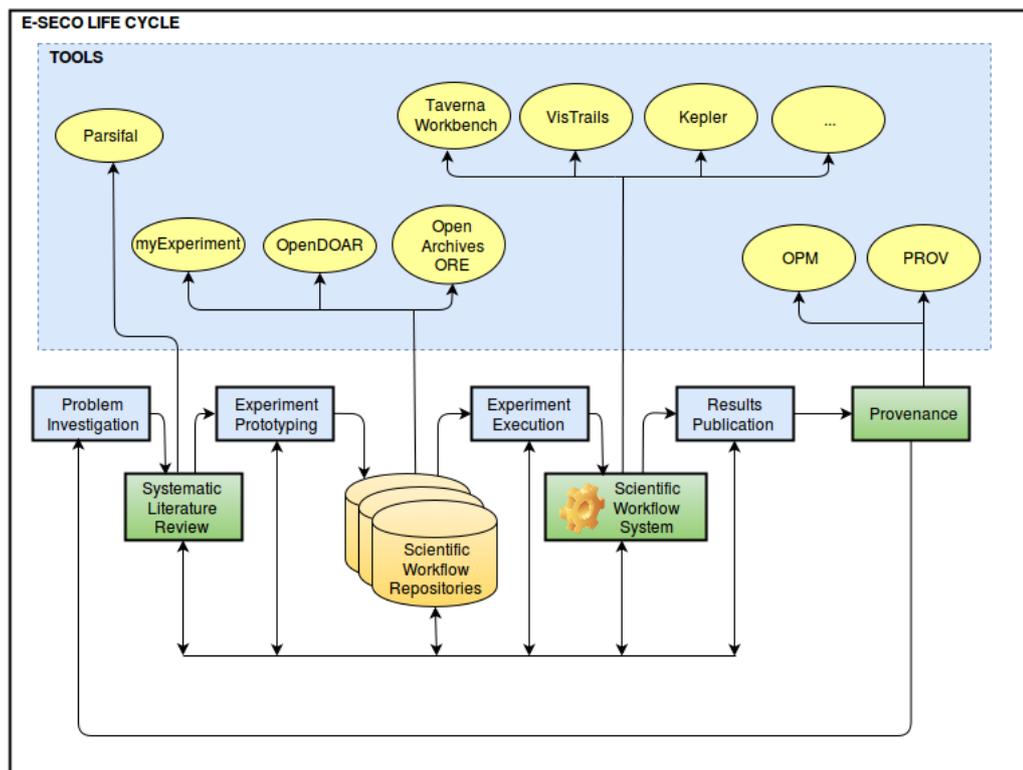


Figura 2.3: Ciclo de vida de um experimento em um ecossistema de software científico

Workflows científicos têm como objetivo automatizar ciclos repetitivos em experimentos científicos de forma a facilitar cientistas em suas pesquisas, ao invés de gerar preocupações com questões que envolvem recursos computacionais (DEELMAN et al., 2009). Em experimentação científica, *workflows* permitem executar uma série de tarefas a partir da utilização de serviços web. Assim, torna-se possível recuperar, integrar, analisar e visualizar os resultados (GOBLE et al., 2010).

SGWfCs foram desenvolvidos visando auxiliar a composição e execução de análises complexas e recursos distribuídos (DEELMAN et al., 2009). Estes sistemas apresentam

¹<https://parsif.al/>

um conjunto de ferramentas que permitem apoiar a definição e execução de *workflows* (SILVA et al., 2006). Como exemplos para SGWfC podem ser citados: Galaxy (GOECKS et al., 2010), Kepler (ALTINTAS et al., 2004), Pegasus (DEELMAN et al., 2004), VisTrails (FREIRE et al., 2006) e Taverna (OINN et al., 2007).

Embora os SGWfCs permitam tratar a utilização de *workflows* em um contexto de *e-Science*, uma perda de conhecimento sobre o experimento pode ocorrer. Isto acontece devido a necessidade de documentação adequada, considerando que estes sistemas costumam tratar apenas a questão de execução de *workflows* (MARINHO, 2011). Alguns SGWfCs permitem capturar os passos do *workflow*, porém utilizam modelos proprietários para capturar traços de proveniência gerados nas execuções (SIRQUEIRA et al., 2016). Assim, torna-se necessário prover recursos de rastreabilidade em relação aos dados produzidos visando possibilitar uma validação posterior (MARINHO, 2011). Este recurso que permite fornecer informação histórica sobre os dados é conhecido como proveniência (FREIRE et al., 2008).

2.2.1 PROVENIÊNCIA DE DADOS

Buneman et al. (2001) classificam proveniência de dados como a descrição das origens de uma peça de dados e o processo pelo qual ele chegou a um banco de dados. Lim et al. (2010) apresentam os dois tipos de proveniência existentes: prospectiva e retrospectiva. A proveniência prospectiva é apresentada como sendo a responsável por capturar uma especificação de um *workflow* abstrato, tendo como objetivo realizar derivações de dados futuras. A proveniência retrospectiva, por sua vez, busca capturar os passos que foram executados. Ambos os tipos de proveniência possuem importância por permitirem uma análise mais abrangente de resultados científicos. Dois modelos padrão são utilizados para capturar proveniência: modelo OPM (MOREAU et al., 2011) e modelo PROV (MOREAU; MISSIER, 2013). Em *workflows*, a proveniência permite reproduzir resultados, compartilhar e reutilizar conhecimento na comunidade científica (DAVIDSON; FREIRE, 2008).

O modelo PROV possui um conjunto de documentos que permitem auxiliar a especificação do modelo de proveniência. Dentre estes documentos, destacam-se o PROV-DM (modelo de captura de dados), PROV-CONSTRAINTS (restrições aplicáveis ao modelo de dados) e o PROV-O (ontologia para mapeamento do modelo de dados).

No núcleo do PROV², a proveniência descreve o uso e produção de entidades por atividades, que podem ser influenciadas de várias maneiras pelos agentes. Esses tipos principais e seus relacionamentos são ilustrados pelo diagrama UML da Figura 2.4.

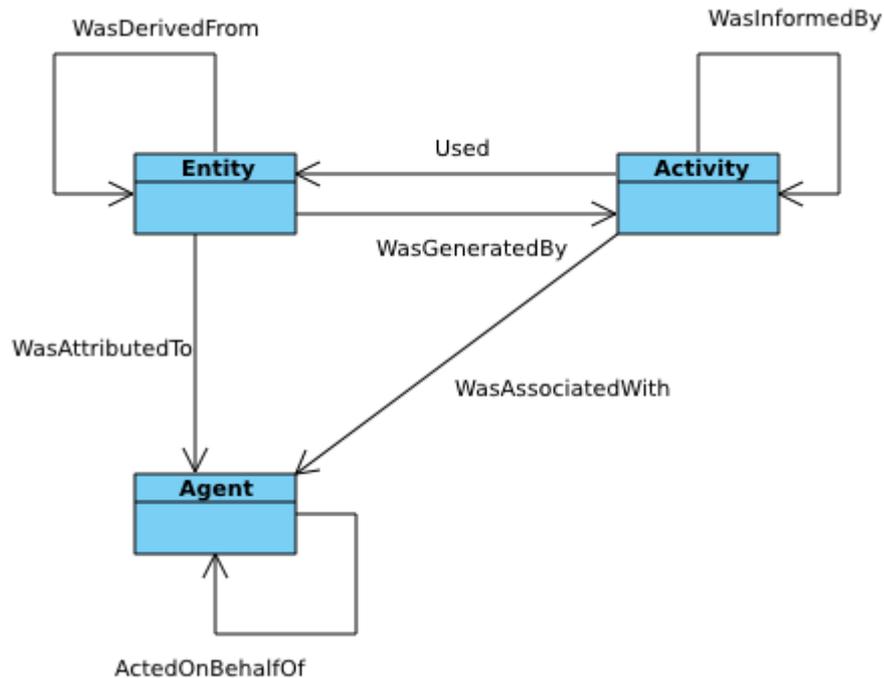


Figura 2.4: Estruturas do Núcleo do PROV

O PROV-DM separa os tipos e relações no modelo. Relações existentes no PROV-DM podem ser primárias ou secundárias como representação opcional. Os tipos apresentados no PROV-DM são:

- Entidade: tipo físico, digital, conceitual ou algo com aspectos fixos. Apresenta um identificador e atributos para representar informações adicionais sobre aspectos físicos da entidade. Entidades podem ser reais ou imaginárias;
- Atividade: algo que ocorre ao longo de um período de tempo e atua sobre entidades. Permite armazenar um identificador, o tempo de início e fim de uma atividade e um conjunto de pares de atributos que associados a valores para representar informações adicionais sobre uma atividade;
- Agente: algo que possui algum tipo de responsabilidade por atividade, para a existência de uma entidade, ou para a atividade de outro agente. Atividades estão

²<https://www.w3.org/TR/2013/REC-prov-dm-20130430/#term-entity>

associadas a agentes (relação “WasAssociatedWith”), enquanto entidades são atribuídas a agentes (relação “WasAttributedTo”).

As relações existentes no PROV-DM não são todas binárias, porém, envolvem dois tipos principais, tais como: primárias ou secundárias. A Figura 2.5-A apresenta as relações primárias, enquanto a Figura 2.5-B apresenta as relações secundárias. Na Figura 2.5-A, os nomes de Relações que aparecem em negrito correspondem ao núcleo das estruturas (apresentado previamente na Figura 2.5). Assim, é possível analisar como um tipo do modelo se relaciona com os outros e o nome das relações. As letras “R” e “L” aparecendo no lado direito de algumas células da Figura 2.5-A indicam que os atributos “prov: role” (função de uma entidade ou agente em relação a uma atividade) e “prov: location” (atributo opcional utilizada para identificar localização, que pode ser geográfica ou não) são permitidos para essas relações. A letra “T” indica que também é permitido um horário opcional. Existem relações PROV-DM que não são binárias e envolvem elemento opcional extra. Tais relações são apresentadas na Figura 2.5-B e agrupam objetos secundários, de acordo com seu tipo.

		Object			
		Entity	Activity	Agent	
Subject	Entity	WasDerivedFrom Revision Quotation PrimarySource AlternateOf SpecializationOf MentionOf	WasGeneratedBy WasInvalidatedBy	^R ^T _L	WasAttributedTo
	Activity	Used WasStartedBy WasEndedBy	^R ^T _L	WasInformedBy	WasAssociatedWith ^R
	Agent	—	—	ActedOnBehalfOf	

		Secondary Object		
		Entity	Activity	Agent
Subject	Entity	MentionOf (bundle)	WasDerivedFrom (activity) Revision (activity) Quotation (activity) PrimarySource (activity)	—
	Activity	WasAssociatedWith (plan)	WasStartedBy (starter) WasEndedBy (ender)	—
	Agent	—	ActedOnBehalfOf (activity)	—

Figura 2.5: Relações do PROV-DM

Na plataforma E-SECO, Sirqueira et al. (2016) utilizam o modelo de proveniência PROV, devido a capacidade de se prover muitas formas de conhecimento neste modelo.

Visando analisar informações que permitem indicar a manutenção e evolução de *workflows*, a ontologia PROV-O³ pode ser utilizada. Porém, a PROV-O original permite apenas capturar a proveniência retrospectiva. Devido a essas necessidades, Sirqueira et al. (2016) apresentaram a ontologia PROV-OEXT, a qual permite a extração e inferência de novos conhecimentos. Como resultado, esta ontologia permite identificar o histórico de *workflows*, indicando *workflows* predecessores e sucessores de um dado *workflow*.

A ontologia PROV-OEXT (SIRQUEIRA et al., 2016) passou a incluir classes como “Experiment” e “Task” como subclasses da classe “Activity”. A classe “Experiment” é responsável por associar dados sobre experimentos com *workflows*, enquanto a classe “Task” é responsável por armazenar informações sobre tarefas disponíveis que permitem compor os *workflows* registrados no sistema. “Workflow”, subclasse de “Entity” permite registrar dados de *workflows* para associá-los a experimentos. Registros sobre portas de comunicação das tarefas também passaram a ser incluídos, com a inclusão das classes “InputPort” e “OutputPort” como subclasses da classe “Task”. Dados sobre grupos de pesquisa também podem ser incluídos nesta nova ontologia, utilizando para isto “ResearchGroup”, subclasse de “Person”.

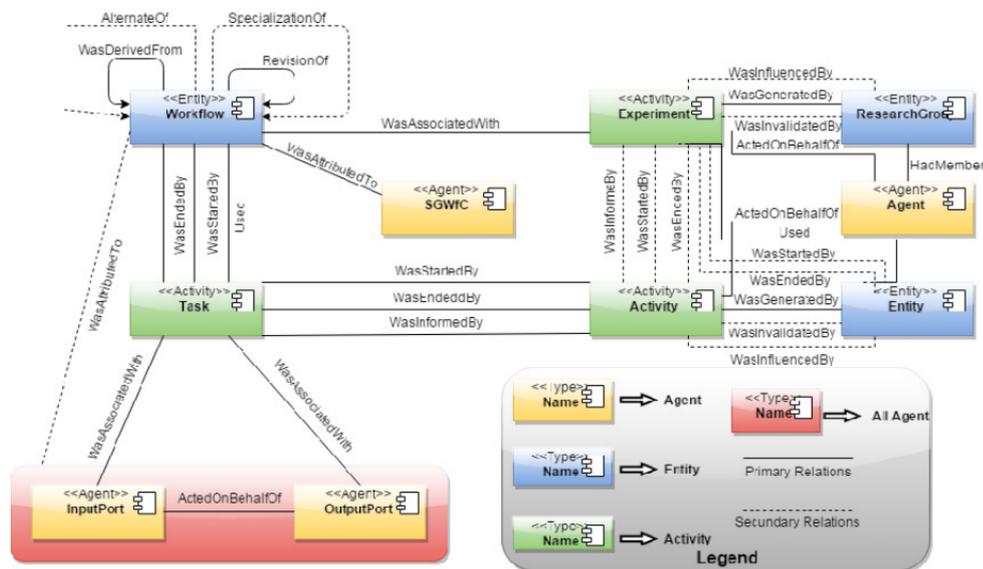


Figura 2.6: Relações causais da ontologia PROV-OEXT (SIRQUEIRA et al., 2016)

A Figura 2.6 apresenta as relações causais da ontologia PROV-OEXT. Nesta figura é possível observar que novas propriedades (*Object Properties*) também foram incluídas no PROV-OEXT. A propriedade “WasAssociatedWith”, por exemplo, foi inferida e permite

³<https://www.w3.org/TR/2013/REC-prov-o-20130430/>

associar *workflows* com experimentos e tarefas com portas de comunicação (entradas e saídas). Outra inferência realizada foi a relação “WasAttributedTo” relacionando portas de comunicação com *workflows*. Uma *property chain*⁴ permite realizar uma transitividade entre múltiplas propriedades. Assim, a *property chain* “EvolutionOf” foi adicionada. Buscando permitir uma inferência de conhecimento sobre evolução e manutenção de *workflows* e de configuração de experimentos, a expansão da ontologia conta também com as relações causais da PROV-O que foram modificadas. “Agent” se relaciona a algo que possui algum tipo de responsabilidade para uma atividade, para a existência de uma entidade, ou para a atividade de outro “Agent”. Dessa forma, é possível observar quais agentes (tipo “Agent”) são responsáveis por cada atividade. No caso, “All Agent” indica agentes que se relacionarão em uma atividade.



Figura 2.7: Inferências do PROV-OEXT (SIRQUEIRA et al., 2016)

Um exemplo de inferências realizadas com a ontologia PROV-OEXT é apresentado na Figura 2.7. Nesta figura é possível observar que o *workflow* selecionado foi executado no SGWfC Kepler. Através das *Properties Chains* “EvolutionTo” e “EvolutionOf”, foi possível analisar quais *workflows* foram influenciados pelo *workflow* analisado. Isto auxilia na reutilização deste *workflow*. Assim, os novos conhecimentos obtidos podem ser fundamentais para auxiliar pesquisadores em novos experimentos. Isto pode contribuir para apoiar as dimensões técnicas e sociais do ecossistema. A dimensão técnica pode ser apoiada a partir de uma análise dos passos de execução do *workflow*. Assim, é possível

⁴https://www.w3.org/TR/owl2-primer/#Property_Chains

analisar dados, como portas de comunicação entre tarefas (serviços), e dados, que permitam analisar evolução e manutenção de *workflows*. Além disso, é possível identificar se o *workflow* foi executado corretamente. A dimensão social, por sua vez, é apoiada a partir de análises sobre a utilização de *workflows* associados a experimentos. Isto permite gerar conhecimento na plataforma, pois é possível identificar como o *workflow* foi utilizado, qual cientista o utilizou e a qual grupo de pesquisa pertence esse cientista.

2.3 ARQUITETURA ORIENTADA A SERVIÇO

Uma Arquitetura Orientada a Serviço, do inglês *Service-Oriented Architecture* (SOA), estabelece um modelo cujo objetivo é aumentar a eficiência, agilidade e produtividade de uma organização para a construção de serviços web (ERL, 2008). SOA consiste de componentes de computação fracamente acoplados, permitindo que as organizações aproveitem os serviços existentes e compartilhem dados entre aplicativos distribuídos (FOUAD et al., 2017). Como resultado, serviços podem interoperar para apoiar as atividades que são conduzidas em uma plataforma que implementa SOA.

2.3.1 INTEROPERABILIDADE

Interoperabilidade pode ser caracterizada por envolver múltiplas entidades, permitindo transferência de dados e troca de informações ou conhecimento visando prover e aceitar serviços (POKRAEV, 2009). No contexto de SOA, interoperabilidade pode ser definida como “a capacidade dos sistemas de software para usar os serviços de software uns dos outros efetivamente” (POKRAEV, 2009).

A criação de um sistema de informação deve representar uma linguagem para interagir com um sistema (POKRAEV, 2009). Morris (1938) caracteriza uma linguagem como sendo constituída de três partes: sintaxe, semântica e pragmática. Baseando nestas partes, Pokraev (2009) apresenta três níveis de interoperabilidade:

- **Interoperabilidade Sintática:** visa garantir que sistemas envolvidos em alguma comunicação usem o mesmo vocabulário e gramática para analisar sentenças trocadas;
- **Interoperabilidade Semântica:** visa garantir que diferentes termos tenham o mesmo significado (que se refiram a mesma coisa no mundo real)

- **Interoperabilidade Pragmática:** está relacionado a intenção de que uma mensagem enviada por um sistema cause o efeito esperado. Assim, serviços que estão neste nível de interoperabilidade compartilham as mesmas expectativas.

2.3.2 COMPOSIÇÃO DE SERVIÇOS WEB SEMÂNTICOS

Serviços são apresentados como programas de software independentes, com características de *design* distintas, enquanto a composição do serviço é vista como um agregado coordenado de serviços (ERL et al., 2017). A Web semântica busca prover uma linguagem que seja capaz de expressar dados e regras para realizar inferências sobre dados. Tais regras visam representar o conhecimento a ser exportado para a Web (BERNERS-LEE et al., 2001). Considerando a importância de descrições semânticas, serviços web semânticos possuem marcadores semânticos de serviços que os tornam compreensíveis e úteis. Assim, esta marcação permite auxiliar a composição e interoperabilidade de serviços (MCILRAITH et al., 2001).

Diferentes linguagens, protocolos e *frameworks* foram propostos com o objetivo de facilitar a reutilização de serviços web (WALI; GIBAUD, 2012). Dentre estas iniciativas, destacam-se UDDI⁵ (*Universal Description Discovery and Integration*), SOA, BPEL4WS (*Business Process Execution Language for Web Service*) (ANDREWS et al., 2003), SOAP⁶ (*Simple Object Access Protocol*) e WSDL (*Web Services Description Language*) (CHRISTENSEN et al., 2001). No entanto, tais iniciativas permitem descrever serviços apenas de maneira sintática. A utilização de anotações semânticas permite definir o que o serviço faz, além de incluir aspectos de comportamento e execução do serviço (WALI; GIBAUD, 2012).

A *Web Ontology Language* (OWL) é uma linguagem de marcação semântica para definição de ontologias (BECHHOFFER, 2009). Buscando adicionar características semânticas aos padrões de serviços web, a ontologia OWL-S foi proposta (MARTIN et al., 2004b). Nesta ontologia a descrição de serviços é apresentada em três módulos: ***ServiceProfile***, ***ServiceModel*** e ***ServiceGrounding***. O ***ServiceProfile*** descreve as funcionalidades básicas do serviço. O ***ServiceModel*** descreve como o serviço trabalha, ou seja, o que ele faz. O ***ServiceGrounding***, por sua vez, descreve como acessar o serviço (WALI;

⁵<http://uddi.xml.org/>

⁶<http://www.w3.org/TR/soap/>

GIBAUD, 2012) (LAKSHMI; DHAS, 2015).

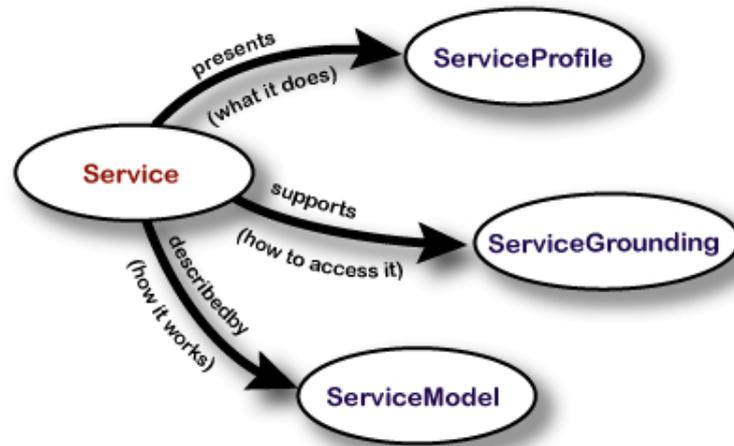


Figura 2.8: Ontologia de serviço (MARTIN et al., 2004a)

A ontologia OWL-S é apresentada na Figura 2.8. Nesta figura as comunicações com as classes *ServiceProfile*, *ServiceGrounding* e *ServiceModel* são apresentadas, respectivamente, com as propriedades *presents*, *supports* e *describedBy*. Assim, informações tais como parâmetros e URIs são descritas pelo *ServiceProfile*, enquanto informações sobre a utilização do serviço são descritas pelas classes *ServiceModel* e *ServiceGrounding*.

Um serviço web oferece uma funcionalidade limitada para seus usuários. Para alcançar o verdadeiro potencial de serviços web, é necessário compor vários serviços web em aplicações mais abrangentes, com funcionalidades mais sofisticadas. Para obter essas novas funcionalidades a partir de serviços existentes, são realizadas composições de serviços (MOGHADDAM; DAVIS, 2014). Desta forma, os serviços obtidos podem ser reutilizados de uma maneira mais eficiente, atendendo a um número maior de requisitos. No contexto de *e-Science*, a composição pode ser interessante quando se considerara a tendência atual em reutilizar e compartilhar dados científicos (TENOPIR et al., 2015).

A composição de serviços pode ser reduzida em três problemas fundamentais: (i) planejar como os serviços interagem e qual funcionalidade oferecem (planejamento), (ii) descobrir serviços que executam as tarefas necessárias no planejamento (descoberta) e (iii) gerenciar as interações entre os serviços (interação) (SYCARA et al., 2011). Neste sentido, o presente trabalho busca tratar tais questões. Desta forma, este trabalho apresenta recursos de visualização que permitem planejar e analisar a interação entre serviços, bem como suas funcionalidades. A partir de um mecanismo que considera os parâme-

tros relacionados a serviços e interoperabilidade em diferentes níveis é possível realizar a descoberta de serviços.

No presente trabalho, a anotação semântica de serviços visa auxiliar a composição com a utilização de recursos de web semântica. Esta anotação visa, neste caso, permitir a análise das funcionalidades dos serviços de maneira mais detalhada.

2.4 COLABORAÇÃO

Ellis et al. (1991) classificaram três dimensões de sistemas que oferecem suporte ao trabalho colaborativo. Estas dimensões foram formuladas inicialmente como comunicação, coordenação e colaboração. Posteriormente originaram o Modelo 3C de colaboração (PIMENTEL; FUKS, 2012) (FUKS et al., 2003), onde o que era chamado de “colaboração” passou a caracterizar operações conjuntas em um espaço de trabalho compartilhado e definido como “cooperação” (FUKS et al., 2007). O Modelo 3C de Colaboração inclui as dimensões de comunicação, coordenação e cooperação (PIMENTEL; FUKS, 2012).

Ferramentas que apoiam as dimensões do Modelo 3C de Colaboração oferecem suporte computacional para a colaboração (FUKS et al., 2007). O Modelo 3C é apresentado na Figura 2.9. A separação em dimensões foca nos aspectos relevantes para a análise de colaboração. Os “3 Cs” (comunicação, coordenação e cooperação) são apoiados pela percepção e também se inter-relacionam para que a colaboração ocorra (PIMENTEL; FUKS, 2012) (FUKS et al., 2007).

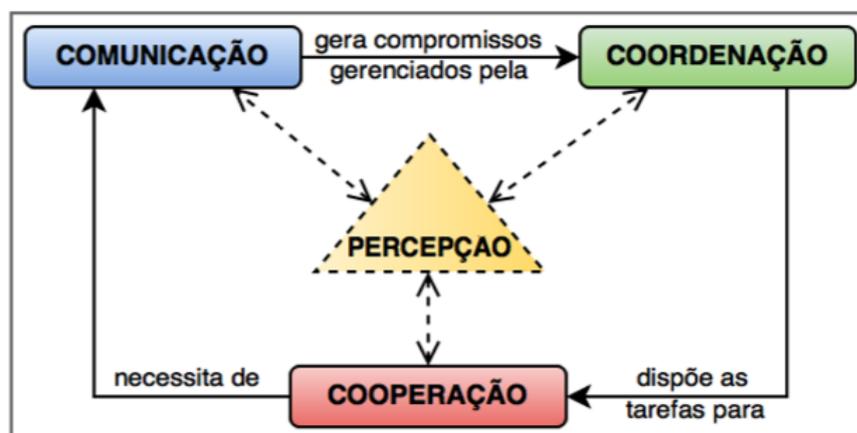


Figura 2.9: Elementos de Colaboração (FUKS et al., 2003)

Na Figura 2.9, a comunicação é caracterizada pela troca de mensagens e informação, pela argumentação e pela negociação entre as pessoas. As tomadas de decisão são reali-

zadas a partir da comunicação (voltada mais para a ação). A organização das atividades é importante para evitar o desperdício da comunicação. A coordenação está relacionada ao gerenciamento de pessoas, atividades e recursos e a cooperação é caracterizada pela atuação conjunta no espaço de trabalho compartilhado (PIMENTEL; FUKS, 2012).

2.5 TRABALHOS RELACIONADOS

Esta seção tem como objetivo apresentar trabalhos que serviram como base para a proposta apresentada. Esta seção está dividida em três subseções que apresentam as diferentes vertentes tratadas neste trabalho. Tais subseções são apresentadas como: (i) Estruturação dos Dados para Busca em ECOS, (ii) Composição de serviços e (iii) Desenvolvimento e Colaboração em ECOS.

2.5.1 ESTRUTURAÇÃO DOS DADOS PARA BUSCA EM ECOS

Quando se desenvolve serviços no contexto de um ECOS, deve-se considerar os aspectos que influenciam a composição de serviços. Uma das questões que deve ser tratada, relaciona-se à distribuição de recursos, já que serviços podem ser obtidos de diferentes instâncias da plataforma E-SECO, bem como de repositórios externos integrados à plataforma. Além disso, padronizar os dados também se mostra importante, principalmente por permitir ao cientista selecionar, visualmente, os serviços que melhor lhe atendem. A literatura apresenta trabalhos que tratam essas questões (JIMÉNEZ; WILKINSON, 2014) (SANTOS et al., 2013) (MOURA et al., 2012). Jiménez and Wilkinson (2014), por exemplo, consideram uma estruturação que visa representar artefatos obtidos de *workflows* e ontologias. Para isto, descrevem serviços e *workflows* através de anotações semânticas. Desta forma, termos ontológicos relevantes podem ser obtidos. Santos et al. (2013) apresentam um repositório que permite armazenamento, publicação e busca de componentes, além de mecanismos de recuperação para componentes e serviços reutilizáveis no contexto de um ECOS. Moura et al. (2012) apresentam um *framework* com dados obtidos de fontes distribuídas. Nele, os dados são coletados e realiza-se um mapeamento que torna possível a descoberta de informações importantes, baseando-se em ontologias.

Embora Jiménez and Wilkinson (2014) e Moura et al. (2012) tratem questões de padronização dos dados de diferentes fontes, tais trabalhos não se relacionam à utilização

de dados de proveniência, de maneira a permitir ao cientista observar em qual contexto o *workflow* ou serviço selecionado foi utilizado. Além disto, não possuem recursos para permitir analisar se diferentes serviços, obtidos de fontes distintas, são capazes de interoperar. Santos et al. (2013) abordam um processo de busca e recuperação de componentes em ECOS, mas não buscam apoiar a composição de serviços neste contexto. Assim, como observado na literatura, o presente trabalho apresenta uma solução capaz de obter e padronizar metadados relacionados a serviços obtidos de diferentes fontes e realizar a anotação semântica dos serviços. Porém, esta solução conta com recursos que permitem analisar proveniência de dados e interoperabilidade entre serviços, considerando um contexto científico. Tais características não foram encontradas na literatura. Assim, cientistas poderão observar em qual experimento determinado serviço ou *workflow* obtido foi utilizado. A interoperabilidade, por sua vez, permite analisar se diferentes serviços se comunicam. Assim, a solução proposta é capaz de apoiar a composição de serviços no contexto de um ECOS.

Delicato et al. (2013) apresentam a plataforma EcoDif. Essa plataforma integra dispositivos heterogêneos e busca criar um ecossistema de IoT (*Internet Of Things*) que busca integrar dispositivos físicos buscando prover controle de dados em tempo real, visualização, processamento e armazenamento. Considerando a necessidade de uma coleção de serviços web acessíveis, os autores observaram que dependências entre aplicações de serviços não eram totalmente conhecidas para o desenvolvimento, considerando o conjunto de serviços oferecidos e compostos. Observaram também a dependência de serviços e dados fornecidos por dispositivos que não podem ser acessíveis no momento de sua utilização. Isto gera uma necessidade de mecanismos que permitam identificar o contexto de aplicações. Na plataforma EcoDif, é proposta uma integração entre diferentes sistemas, informações, usuários e aplicações, criando um ecossistema de IoT e utilizando serviços obtidos de diferentes dispositivos. Entretanto, não realizam uma análise sobre a rede de desses componentes obtidos e também não tratam questões sociais relacionadas ao ecossistema. A Tabela 2.1 busca auxiliar a comparação dos trabalhos em relação à estruturação de Dados em ECOS. Para isto, os trabalhos são comparados em relação aos seguintes critérios: realização de anotação semântica de serviços, ECOS, padronização de dados de diferentes repositórios, busca e recuperação de componentes em ECOS, e distribuição de recursos. A solução proposta busca tratar todas estas questões.

Tabela 2.1: Trabalhos Relacionados à Estruturação de Dados em ECOS

	Anotação Semântica de Serviços	ECOS	Padronização de Dados de Diferentes Repositórios	Busca e Recuperação de Componentes em ECOS	Distribuição de Recursos
(Jiménez e Wilkinson, 2014)	X		X		X
(Moura et al., 2012)			X		X
(Santos et al., 2013)		X		X	X
(Delicato et al., 2013)		X			X
Solução Proposta	X	X	X	X	X

A partir da Tabela 2.1, a solução apresentada neste trabalho busca tratar questões como a busca e recuperação de componentes no contexto de ECOS, considerando a necessidade de padronização de dados existente devido à distribuição dos recursos. Além disso, uma anotação semântica dos serviços obtidos é realizada.

2.5.2 COMPOSIÇÃO DE SERVIÇOS

Mier et al. (2015) apresentam uma proposta para composição de serviços baseada em grafos. Neste trabalho é abordada a descoberta de serviços relevantes utilizando um algoritmo que analisa as entradas e saídas de um serviço de maneira semântica. Um *framework* para a geração de um grafo de composição é também apresentado. O método utilizado compara relações semânticas, a partir de entradas e saídas de uma requisição, a um conjunto de serviços web.

No caso do processo de desenvolvimento da Plataforma E-SECO, considera-se também a possibilidade de obter dados de proveniência sobre serviços. Por este motivo, a inclusão de palavras-chave relacionadas à descrição de um serviço pode ser útil para uma descoberta de serviços relevantes, considerando um contexto de utilização. Isto poderia auxiliar no ranqueamento de serviços relevantes visando reutilizar serviços para a composição. Por possuir um serviço de suporte à interoperabilidade em diferentes níveis, a plataforma E-SECO permite também que um contexto de utilização possa ser considerado na busca por

serviços interoperáveis. O suporte à interoperabilidade da E-SECO permite uma análise de serviços que possam ser interoperáveis em diferentes níveis (sintático, semântico e pragmático). Assim, além de considerar os graus de alinhamento entre serviços, como é realizado por Mier et al. (2015), a análise de interoperabilidade permite uma obtenção de um maior conjunto de serviços que possam ser relevantes para a composição.

Chen et al. (2017) analisam que muitos trabalhos utilizam apenas verbos e substantivos para descobrir serviços de maneira semântica. Porém, avaliam que a utilização de adjetivos e advérbios é capaz de contribuir para uma estimativa de similaridade mais efetiva. Assim, apresentam uma medida de similaridade semântica que inclui a inclusão de adjetivos e advérbios para analisar a similaridade semântica entre serviços disponíveis e uma requisição. Na plataforma E-SECO, a inclusão de adjetivos e advérbios para análise de similaridade se mostrou importante devido à utilização de dados de proveniência na anotação semântica dos serviços. Considera-se que serviços que já foram utilizados com sucesso em experimentos anteriores podem ser mais relevantes para cientistas. Adjetivos contidos na descrição textual do serviço podem permitir que um serviço seja classificado como sendo mais relevante para a composição. Desta forma, o termo “successful”, por exemplo, poderia indicar que um serviço foi utilizado com sucesso em um experimento anterior. Isto pode ranquear melhor um serviço, indicando uma maior possibilidade de reutilização em uma composição.

Naseri and Ludwig (2013) também utilizam dados de proveniência relacionados a serviços, *workflows*, suas especificações e detalhes de execução (tais como parâmetros de entrada e saída e objetos de dados) para a composição de serviços. Estes dados são utilizados para realizar o processo de composição de maneira mais eficiente. A importância da composição para prover a reutilização é destacada nesse trabalho. Dados de proveniência são usados para compor e selecionar serviços. Uma das camadas da arquitetura apresentada neste trabalho é responsável por capturar dados de proveniência da execução de um *workflow*. Esta captura inclui especificação dos serviços e parâmetros relacionados à execução. Embora este trabalho utilize dados de proveniência visando apoiar a composição de serviços, os dados de proveniência obtidos não estão relacionados a um contexto de *e-Science*. Desta forma, informações sobre um determinado experimento ou pesquisador não são obtidas. Na plataforma E-SECO os dados de proveniência visam oferecer mais informações contextuais aos *stakeholders* que a utilizam. Assim é possível tomar decisões

Tabela 2.2: Trabalhos Relacionados à Composição de Serviços

	Descoberta Semântica de Serviços	Composição de Serviços	Proveniência de Dados	Anotação Semântica de Serviços
(Mier et al., 2015)	X	X		
(Chen et al., 2017)	X	X		
(Naseri e Ludwig, 2013)		X	X	X
Solução Proposta	X	X	X	X

sobre reutilizar ou realizar composições para determinado serviço. A Tabela 2.2 busca comparar trabalhos que tratam questões relacionadas à composição de serviços. Para isto, os trabalhos são comparados em relação aos seguintes critérios: descoberta semântica de serviços, composição de serviços, proveniência de dados e anotação semântica de serviços.

Conforme observado na Tabela 2.2, a solução apresentada no presente trabalho considera a anotação semântica dos serviços utilizando dados de proveniência obtidos do meta-repositório de proveniência existente na plataforma E-SECO. Esses dados são incluídos na descrição dos serviços, buscando auxiliar a descoberta de serviços relevantes. A descoberta de serviços na solução é realizada de maneira semântica, sendo apoiada também por um serviço para análise de interoperabilidade em diferentes níveis. Um processo para composição de serviços também é apresentado na solução proposta.

2.5.3 DESENVOLVIMENTO E COLABORAÇÃO EM ECOS

Com o objetivo de apoiar o desenvolvimento de software em ECOS, um processo é apresentado por Fontão et al. (2016). Neste trabalho é tratada a necessidade de reformulação de processos por uma organização, com foco em ECOS para dispositivos móveis (MSECO). Para isso, um processo que visa auxiliar desenvolvedores externos no desenvolvimento de aplicativos, que estejam de acordo com os objetivos da organização, é apresentado. Embora a necessidade de se reestruturar processos seja apresentada, este trabalho não aborda a descoberta de serviços dentro de um processo de criação de serviços. Aspectos relacionados à proveniência de dados e a realização de anotação semântica visando compor novos serviços também não são tratados.

Além da capacidade de analisar proveniência de serviços e interoperabilidade entre

serviços, providas pela plataforma E-SECO, o apoio à composição de serviços envolve também a colaboração entre *stakeholders*. A análise de redes sociais a partir de visualização, quando utilizada neste contexto, pode identificar potenciais pesquisadores geograficamente distribuídos capazes de contribuir para o processo de desenvolvimento. A análise de redes sociais permite oferecer características colaborativas que apoiam o contexto social de um ECOS.

Manikas (2016) apresenta uma pesquisa através da qual identifica propostas relacionadas a redes sociais e colaborativas, no contexto de ECOS (SANTOS; WERNER, 2012) (SADI; YU, 2014). Santos and Werner (2012), por exemplo, apresentam uma rede socio-técnica, envolvendo as dimensões técnica e social de um ECOS. A evolução do processo de desenvolvimento de software é abordada por Sadi and Yu (2014). Na solução proposta pelos autores, as redes de atores são utilizadas para modelar o contexto sociotécnico de desenvolvimento. Como resultado, permite analisar as forças envolvidas no desenvolvimento de software, considerando aspectos técnicos e sociais. Entretanto, não exploram as especificidades do processo de composição de serviços, nem analisam as relações necessárias para apoiar esse processo como, por exemplo, interoperabilidade, reutilização, independência funcional, ausência de estado, entre outros princípios de SOA. Esse processo deve ser capaz de ajudar equipes de cientistas a reutilizar os serviços para a composição de novos serviços. No contexto de um ECOS, a composição de serviços obtida de diferentes fontes é fundamental, tendo em vista a complexidade e a diversidade dos requisitos que necessitam ser atendidos. Visando atender a dimensão social da plataforma E-SECO, busca-se identificar pesquisadores capazes de auxiliar na tomada de decisão. A colaboração entre pesquisadores de diferentes equipes gera oportunidades para trocar experiências sobre diferentes experimentos realizados em contextos semelhantes. É possível também difundir o conhecimento gerado em uma pesquisa para outras equipes. Assim, a interação na plataforma permite apoiar questões sociais, conectando pessoas geograficamente distantes e permitindo que elas comuniquem e colaborem. Desta forma, as equipes distribuídas geograficamente, podem contribuir com a tomada de decisão através da avaliação das composições realizadas. A Tabela 2.3 permite comparar trabalhos que tratam questões de desenvolvimento e colaboração em um ECOS. Para isso, são utilizados como critérios os processos de desenvolvimento aplicado a ECOS, a análise de redes sociais e colaborativas em ECOS e os aspectos técnicos e sociais também presentes em um ECOS.

Tabela 2.3: Trabalhos Relacionados ao Desenvolvimento e Colaboração em um ECOS

	Processo de Desenvolvimento de Software Aplicado a ECOS	Redes Sociais e Colaborativas em ECOS	Aspectos Técnicos e Sociais de um ECOS
(Fontão et al., 2016)	X		
(Santos e Werner, 2012)		X	X
(Sadi e Yu, 2014)	X	X	X
Solução Proposta	X	X	X

Como observado na Tabela 2.3, a solução apresentada neste trabalho se associa ao processo de desenvolvimento no contexto de um ECOS considerando aspectos técnicos e sociais nesse contexto. Para tratar os aspectos sociais, utiliza-se uma análise de redes sociais e colaborativas.

2.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Este capítulo apresentou os principais conceitos envolvidos na proposta deste trabalho, tais como: Ecossistemas de Software, *e-Science*, Arquitetura Orientada a Serviço e Colaboração. Trabalhos relacionados à proposta foram apresentados. Buscou-se analisar os diferentes aspectos que permitem apoiar a composição de serviços em um ECOSC. Um mecanismo que considera tais aspectos será apresentado no próximo capítulo, bem como o processo para apoiar a composição na plataforma.

Diante do cenário apresentado, o presente trabalho busca uma solução integrada ao processo de desenvolvimento de serviços que apoie as diferentes necessidades da composição de serviços em um ECOSC. Esta solução visa apoiar a composição de serviços utilizando como arcabouço para a dimensão técnica da plataforma os dados de proveniência, bem como serviços de recuperação e o suporte à interoperabilidade no contexto de uma plataforma de ecossistema de software científico. Buscando tratar questões sociais na plataforma, esta solução permite também analisar redes sociais científicas e oferecer o suporte à visualização das relações obtidas como resultado desta análise. Como resultado, buscaram-se oportunidades de colaboração entre pesquisadores geograficamente distribuídos para apoiar a composição de serviços em um ECOSC.

3 COMPOSIÇÃO DE SERVIÇOS EM UM ECOSISTEMA DE SOFTWARE CIENTÍFICO

Este capítulo apresenta um mecanismo para apoiar a composição de serviços na plataforma E-SECO. Inicialmente o processo de desenvolvimento de serviços é apresentado. Posteriormente, são destacadas as atividades de busca e recuperação de serviços e composição de serviços. Tais atividades compõem o processo de desenvolvimento apresentado. Além disto, a proposta apresentada considera também aspectos sociais da plataforma. Desta forma, interações entre pesquisadores podem apoiar o processo de composição de serviços.

3.1 PROCESSO DE DESENVOLVIMENTO DE SERVIÇOS NA PLATAFORMA E-SECO

Ecosistemas de software podem ser complexos, e envolver diferentes tecnologias e conceitos (BOSCH; BOSCH-SIJTSEMA, 2010). Como resultado, um ECOSC pode necessitar de uma diversidade de serviços para apoiar os cientistas nas diferentes etapas do ciclo de vida do experimento. Para auxiliar a descoberta e o desenvolvimento de serviços, pode-se utilizar um processo que visa recuperar, criar, evoluir e/ou compor serviços no contexto de um ECOSC. A Figura 3.1 apresenta uma visão geral do processo, utilizando a notação BPMN¹, que será descrito a seguir.

O processo se inicia com o planejamento (Figura 3.1-A), através do qual é gerada uma especificação das características desejadas para o serviço como, por exemplo, seus requisitos funcionais. Ao final dessa atividade, tem-se a especificação inicial que será utilizada na atividade seguinte. A segunda atividade é a busca e recuperação de serviços (Figura 3.1-B). Seus objetivos são encontrar serviços na plataforma E-SECO que atendam à especificação realizada na atividade anterior, e, caso sejam encontrados, são recuperados bem como os artefatos a ele relacionados.

Como resultado da atividade de busca e recuperação de serviços podem ocorrer quatro casos. No primeiro, nenhum serviço que atenda à especificação foi encontrado e, com isso, o serviço deverá ser desenvolvido desde o início (Figura 3.1-C). No segundo caso, encontra-se

¹<http://www.bpmn.org/>

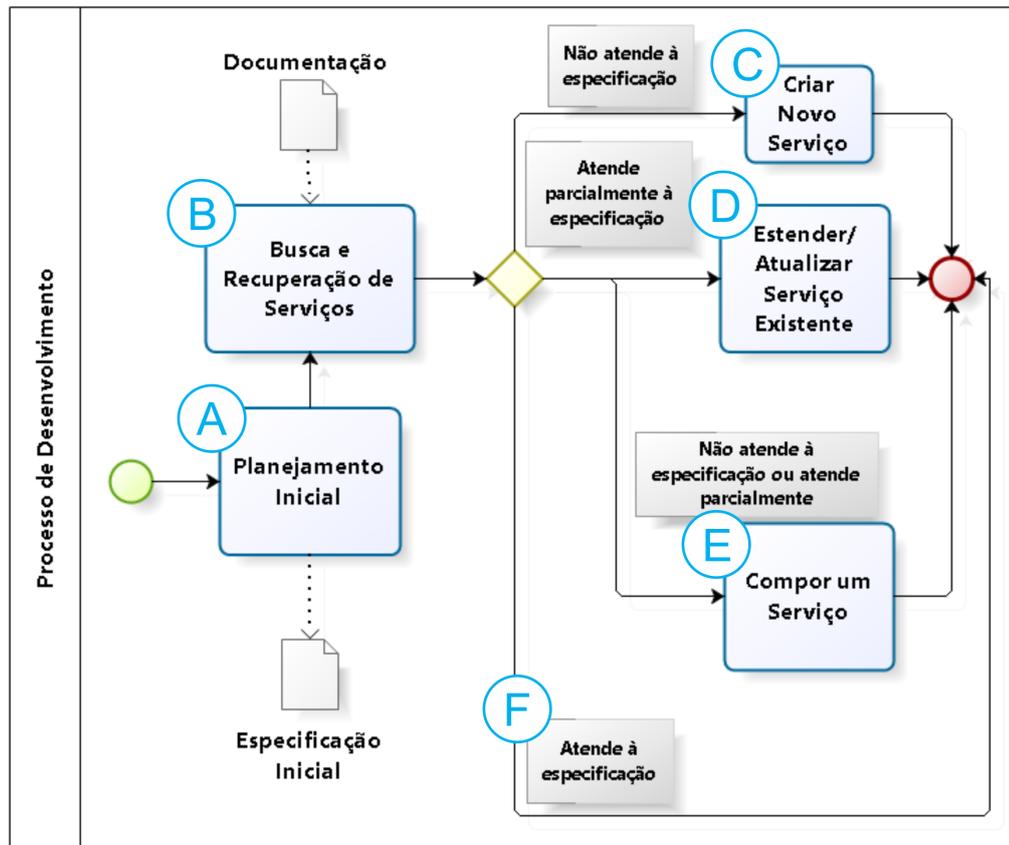


Figura 3.1: Processo de Desenvolvimento de Serviços em um Ecosistema de Software Científico (estendido de (ARAKAKI et al., 2016))

um serviço que atende parcialmente à especificação e, com isso, o serviço será desenvolvido a partir dos artefatos recuperados, de modo a evoluí-lo (Figura 3.1-D). No terceiro caso, o serviço não atende à especificação ou são encontrados um ou mais serviços que atendem parcialmente à especificação. Com isso, pode-se realizar a composição de serviços entre eles visando sua reutilização (Figura 3.1-E). O terceiro caso diferencia-se do primeiro porque permite reutilizar serviços pré-existentes. Além disso, diferencia-se do segundo porque a extensão e atualização de um serviço construído por desenvolvedores externos à plataforma geralmente não é possível, o que gera a necessidade de uma composição. A composição permite que serviços externos e internos possam ser associados, apoiando a extensibilidade na plataforma. A extensibilidade deve ser considerada por contribuir para o sucesso do ecossistema. Um quarto caso ocorre quando o serviço encontrado atende totalmente à especificação, desta forma ele poderá ser utilizado em um experimento (Figura 3.1-F).

Conforme apresentado por Arakaki et al. (2016), a atividade de criação contém as tarefas que devem ser executadas para o desenvolvimento de um novo serviço. Essas tarefas contemplam os aspectos específicos da plataforma E-SECO, como, por exemplo,

a modelagem do serviço utilizando um *modelo de features* e o desenvolvimento do serviço utilizando conceitos de LPS e SOA.

A atividade de composição é realizada quando não se encontra um serviço adequado para um experimento. Ela pode ser realizada também quando se encontra um ou mais serviços que atendem parcialmente à especificação inicial. A composição se associa à capacidade de automatizar um processo de negócios. Tem como objetivo transformar grandes problemas em problemas menores. Assim, soluções lógicas são decompostas, e recompostas, em outra configuração com o objetivo de resolver um problema específico (ERL, 2008).

A próxima seção trata da evolução da atividade de busca e recuperação de serviços, que é uma tarefa relevante para permitir a composição de serviços em um ECOSC.

3.2 ATIVIDADE DE BUSCA E RECUPERAÇÃO DE SERVIÇOS

Visando buscar e recuperar serviços que possam ser reutilizados em um ECOS, apresentamos um mecanismo para realizar esta atividade. Para isso, são considerados tanto os serviços encontrados em instâncias da plataforma E-SECO, como em repositórios externos.

Como ilustrado na Figura 3.2, a busca se inicia com a inserção dos dados da atividade de planejamento inicial. Neste momento, a busca é dividida em duas etapas que funcionam em paralelo, uma interna, que atua nos repositórios de serviços das instâncias da plataforma E-SECO e outra que atua nos repositórios externos.

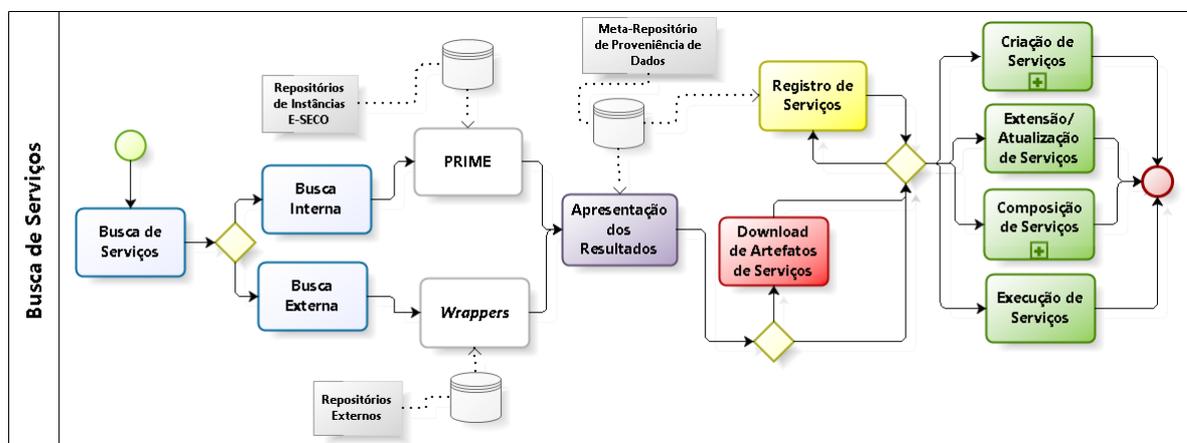


Figura 3.2: Etapa de Busca e Recuperação de Serviços na Plataforma E-SECO

A etapa de busca interna propõe a utilização da abordagem PRIME (**PR**agmatic

Interoperability to MEaningful collaboration) (NEIVA et al., 2015) desenvolvida para apoiar a interoperabilidade. PRIME é uma abordagem que, ao receber um conjunto de dados, ela os processa e realiza a busca de serviços nas diversas instâncias da E-SECO. Posteriormente, PRIME retorna todos os serviços encontrados ordenados do mais indicado ao menos indicado, segundo os parâmetros da busca. Estes parâmetros capturam características sintáticas, semânticas e pragmáticas dos serviços.

Cabe ressaltar que esta abordagem considera apenas os serviços registrados na plataforma E-SECO. Neste contexto, a busca em repositórios externos considera a utilização de APIs disponíveis. Visando a padronização da apresentação dos dados, utiliza-se *wrappers* para converter os resultados específicos de cada repositório em um formato padrão. No presente trabalho, os *wrappers* estão localizados no cliente dos repositórios. Tendo como objetivo padronizar os dados obtidos através de APIs. No contexto de um ECOSC, esta visão padronizada dos dados pode facilitar a reutilização de serviços em experimentos científicos. Desta forma, é possível visualizar serviços disponíveis em diferentes repositórios.

Assim como apresentado por Santos et al. (2013) no contexto de ECOS um componente é classificado como qualquer artefato de software produzido durante o processo de desenvolvimento. Embora o foco deste trabalho esteja na apresentação de um mecanismo para buscar e recuperar serviços, este mecanismo permite também a recuperação de *workflows* científicos e outros artefatos relacionados aos serviços. Desta forma, a busca pode retornar diferentes componentes, a partir de repositórios internos e externos. Ao final das etapas de busca os resultados são listados à equipe de desenvolvimento. A partir destes resultados é possível realizar o *download* dos componentes relacionados a determinado serviço, bem como analisar as informações disponíveis sobre o mesmo.

Além disso, é possível analisar dados de proveniência relacionados a serviços e *workflows* utilizados em experimentos anteriores. Estes dados são obtidos a partir do meta-repositório de proveniência existente na plataforma. A partir deles, cientistas podem analisar os parâmetros necessários para a execução de um *workflow*, bem como o histórico de versões para cada *workflow*. Assim, é possível, por exemplo, verificar a necessidade de migrar um experimento para uma nova versão de *workflow*. Informações sobre falhas em tarefas também podem ser obtidas, considerando um histórico de execução, com dados capturados pelo módulo de proveniência. De posse de todas as informações, um cientista pode decidir se algum serviço encontrado pode ser utilizado em seus experimentos. Caso

um serviço atenda a necessidade, ele pode ser utilizado e adicionado ao repositório da plataforma com informações de proveniência associadas. Como exemplo, podem ser adicionados dados relacionados a qual *workflow* o serviço está associado, detalhes sobre a execução correta do serviço, a versão utilizada e detalhes relacionados ao experimento. Isto permitirá que, posteriormente, outros cientistas possam reutilizar o serviço. Em novos experimentos, os cientistas poderão verificar estes dados para verificar se o serviço é adequado ao experimento.

A Figura 3.3 apresenta a arquitetura da plataforma E-SECO com a adição do módulo de composição de serviços (MCS). A Arquitetura inclui o Ambiente de Desenvolvimento E-SECO, o Núcleo da Plataforma, a Rede P2P (Rede ponto-a-ponto), o Módulo de Proveniência de Dados e as Camadas de Integração e Visualização. O Ambiente de Desenvolvimento permite que desenvolvedores internos e externos desenvolvam novas funcionalidades para a plataforma. As funcionalidades necessárias e relatos de problemas são solicitados por pesquisadores (usuários finais da plataforma). O Núcleo da Plataforma contém o repositório de serviços da plataforma e o Repositório da LPS, através dos quais cada serviço é associado a um Modelo de *Features* e a uma Ontologia de Domínio. A Rede P2P permite que cada instância que utiliza a plataforma funcione tanto como cliente quanto como servidor, permitindo assim, compartilhar serviços e dados. Os serviços armazenados no Núcleo da Plataforma são ranqueados pelo Módulo de Interoperabilidade. Esse módulo possui uma ontologia para descrição de serviços obtidos de diferentes instâncias da plataforma, disponível em um arquivo localizado na Rede P2P. O módulo de Proveniência permite obter dados a partir da execução de um *workflow*. As aplicações externas se associam às diferentes fontes de dados integradas à plataforma, que estão associadas ao ciclo de vida de um experimento científico. A camada de integração possui clientes para as APIs que podem ser utilizados e estendidos por desenvolvedores. A camada de visualização, por sua vez, fornece um ambiente através de uma interface web para a condução de um experimento científico.

O presente trabalho busca apresentar o MCS (destacado em cor vermelha) na Figura 3.3. Este módulo compreende o Gerente de Dados de Repositórios Externos (GDRE), o Gerente de Dados de Repositórios Internos (GDRI), o Gerente de Serviços (GS) e o Gerente de Composição de Serviços (GCS). O GDRE é responsável por obter dados de aplicações externas, incluindo dados de proveniência fornecidos por outras instâncias do

E-SECO, além dos repositórios externos de serviços e *workflows*. Os dados obtidos são então repassados aos *wrappers* que são responsáveis por converterem os dados em um formato específico. O GDRI é responsável por obter serviços a partir de cada instância da plataforma E-SECO utilizando a abordagem PRIME. A abordagem PRIME considera que diferentes atividades podem exigir o suporte para diferentes tipos de interoperabilidade. Por exemplo, um cientista que deseja utilizar um serviço que seja o mais indicado para compor outro serviço, visando a prototipação de um *workflow* científico, pode optar pelo suporte à interoperabilidade pragmática, porém, uma busca sintática pode ser realizada quando ele requer apenas informações relacionadas ao formato sintático dos dados para interoperar os serviços.

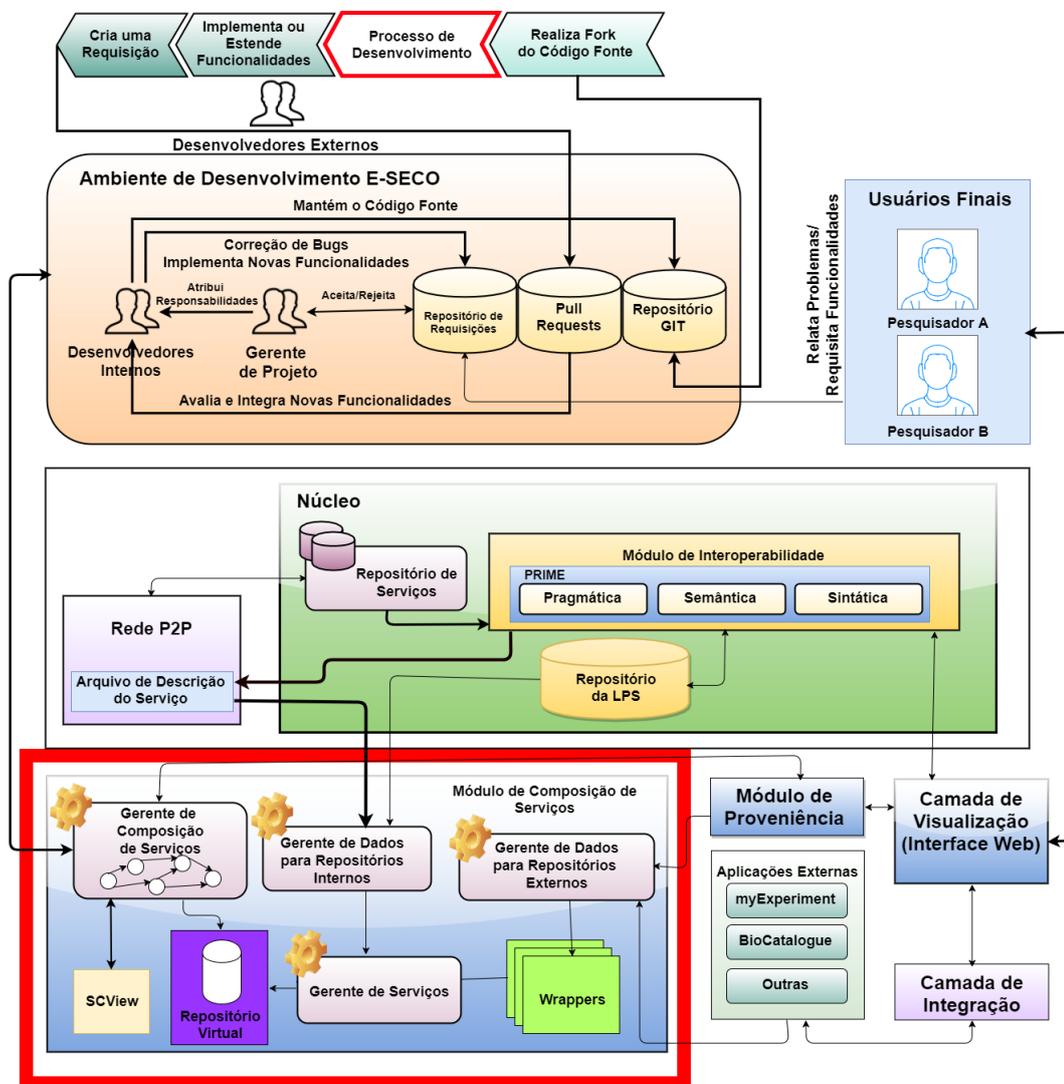


Figura 3.3: Arquitetura Simplificada da Plataforma E-SECO com a adição do módulo de composição de serviços (estendido de (SIRQUEIRA et al., 2016))

Posteriormente, o Repositório Virtual é responsável por representar os serviços encontrados em um formato padrão, definido por uma ontologia que permite descrever os serviços. Isso torna possível a composição de novos serviços e a reutilização de serviços existentes. Utilizou-se o termo “Repositório Virtual”, considerando que a literatura pode fazer uma distinção entre os termos “Registro de Serviços” e “Repositório de Serviços”. Na literatura o termo “Repositório” é geralmente associado à persistência de dados em bases de dados, enquanto o termo “Registro” associa-se mais a utilização em tempo de execução (ERL, 2008). No caso deste trabalho, optou-se por utilizar o termo “Repositório Virtual” para indicar que estes serviços não são persistidos em uma base de dados quando anotados semanticamente. Após realizar uma composição e avaliá-la, o desenvolvedor pode adicionar o serviço à LPS. A inclusão de serviços à LPS considera o processo de construção de serviços definido na plataforma, onde o serviço é associado a um Modelo de *Features* e a uma Ontologia de Domínio (ARAKAKI et al., 2016).

O meta-repositório de proveniência da E-SECO inclui dados de proveniência prospectiva (especificações de uma tarefa ou serviço computacional) e retrospectiva (os passos que foram executados para gerar um produto de dado). O suporte para ambos os tipos de proveniência permite uma análise mais abrangente de resultados científicos (LIM et al., 2010). No caso da plataforma E-SECO, é possível, por exemplo, analisar onde o *workflow* foi executado e utilizar *property chains* para analisar sua evolução e influência sobre outros *workflows*, o que auxilia a manutenção e evolução de *workflows*. Desta forma, o cientista pode analisar o contexto de utilização de um serviço em experimentos realizados anteriormente. Como resultado, é possível verificar se o *workflow* ou serviço buscado pode atender suas necessidades.

Além do repositório de proveniência existente na E-SECO, a ontologia PROV-OEXT (SIRQUEIRA et al., 2016), alinhada a este meta-repositório, permite a geração de inferências e a extração de informação dos serviços capturados pela E-SECO, armazenados no repositório. Com essas informações, a plataforma E-SECO pode sugerir ao pesquisador mudanças estratégicas de reutilização ou novas composições de serviços, utilizando como base os dados históricos.

O GS permite que os serviços encontrados sejam anotados semanticamente, utilizando uma ontologia de domínio. A partir de um serviço encontrado nos repositórios, o GS é capaz de gerar um arquivo em formato OWL-S para cada operação relacionada a um serviço,

a partir de sua descrição, seus parâmetros de entrada e saída e os tipos de dados utilizados como parâmetros. Jiménez and Wilkinson (2014) utilizaram uma descrição textual para serviços que inclui (quando possível) o nome do serviço, a descrição do serviço, o nome da operação e a descrição da operação. No caso do presente trabalho, onde se considera a necessidade de desenvolvimento de serviços para apoiar experimentos científicos, optou-se também por incluir, quando possível, dados relacionados à proveniência dos serviços.

O GCS auxilia o usuário durante o processo de composição de serviços. A partir dos serviços registrados no repositório virtual é possível analisar quais são relevantes para a composição. Para isso, um algoritmo que permite obter serviços relevantes, a partir de um conjunto de entradas e saídas de serviços e palavras-chaves, é utilizado. Além disso, nesta etapa também é possível buscar serviços utilizando a abordagem PRIME. Assim, é possível obter serviços que são interoperáveis a nível sintático, semântico e pragmático. A busca por serviços relevantes permite analisar como os serviços web serão integrados, e como diferentes serviços podem ser associados para resolver determinado problema. A composição de serviços na E-SECO visa garantir a manutenção de aspectos de qualidade, como “reusabilidade”, “capacidade de composição” e “interoperabilidade” na plataforma. Isto é importante porque a área de *e-Science* utiliza recursos compartilhados entre cientistas que podem estar geograficamente distribuídos. Para auxiliar a análise de interoperabilidade na plataforma, o serviço SCView (MARQUES et al., 2017) pode ser utilizado pelo GCS. Desta forma, o desenvolvedor pode analisar visualmente se serviços são interoperáveis ou se podem ser recompostos em novos serviços.

A reusabilidade de serviços é um princípio básico para se alcançar muitos objetivos estratégicos associados a SOA (ERL, 2008). Neste sentido, consideramos que a plataforma E-SECO permite a obtenção de uma maior quantidade de serviços, por estar integrada a diferentes repositórios. Com a possibilidade de analisar características específicas de um serviço, torna-se possível obter serviços mais neutros ou agnósticos para serem registrados. Serviços agnósticos possuem uma lógica independente do processo de negócio, tecnologia ou plataforma de aplicação e tendem a aumentar o reuso potencial (ERL, 2008). A utilização de dados de proveniência é capaz de auxiliar uma análise neste sentido. Estes dados permitem analisar o contexto de utilização de um serviço em experimentos anteriores, o que possibilita, por exemplo, perceber se a lógica de um serviço independe de seu processo de negócios.

A capacidade de composição de serviços se relaciona à reusabilidade. Serviços com um potencial de reuso aumentam as chances de que eles sejam repetidamente compostos (ERL, 2008). Busca-se realizar uma evolução do processo de desenvolvimento de serviços na plataforma E-SECO. A composição de serviços é realizada neste sentido, por permitir reutilizar serviços existentes.

Para garantir a interoperabilidade de um sistema, a formalização da captura e a incorporação de conceitos semânticos se tornam muito importantes (CHEN et al., 2017). Visando garantir a interoperabilidade entre os serviços na plataforma, é realizada uma descoberta de serviços que considera não apenas aspectos semânticos, mas também aspectos sintáticos e pragmáticos.

Em uma arquitetura orientada a serviços, é importante garantir que requisitos não funcionais tais como interoperabilidade entre serviços e reusabilidade sejam considerados. Para garantir que esses requisitos sejam mantidos na plataforma, uma evolução do processo de desenvolvimento de serviços pode ser necessária. A interoperabilidade permite apoiar a capacidade de composição de serviços. Quando o processo de desenvolvimento passa a permitir a composição de novos serviços e apoiar a reutilização neste contexto, a dimensão técnica da plataforma também é apoiada.

3.3 COMPOSIÇÃO DE SERVIÇOS

O processo de composição de serviços na plataforma é ilustrado na Figura 3.4. O processo utiliza notação BPMN e apresenta todas as etapas necessárias para auxiliar o desenvolvedor na criação de serviços compostos. Inicialmente, os serviços buscados são anotados semanticamente no Repositório Virtual (Registro de Serviços com anotação semântica no formato OWL-S). O desenvolvedor realiza uma requisição para a busca dos serviços mais relevantes. Esta requisição, quando informada, é comparada aos parâmetros da interface de um conjunto de serviços registrados. Tais serviços são obtidos a partir do repositório virtual através do “alinhamento semântico” entre os termos (*Service Matching*). Caso o desenvolvedor considere importante obter serviços que sejam interoperáveis, em diferentes níveis, pode optar por utilizar a abordagem PRIME, descrita anteriormente na seção 3.2. Esta comparação entre requisição e parâmetros relacionados aos serviços disponíveis é realizada na etapa de descoberta de serviços.

A descoberta de serviços permite identificar os mais relevantes serviços para uma

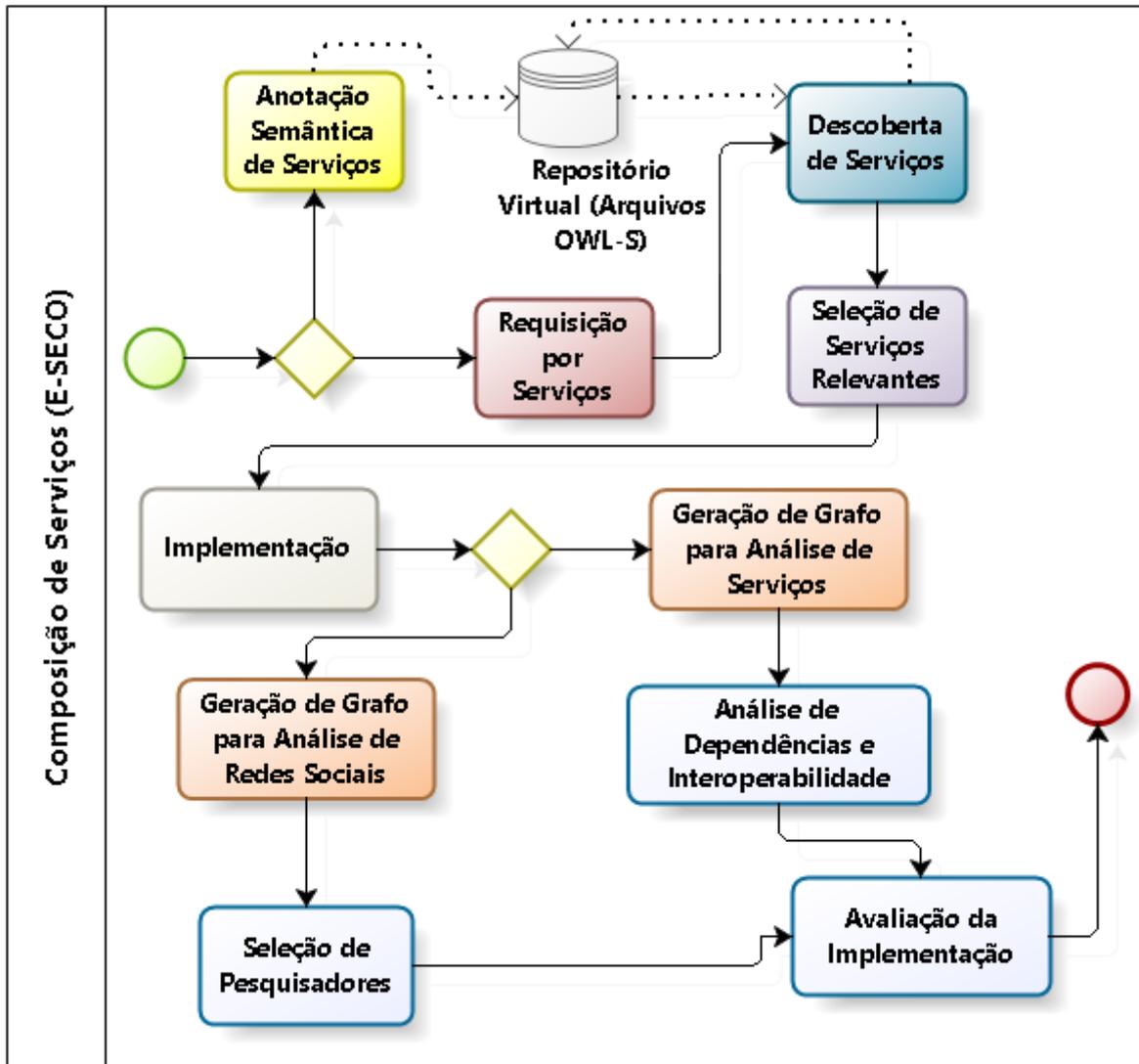


Figura 3.4: Processo de Composição de Serviços na Plataforma E-SECO

requisição. É realizada através da comparação entre a requisição e os parâmetros de relacionados à interface dos serviços disponíveis no repositório (CHEN et al., 2017). Ao realizar uma requisição, o cientista utiliza uma interface web onde informa dados relacionados à interface do serviço (parâmetros associados a entradas e saídas do serviço) e palavras-chave relacionadas. Na web semântica, a base de conhecimento *WordNet* (MILLER, 1995) é comumente usada para a descoberta de serviços web, quando se considera aspectos semânticos (CHEN et al., 2017). Esta base associa substantivos, verbos, adjetivos e advérbios a sinônimos e hiperônimos através de relações semânticas. Hiperônimo é apresentado como uma relação semântica entre duas palavras, porém estando o hiperônimo em um nível hierárquico superior a outra palavra (hipônimo) (MILLER et al., 1990) (HEARST, 1992). Métodos de extração de hiperônimos permitem construir hierarquias

entre conceitos (LAFOURCADE, 2002). Hiperônimos são frequentemente referenciados na Engenharia de Software como “especialização” (LAFOURCADE, 2002).

Quando se considera apenas parâmetros entre serviços, a relevância é avaliada por um algoritmo que analisa a combinação entre entradas e saídas semânticas. O ranqueamento dos serviços considera os “graus de alinhamento semântico” (*Service Matching*) entre a interface do serviço e palavras chaves informadas na requisição aos serviços disponíveis. Estes graus são apresentados na literatura, em ordem de relevância, como “exact”, “plugin”, “subsume” e “fail” (PAOLUCCI et al., 2002). O grau “exact” indica que a saída de um serviço é um conceito equivalente a entrada de outro serviço. O grau “plugin” indica que a saída de um serviço é um sub-conceito (hipônimo) em relação a entrada de outro serviço. O grau “subsume” indica que a saída de um serviço é um super-conceito (hiperônimo) em relação a entrada de outro serviço. Quando não existe alinhamento “exact”, “plugin” ou “subsume”, os conceitos são incompatíveis. Para conceitos incompatíveis é atribuído o grau de alinhamento “fail”.

Dentre estes graus de alinhamento (*matching*), combinações do tipo “exact” e “plugin” são vistas como os únicos graus expressivos para analisar a relevância de um serviço (MIER et al., 2015). Para verificar se a interface de um serviço é equivalente à requisição, ou seja, possuem grau de compatibilidade “exact” verifica-se, para cada saída da requisição, a compatibilidade com a entrada de algum dos serviços registrados. Para verificar se esse grau é do tipo “plugin”, verifica-se, para cada saída da requisição, se é um subconceito de algum serviço disponível. Serviços cujas entradas e saídas sejam semanticamente equivalentes à requisição (sinônimos), também são classificados como relevantes. Para analisar a relação entre conceitos e subconceitos, foram utilizados hiperônimos entre os termos. Sinônimos e hiperônimos relacionados a um conceito foram obtidos da *WordNet* a partir da biblioteca MIT JWI (FINLAYSON, 2014).

A abordagem PRIME, também pode ser utilizada para a seleção de serviços relevantes. A partir dela é possível realizar uma requisição que considera aspectos sintáticos, semânticos e pragmáticos para buscar serviços relevantes. Os dados cadastrados apresentados são registrados e associados a um serviço específico, permitindo que a descoberta de serviços considere também aspectos pragmáticos como, por exemplo o tipo de licença, a configuração mínima de hardware necessária para execução ou qual grupo de pesquisa desenvolveu o serviço. Desta forma, há uma possibilidade de identificar a intenção do

desenvolvedor.

O desenvolvedor pode então selecionar os serviços ranqueados como mais relevantes para realizar a composição de serviços. Se considerar necessário, poderá registrar e utilizar *channels* (CAO et al., 2016) a partir do repositório virtual para converter entradas e saídas dos serviços científicos obtidos. No contexto deste trabalho, *channels* são conectores que permitem fornecer novas opções de conexão entre serviços. A partir deles é possível, por exemplo, converter valores ou modificar parâmetros. A partir da seleção de serviços e *channels*, o desenvolvedor poderá modelar o serviço a ser composto.

Após selecionar os serviços relevantes identificados o desenvolvedor realiza a implementação da composição. Para auxiliar a atividade de composição, este trabalho apresenta o serviço SCView. Este permite a visualização das composições realizadas e a identificação de pesquisadores que podem avaliar uma composição. Visualizar a composição permite aos *stakeholders* (desenvolvedores e cientistas) analisar as dependências e as relações de interoperabilidade entre os serviços. A partir da seleção de um serviço entre os ranqueados, desenvolvedores podem analisar visualmente quais serviços estão associados ao selecionado. Para esta representação, um grafo $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ é utilizado. Onde \mathbf{V} é um conjunto de vértices que representa todos os serviços associados à seleção. Cada serviço $s \in V$ é apresentado como um nó. \mathbf{E} , por sua vez, representa um conjunto com todas as arestas presentes no grafo, onde $E = D \cup I$. \mathbf{D} é um conjunto de arestas direcionadas que indicam dependências entre serviços. Cada aresta $\mathbf{d} \in \mathbf{D}$ indica que um serviço s_1 depende de um serviço s_2 . \mathbf{I} , por sua vez, representa um conjunto de arestas não direcionadas que indicam interoperabilidade entre serviços. Cada aresta $\mathbf{i} \in \mathbf{I}$ indica que dois serviços são capazes de interoperar. Este grafo possui arestas direcionadas (indicando dependências) e não direcionadas representando que eles interoperam.

A Figura 3.5 apresenta um grafo gerado pelo SCView. Como apresentado nesta figura, serviços compostos são apresentados por nós em cor vermelha. Serviços recuperados do repositório “BioCatalogue” são representados por nós na cor preta, enquanto serviços recuperados a partir de instâncias da plataforma E-SECO são representados por nós na cor verde. Arestas que representam a interoperabilidade são ilustradas por linhas em cor preta e tracejadas. As dependências por sua vez são ilustradas por linhas contínuas em cor vermelha.

Este grafo foi gerado para o serviço “schema”, selecionado pelo botão “Dependencies”

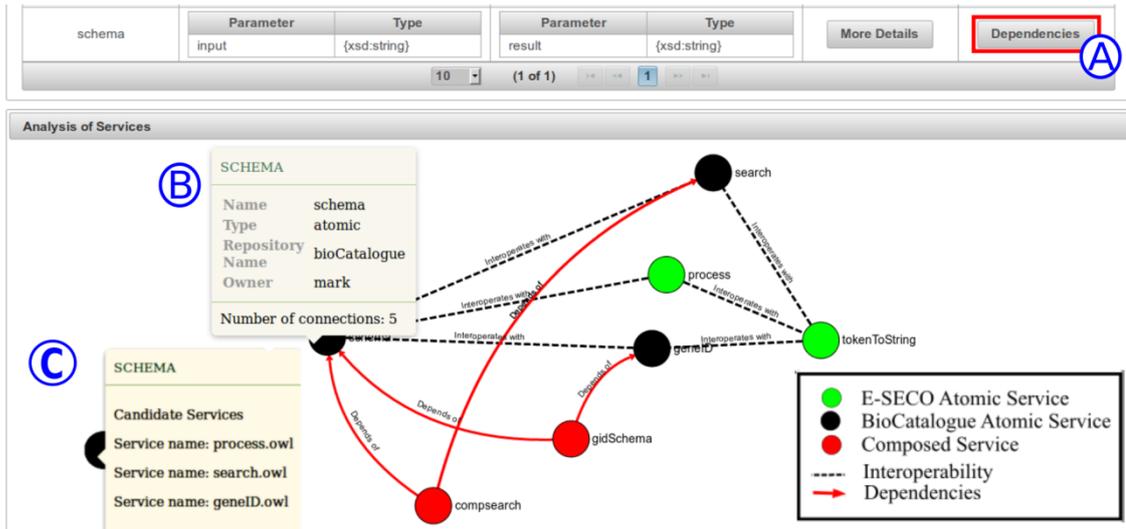


Figura 3.5: Grafo gerado pelo SCView com relações de interoperabilidade e dependências entre serviços

(Figura 3.5-A). Dados sobre o serviço podem ser analisados (Figura 3.5-B). Caso o serviço não atenda à requisição, o desenvolvedor pode selecionar outro serviço mais adequado. Para tanto, basta selecionar com o botão direito determinado nó (serviço) do grafo e, visualizar outros serviços candidatos à substituição (Figura 3.5-C). Quando esta ação é realizada, serviços compatíveis semanticamente com as entradas e saídas do serviço atual são apresentadas. Assim, um novo serviço pode ser escolhido para a composição. O desenvolvedor pode então selecionar este serviço no registro de serviços e gerar uma nova composição. Isto é realizado através de um sequenciamento de serviços disponíveis que pode ser realizado na plataforma. O serviço composto é, então, anotado semanticamente e registrado no repositório, substituindo a composição anterior. Neste momento, a solução considera não apenas as métricas de dependência funcional, mas a capacidade de interoperar com outros serviços (locais e geograficamente distribuídos). Além disso, é possível realizar uma nova busca adicionando parâmetros relacionados à interface do serviço e novas informações contextuais, visando obter serviços pragmaticamente interoperáveis. Caso o serviço atenda à requisição, o desenvolvedor pode utilizar o serviço e gerar uma nova composição a partir dos serviços registrados. Posteriormente, o serviço composto terá sua funcionalidade avaliada junto aos pesquisadores.

A capacidade de identificar pesquisadores capazes de avaliar a funcionalidade de um serviço também existe no SCVIEW. Para tanto, o desenvolvedor utiliza um filtro com opções para gerar redes de colaboração. A Figura 3.6 apresenta as opções presentes nesse

filtro. Assim, é possível selecionar o tipo de colaboração realizada entre os pesquisadores (Produção Bibliográfica, Orientações Concluídas, Projetos e Produção Técnica). Além disso, é possível realizar uma busca em três diferentes níveis (universidades, grupos e pesquisadores). O serviço permite também que dados sobre um pesquisador específico sejam encontrados, para isto, o serviço identifica e seleciona na rede o pesquisador selecionado e suas relações. Outra característica importante é a capacidade de realizar uma análise evolutiva, selecionando um ano específico ou em um intervalo entre anos para análise da rede. Desta forma, é possível identificar colaborações que ocorreram em um período específico. Assim, o usuário é capaz de identificar como as colaborações evoluíram ao longo do tempo. Isto permite, por exemplo, analisar o nível de colaboração de determinado grupo e quem são os pesquisadores que mais colaboraram em determinado período de tempo.

Figura 3.6: Filtro utilizado para identificar interações entre pesquisadores

3.3.1 ANÁLISE DE DEPENDÊNCIAS ENTRE SERVIÇOS

Para apoiar a dimensão técnica de um ECOS, utiliza-se o serviço SCView. Este serviço foi construído para visualizar composições de serviços, obtidos de diferentes instituições de pesquisas. Para apoiar a análise sobre as dependências do serviço, métricas de centralidade são utilizadas. Várias métricas de centralidades caracterizam a estrutura de uma rede complexa, tais como: *Closeness*, *Betweenness* e *Node Degree* (WASSERMAN; FAUST, 1994). Estas métricas podem definir os nós mais importantes da rede. A importância do nó é dada com base na semântica da rede.

Buscando caracterizar os serviços mais importantes em uma rede que inclui relações de

dependências e interoperabilidade entre serviços (arestas do grafo), utilizou-se as métricas de centralidade global *Closeness* e *Betweenness*. Métricas globais são assim apresentadas por não considerarem apenas os nós (no caso, serviços) vizinhos ao nó analisado, mas por considerarem a rede como um todo. Analisar a importância destes serviços permite avaliar o impacto causado na rede, caso ocorra alguma falha em um dos serviços. Esta métrica permite capturar o quão central é o vértice (serviço) em relação à rede de composição como um todo. Assim, são consideradas todas as arestas, tanto as que indicam interoperabilidade, quanto aquelas que indicam dependências associadas à composição de serviços, desconsiderando o sentido das arestas que indicam dependência.

A métrica de *Closeness* é apresentada na Figura 3.7-A. Para tanto, V representa o conjunto de serviços, “ n ” representa o número de serviços $d(x,y)$ é uma função que calcula a distância entre serviços “ x ” e “ y ”, onde “ x ” é o serviço para o qual a métrica de *Closeness* está sendo calculada. Assim, é possível analisar a importância de um serviço a partir de sua centralidade na rede. Utilizando esta métrica, é possível identificar a distância média entre um serviço e todos os outros serviços da rede.

A métrica de *Betweenness* é apresentada na Figura 3.7-B. Esta métrica considera todos os caminhos mínimos do grafo e o número de caminhos mínimos que passam pelo nó (serviço) analisado. Para tanto, “ s ” e “ t ” representam dois serviços pertencentes ao grafo. Os caminhos mínimos não incluem o serviço que está sendo analisado (serviço “ v ”). Assim, no contexto deste trabalho, a fórmula apresentada na Figura 3-B representa a razão entre o número de caminhos mais curtos entre “ s ” e “ t ” que passam por “ v ” e o número total de caminhos mais curtos entre “ s ” e “ t ”.

$$\begin{array}{cc}
 \textcircled{A} & \textcircled{B} \\
 C_x = \frac{\sum_{y \in V - \{x\}} d(x,y)}{n-1} & C_v = \sum_{s,t \in V; s,t \neq v} \frac{\sigma_v(s,t)}{\sigma(s,t)}
 \end{array}$$

Figura 3.7: Métricas de centralidade global (WASSERMAN; FAUST, 1994)

Para as métricas de *Closeness* e *Betweenness* utilizou-se a técnica de normalização min-max (HAN et al., 2011) buscando facilitar o entendimento do valor encontrado. A normalização permite que os valores apresentados estejam em um intervalo entre 0 e 1. A Figura 3.8 apresenta a técnica utilizada, onde os valores min_A e max_A são, respectivamente,

os valores mínimo e máximo para o atributo A, onde A corresponde ao valor de *Closeness* ou *Betweenness*). Neste caso, a normalização irá mapear um valor de centralidade “v” no intervalo entre new_min_A (igual a 0) e new_max_A (igual a 1). Um baixo valor para *Closeness* ou um alto valor para *Betweenness* permitem identificar serviços com alta centralidade na rede (serviços mais importantes).

$$v' = \frac{v - min_A}{max_A - min_A} (new_max_A - new_min_A) + new_min_A.$$

Figura 3.8: Normalização min-max (HAN et al., 2011)

Para caracterizar as dependências utilizou-se as métricas de *fan-in* e *fan-out*. Marin et al. (2004) definem *fan-in* de um método **m** como o número de métodos distintos que podem invocar **m** e *fan-out* de um método **m** como o número de métodos distintos que podem ser invocados por **m**. No caso deste trabalho, onde consideramos *web services*, *fan-in* se relaciona ao número de serviços que chamam o serviço analisado, enquanto *fan-out* representa o número de serviços atômicos chamados por um serviço composto. As métricas aqui apresentadas são propriedades dos nós, não sendo relacionadas a geração de arestas. A função das métricas é permitir identificar os serviços que podem ser mais importantes na rede. As arestas do grafo possuem relações independentes destas métricas (interoperabilidade e dependências funcionais). No entanto, as métricas de *fan-in* e *fan-out* permitem identificar o número de dependências de um serviço, sem necessidade de uma análise visual do grafo.

A Figura 3.9 apresenta a visualização dos valores encontrados para analisar as dependências de um serviço específico. Nessa figura é possível identificar dados do serviço como nome, tipo, repositório de origem e nome do desenvolvedor. No caso do serviço “GetGeneId”, apresentado na Figura 3.9, é possível identificar que ele possui dependência de 2 serviços (a partir da métrica de *fan-out*). Por ser um serviço composto, as métricas de *Closeness* e *Betweenness* permitem identificar a importância do serviço na rede. O valor calculado para *Closeness* é inferior a “0,5” enquanto o valor de *Betweenness* é superior a “0,5”. Isso pode indicar que este serviço é importante para a rede por interoperar com vários outros e ter dependências. Arestas de dependência mais grossas se associam à alta centralidade de serviços atômicos associados (utilizando a métrica de *Closeness*).

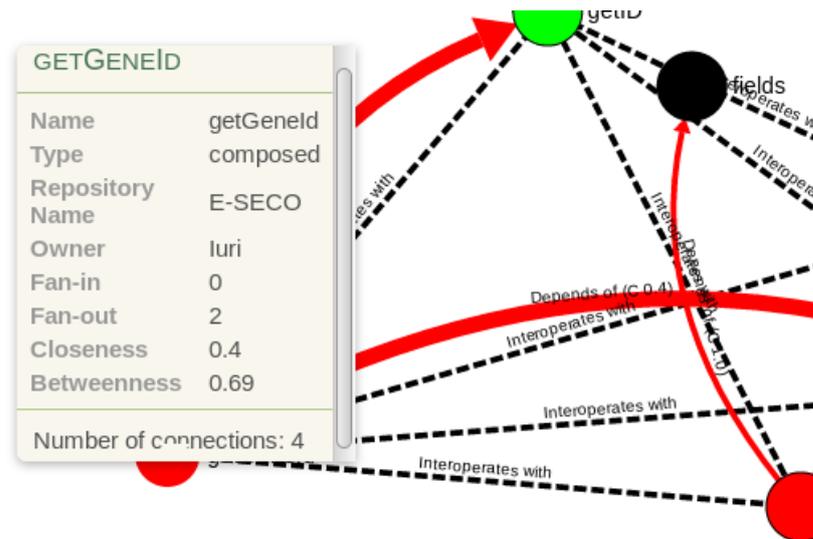


Figura 3.9: Análise de Métricas para um Serviço Específico

3.3.2 ANÁLISE DE REDES SOCIAIS CIENTÍFICAS

A visualização de redes sociais científicas é utilizada na plataforma E-SECO para auxiliar na identificação de potenciais colaboradores entre pesquisadores. A Figura 3.10 apresenta um exemplo de análise de uma rede científica gerada através do serviço SCView. Neste exemplo, são analisadas as colaborações de um pesquisador da Universidade Federal do Pernambuco (UFPE). As arestas em cor azul, identificadas na Figura 3.10, representam produções bibliográficas. As outras cores identificadas na figura: vermelho, laranja e verde, representam, respectivamente, produções técnicas, projetos e orientações concluídas. Para o nó selecionado é possível identificar, neste exemplo, duas potenciais relações de colaboração. Ambas, por produções bibliográficas. A rede gerada neste exemplo se relaciona a pesquisadores, porém, o serviço permite também analisar redes em termos de grupos identificados e universidades. Para gerar redes de colaboração em outros níveis (diferentes de pesquisador), o desenvolvedor deve selecionar como nível “Groups” ou “Universities”. Desta forma, poderá observar quais grupos ou universidades colaboraram entre si, com os tipos de relações entre eles. Embora este recurso esteja presente no serviço SCView, neste trabalho o foco está na identificação de colaboração a nível de pesquisadores (“Researchers”).

Ströele et al. (2017) caracterizam o grau de importância dos pesquisadores por seu potencial de colaboração. Dessa forma, duas métricas são utilizadas, uma para analisar a centralidade global dos pesquisadores e outra para avaliar a centralidade local. Embora

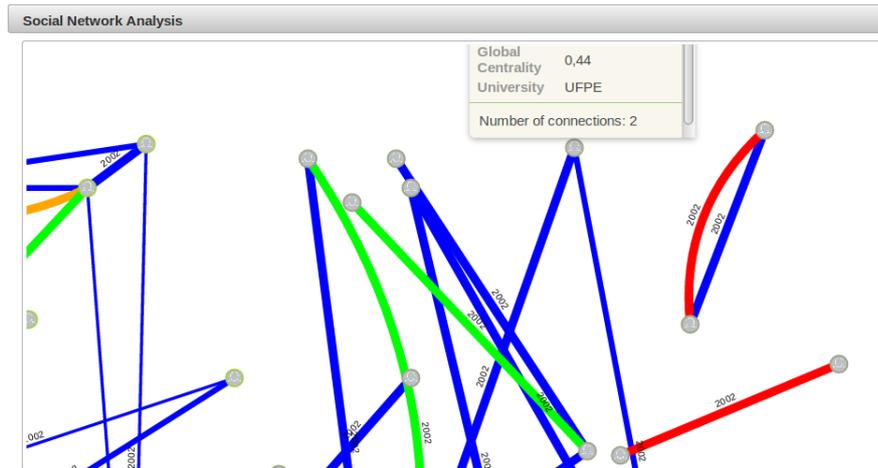


Figura 3.10: Identificação de Colaborações em uma Rede Social Científica

Ströele et al. (2017) apresentem uma ferramenta para análise de rede social científica, esta ferramenta não considera aspectos relacionados a uma plataforma de ECOSC. Desta forma, o serviço SCVIEW utiliza a análise de potenciais colaboradores (pesquisadores) para apoiar a composição de serviços em um ECOSC. Neste trabalho, a métrica *Closeness* é adotada como métrica global, analisando o potencial de colaboração dos pesquisadores, considerando a rede como um todo. A centralidade local, por sua vez, analisa as conexões dos vizinhos de um nó. Essa métrica tem como objetivo caracterizar se o nó considerado está significativamente conectado a seus vizinhos (se possui influência local). Neste sentido, a métrica *Node Degree* é utilizada para analisar a colaboração local dos pesquisadores, identificando seu potencial de colaboração com seus vizinhos. A métrica de *Node Degree* é apresentada na Figura 3.11.

$$LC_x = \frac{2 \times g(x)}{n(n-1)}$$

Figura 3.11: Métrica local para análise de colaboração entre pesquisadores (WASSERMAN; FAUST, 1994)

Na Figura 3.11, “ $g(x)$ ” representa o número de conexões do pesquisador, e “ n ” representa o total de pesquisadores na rede. A variável “ x ” representa o pesquisador para o qual a centralidade local está sendo calculada. Elementos com alta centralidade local representam pesquisadores com alta colaboração entre seus pares. Geralmente, esses pesquisadores representam um grupo de pesquisa em suas instituições. Por outro lado, os pesquisadores com alta centralidade global têm alta colaboração com a maioria dos pes-

quisadores de redes sociais, ou seja, são capazes de propagar suas ideias mais facilmente para vários pesquisadores da rede científica e, portanto, podem interagir com objetivos específicos.

Para identificar as comunidades de pesquisa científica foram também utilizadas estratégias de agrupamento. O algoritmo de agrupamento tem como objetivo identificar grupos de pessoas no grafo social que possuem um relacionamento forte entre eles. Para realizar esta identificação, o algoritmo segue a estratégia de analisar fluxos de conhecimento em rede social. Assim, pessoas com um grande fluxo de informações entre elas possuem tendência a pertencer a um mesmo grupo. O cálculo do fluxo máximo entre os elementos foi feito usando o algoritmo de Edmonds-Karp (EDMONDS; KARP, 1972). Neste trabalho estes grupos foram identificados previamente e armazenados em uma base de dados. Isto tornou possível a identificação, filtragem e análise de interações entre os grupos de pesquisa.

Para permitir que os potenciais colaboradores interajam através da plataforma, um serviço de colaboração foi desenvolvido. Esse serviço permite que o desenvolvedor identifique os pesquisadores relacionados ao processo de construção, e estabeleça uma comunicação através de uma troca de mensagens que ocorre de forma síncronas. Assim, a partir de um espaço de trabalho compartilhado é possível que o desenvolvedor atue como um coordenador para avaliar uma composição junto aos pesquisadores.

A avaliação das composições é realizada utilizando elementos de colaboração associados à percepção. No contexto deste trabalho, a percepção se associa à compreensão e execução correta das atividades. A coordenação das atividades é realizada pelo desenvolvedor. Assim, o desenvolvedor possui o papel de identificar pesquisadores através da análise de uma rede social (ARS). O suporte à comunicação é oferecido pela troca de mensagens, argumentação e negociação entre os participantes. Esta interação ocorre através do bate-papo. A cooperação é caracterizada pela atuação simultânea entre pesquisadores e o desenvolvedor no espaço de trabalho compartilhado. O conhecimento coletivo permite unir conhecimentos individuais para formar grupos com um objetivo específico (PIMENTEL; FUKS, 2012). Associar conhecimentos individuais torna o conhecimento coletivo mais rico, permitindo identificar relações que não são possíveis de ser apresentadas de forma isolada (PIMENTEL; FUKS, 2012). A utilização de mecanismos de percepção busca apoiar a comunicação, cooperação e coordenação na plataforma E-SECO. Tais mecanis-

mos podem prover a colaboração na plataforma. Assim, é possível utilizar o conhecimento coletivo obtido através da colaboração para apoiar a composição de serviços. Busca-se assim, utilizar recursos relacionados à dimensão social da plataforma para apoiar questões técnicas.

3.3.3 PROJETO DETALHADO

Para apresentar como o SCView está organizado na plataforma, a Figura 3.12 apresenta um diagrama de componentes que caracteriza como o serviço permite apoiar a composição. Através desse diagrama, é possível observar como o suporte à composição de serviços (“Service Composition Support”) utiliza serviços disponíveis na plataforma para análise de interoperabilidade entre serviços (PRIME) e para obter dados de proveniência (WSProVersion). Para auxiliar a análise de redes sociais e a visualização de serviços, são utilizadas métricas específicas, representadas pelo componente “Graph Metrics”. Dados relacionados a pesquisadores e proveniência são obtidos a partir do repositório da plataforma (E-SECO DB). O serviço SCView utiliza uma interface que contém métodos que permitem construir grafos para composição de serviços e análise de redes sociais científicas.

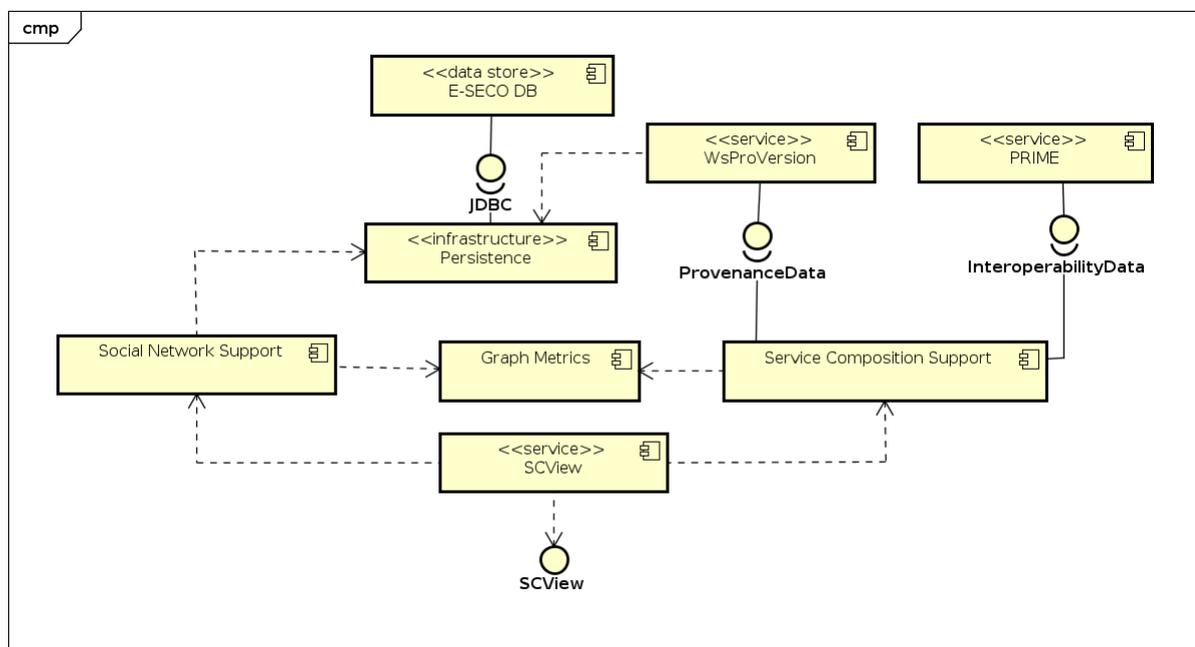


Figura 3.12: Diagrama de Componentes Relacionados ao SCView

Um diagrama de classes que contém as principais classes relacionadas ao serviço SC-VIEW é apresentada na Figura 3.13. Estas classes são utilizadas para permitir gerar a vi-

sualização na plataforma. Dentre as classes apresentadas no diagrama, a classe “GraphSN” é responsável por gerar o grafo com a visualização da análise de redes sociais. Para isto, ela utiliza duas classes. A primeira (“EdgeSN”) é utilizada para representar as arestas, enquanto a segunda (“NodeSN”) é utilizada para representar os nós. Os nós da rede podem representar pesquisadores, grupos de pesquisa ou universidades, de acordo com o nível de visualização selecionado.

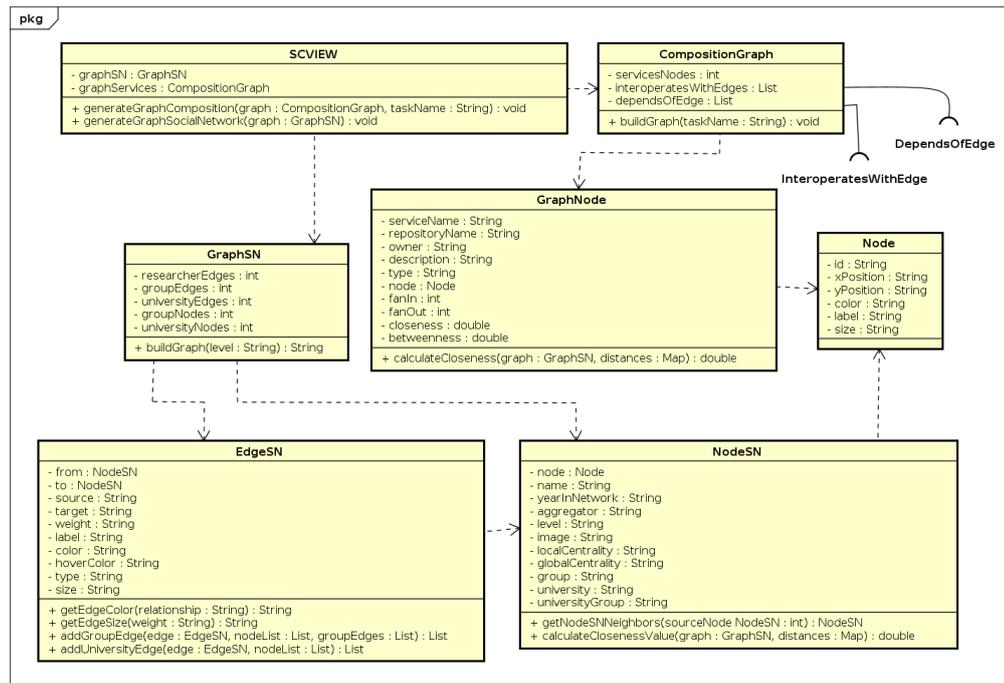


Figura 3.13: Diagrama de Classes (Serviço SCView)

Para analisar as composições de serviços, utiliza-se a classe “CompositionGraph”. Nesta classe os nós são serviços representados pela classe “GraphNode” e utilizam diferentes tipos de arestas. Estas arestas permitem identificar relações de dependências funcionais e de interoperabilidade (representadas pelas interfaces “DependsOfEdge” e “InteroperatesWithEdge”).

A representação da rede social no repositório da plataforma é apresentada na Figura 3.14, utilizando um Diagrama Entidade-Relacionamento (DER). Esta Figura apresenta as tabelas adicionadas para análise de redes sociais científicas. A entidade “node” é utilizada para armazenar diferentes nós, onde cada nó pode representar um pesquisador, um grupo de pesquisa ou uma universidade, de acordo com o nível de visualização selecionado. As arestas são representadas pela entidade “relationship_edge”.

Todas as tabelas adicionadas ao repositório, bem como todo o código desenvolvido, fo-

descrição dos serviços pode auxiliar um desenvolvedor na tomada de decisão em relação à reutilização de um serviço. Os serviços são então registrados no repositório virtual. Caso considere necessário, o desenvolvedor pode incluir dados sobre as características necessárias para utilização do serviço. Isto permite obter serviços que são pragmaticamente interoperáveis. A partir de um conjunto de serviços disponíveis é possível realizar a composição. Após registrar serviços compostos no repositório, o desenvolvedor é capaz de analisar como as composições foram realizadas. Para isto o serviço SCView permite visualizar graficamente dependências entre serviços, relacionadas às métricas entre eles. Também é possível analisar a relação de interoperabilidade entre os serviços. Por último, o desenvolvedor identifica potenciais colaborações entre pesquisadores. Desta forma é possível avaliar uma composição com os pesquisadores identificados.

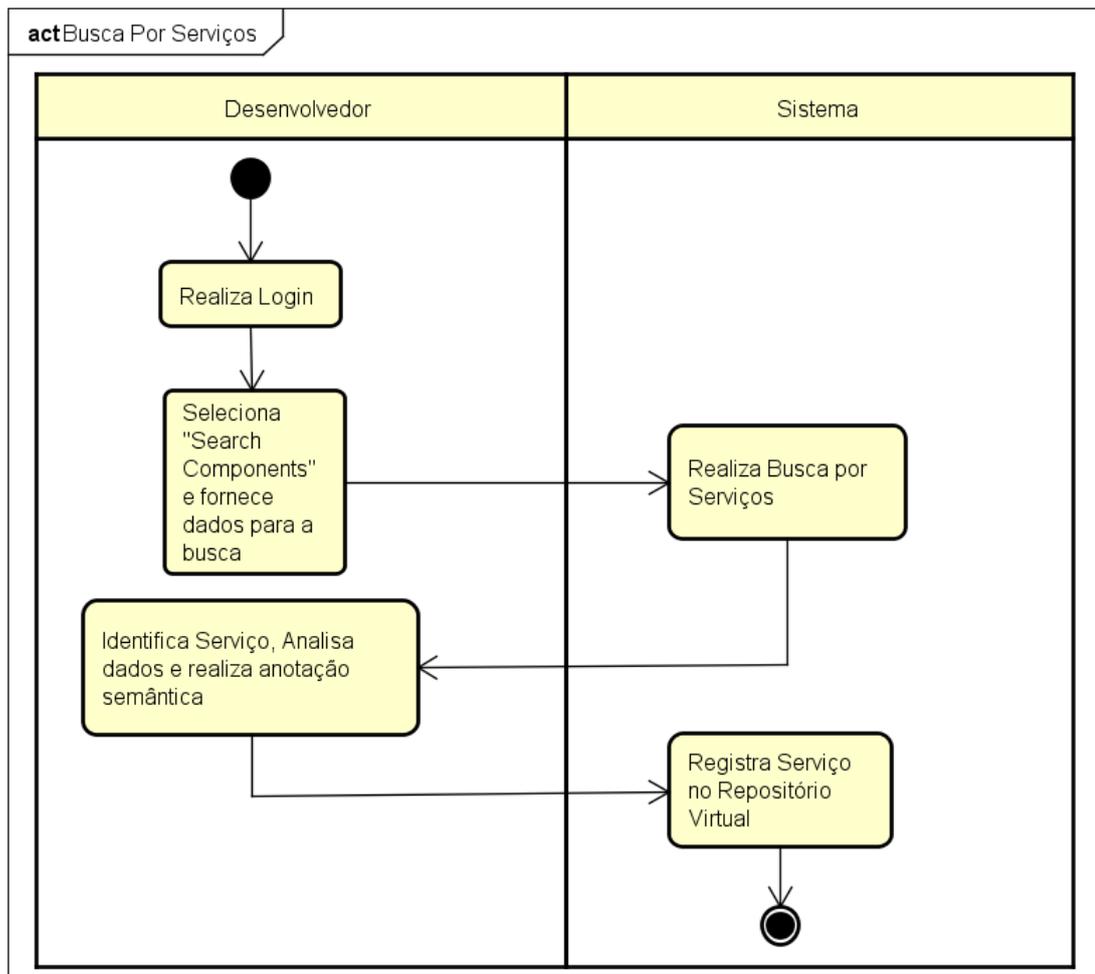


Figura 3.15: Diagrama de Atividades para a Busca por Serviços

As Figuras 3.15 e 3.16 apresentam diagramas de atividades para a busca e composição de serviços na plataforma. A Figura 3.15 destaca os papéis do desenvolvedor e do sistema

no processo de busca por serviços. Nessa figura é possível observar que o processo de busca por serviços no processo de desenvolvimento é realizado por um desenvolvedor, sem necessidade de interação com um pesquisador. Embora o pesquisador também possa realizar busca por serviços e analisar dados de proveniência para utilizar um serviço em um experimento, o presente trabalho está focado no processo de desenvolvimento de serviços na plataforma.

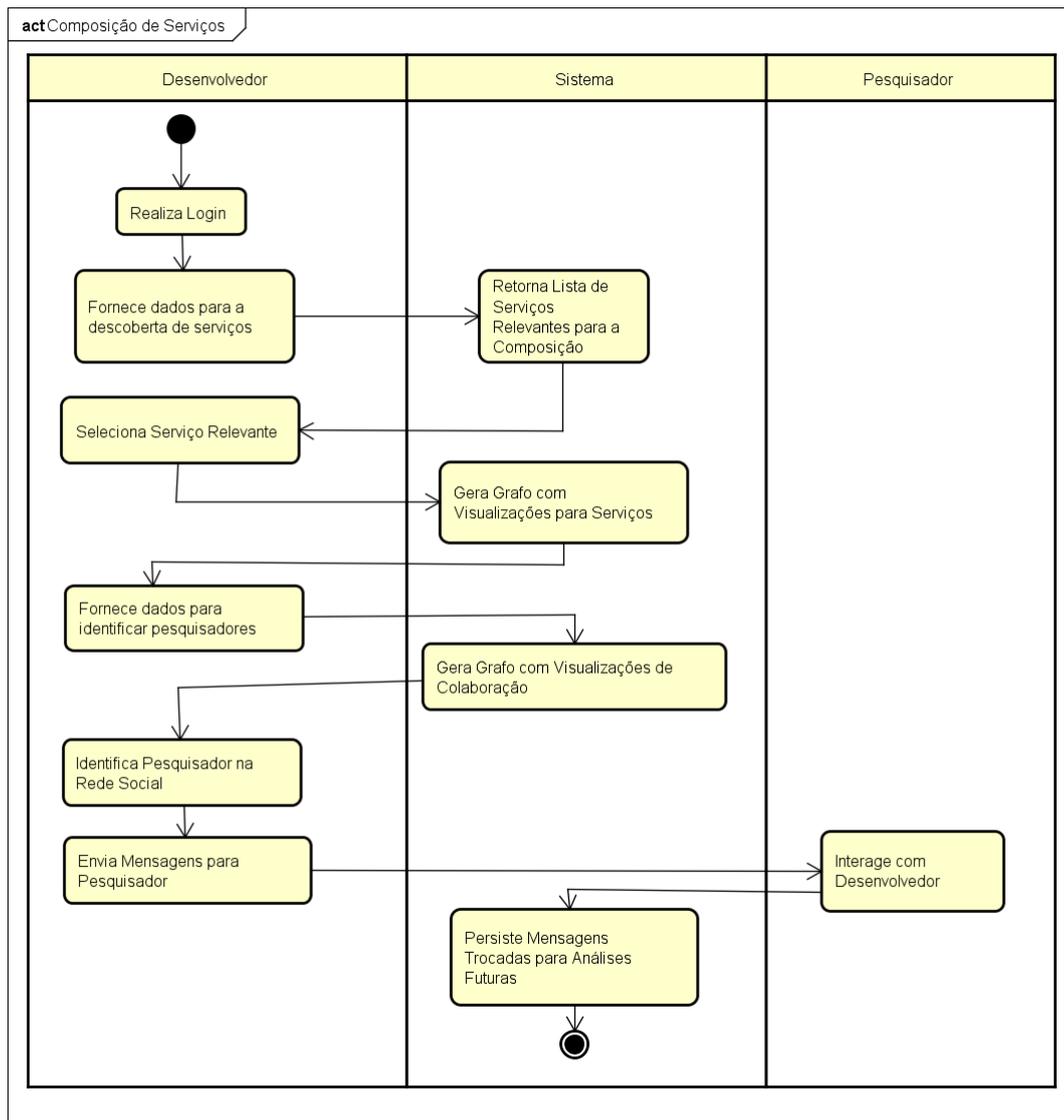


Figura 3.16: Diagrama de Atividades para a Busca por Serviços

A Figura 3.16, por sua vez, destaca o processo de composição de serviço, ressaltando os papéis do desenvolvedor, do sistema e do pesquisador. Nessa figura é possível observar que, ao contrário do processo de busca por serviços, a composição necessita do pesquisador para interagir com o desenvolvedor e avaliar uma composição realizada.

Cenário: um desenvolvedor deseja buscar um serviço relacionado à recuperação de dados em uma base de dados de proteínas. A partir dos resultados obtidos, o desenvolvedor verifica se o serviço atende às necessidades dos cientistas envolvidos e se o mesmo pode ser reutilizado ou estendido. O desenvolvedor analisa os metadados relacionados ao serviço e a existência de dados de proveniência para cada operação do serviço. Realiza a anotação semântica e utiliza os serviços anotados para compor um novo serviço. Por fim, utiliza o serviço SCView para avaliar as composições realizadas.

Utilização do mecanismo apresentado: o desenvolvedor inicialmente acessa a interface web da plataforma E-SECO. Na tela de busca, é possível informar um termo a ser buscado (Figura 3.17-A), e/ou utilizar uma busca sintática, semântica ou pragmática para localizar serviços armazenados em alguma instância da plataforma E-SECO (Figura 3.17-B). Para realizar a busca, o desenvolvedor informa os dados e clica sobre o botão “Search” (Figura 3.17-C). A busca então será realizada em instâncias da plataforma e em repositórios externos integrados. Os resultados obtidos serão apresentados em um formato padrão.

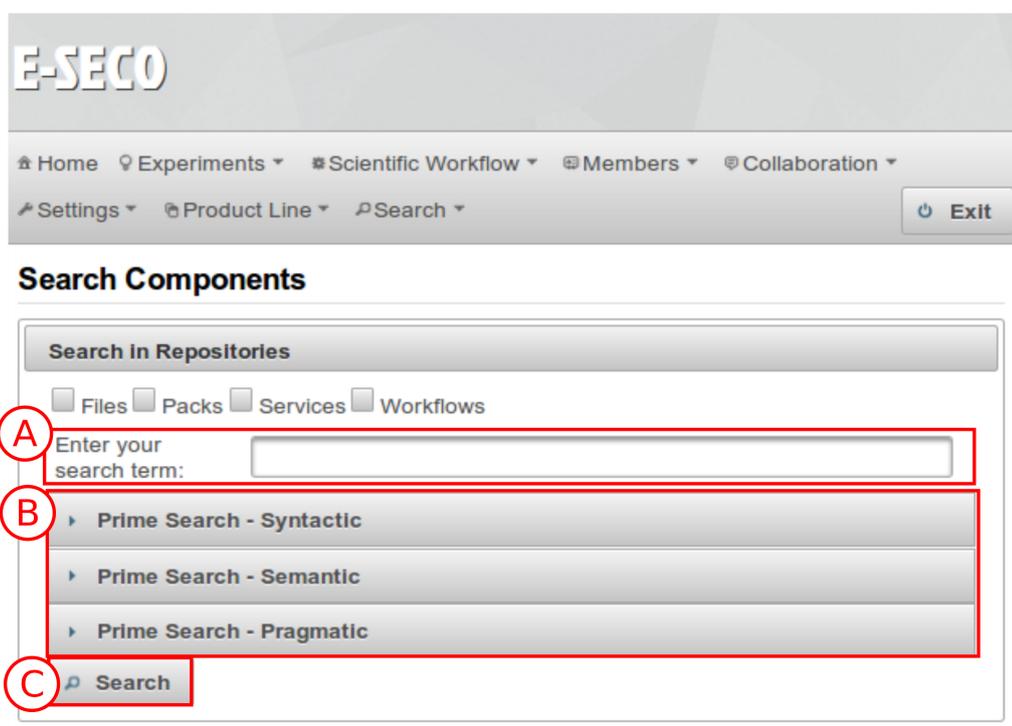


Figura 3.17: Tela de Busca por Serviços em Diferentes Repositórios

Neste exemplo, o desenvolvedor seleciona uma busca por serviços e informa o termo de busca como “Protein Database” (Figura 3.18). Além disto, deseja-se obter serviços cujo contexto de utilização seja público. Isto é obtido a partir da interoperabilidade

pragmática, utilizando a abordagem PRIME. Para isto, são buscados apenas serviços como tipo de licença “Public”. Após informar esses dados, o desenvolvedor clica em “Search” e inicia a busca.

The screenshot shows the E-SECO web application interface. At the top, there is a navigation menu with items: Home, Experiments, Scientific Workflow, Members, Collaboration, Settings, Product Line, Search, and an Exit button. Below the navigation is a section titled "Search Components". Inside this section, there is a sub-section "Search in Repositories" with radio buttons for "Files", "Packs", "Services" (which is selected), and "Workflows". Below this is a text input field labeled "Enter your search term:" containing the text "Protein Database". There are three expandable sections: "Prime Search - Syntactic", "Prime Search - Semantic", and "Prime Search - Pragmatic" (which is expanded). Under the "Prime Search - Pragmatic" section, there are several input fields: "Non Functional Requirement" (empty), "Context" (empty), "Artifact" (empty), "Domain" (empty), and "License" (containing the text "public").

Figura 3.18: Parâmetros Utilizados para a Busca em Repositórios na Plataforma E-SECO

Os serviços encontrados são então listados (Figura 3.19), permitindo ao usuário selecionar aqueles que, possivelmente, podem atender sua necessidade. Para analisar um serviço, o usuário pode selecionar o botão “More Details” para visualizar metadados ou selecionar o botão “Download” para obter a documentação do serviço selecionado.

Neste cenário, nenhum serviço capaz de buscar uma sequência de proteína em bases de dados de proteínas foi encontrado nas instâncias E-SECO. Apesar disso, os serviços obtidos de instâncias da plataforma foram apresentados por ordem de relevância em relação ao tipo de licença utilizado. O serviço “Nuclear Protein Database Query Service” foi encontrado na base de dados “BioCatalogue”. Por meio do mecanismo de busca, foi possível evitar o retrabalho com o desenvolvimento de um serviço já existente, o que pode agilizar o

Component Name	Repository Name	Component Type	Description	Owner	Component Documentation
Nuclear Protein Database Query Service	BioCatalogue	Service - SOAP	This service provides various ways to retrieve information from The Nuclear Protein Database, a d...	mark	More Details Download
Nuclear Protein Database Query Service	BioCatalogue	Service - REST	This service provides various ways to retrieve information from The Nuclear Protein Database, a d...	mark	More Details Download
jackhmmer	BioCatalogue	Service - SOAP	Iteratively search a protein sequence against a protein database	anthony.bretraudeau	More Details Download

Figura 3.19: Serviços Encontrados a Partir da Busca

desenvolvimento do experimento.

Neste exemplo, o desenvolvedor tem a opção de analisar um serviço detalhadamente, clicando na opção “More Details”. Esta ação possibilita analisar dados de proveniência e metadados associados ao serviço. Além disso é possível realizar a anotação semântica para cada operação disponível. Os metadados permitem analisar, dentre outras coisas, a data de criação do serviço, origem e *status* de utilização. A Figura 3.20 apresenta os metadados obtidos para o serviço selecionado a partir do repositório “BioCatalogue”, tais como: dados sobre o desenvolvedor e sobre o serviço selecionado, fornecidos pelo repositório de origem.

Provenance Data	
Metadata	
Service Semantic Annotation	
Owner	mark
Owner Country	United Kingdom
Owner City	
Component Created At:	2009-09-10T14:53:25Z
Component Updated At:	
Component Archived At:	---
Monitoring Status:	PASSED
Last Checked:	2017-01-27T08:38:52Z
License Type:	

Figura 3.20: Metadados obtidos para o Serviço Selecionado

Dados de proveniência sobre o serviço podem ser obtidos, o que permitirá ao cientista saber se o serviço já foi utilizado em algum experimento anteriormente. A partir do serviço selecionado, o sistema analisa todas as operações e realiza a busca por cada uma delas no meta-repositório de proveniência. Assim, é possível saber, por exemplo, quais *workflows* utilizaram determinada tarefa (operação de um serviço), se o *workflow* executou corretamente a tarefa, bem como em qual experimento o *workflow* foi utilizado. Exemplos de dados obtidos para a tarefa “geneID” são apresentados na Figura 3.21. “geneID” é

uma das operações do serviço “Nuclear Protein Database Query Service”. Este serviço foi executado anteriormente em um *workflow*, a partir do Sistema de Gerenciamento de *Workflow* Científico Kepler e associado a um experimento, a partir da plataforma E-SECO.

Component Name	Nuclear Protein Database Query Service
Used	<ul style="list-style-type: none"> ○ Task geneID was used in workflow GeneExtraction Task Id: 10 Workflow Id: 10
Was Informed By	<ul style="list-style-type: none"> ○ Task geneID was successful for activity Nuclear Protein Database Query Service Task Id: 10 Activity Id: 4
Was Associated With	<ul style="list-style-type: none"> ○ Workflow GeneExtraction was attributed to experiment Gene Extraction Experiment Experiment Id: 8 Workflow Id: 10
Acted Behalf Of	<ul style="list-style-type: none"> ○ Task geneID acted on behalf of task geneIDCon Input Port Id: 285 Output Port Id: 145 ○ Task geneID acted on behalf of task geneIDCon Input Port Id: 289 Output Port Id: 149
Was Ended By:	<ul style="list-style-type: none"> ○ Task geneID ended to activity Nuclear Protein Database Query Service Task Id: 10 Activity Id: 4

Figura 3.21: Dados de Proveniência obtidos para a Tarefa “geneID”

Após analisar os dados de proveniência é possível anotar semanticamente cada operação do serviço selecionado. Isto é realizado pelo Gerente de Serviços e permite que serviços atômicos sejam adicionados ao repositório virtual, potencializando sua reutilização. A Figura 3.22 apresenta as operações que podem ser anotadas para o serviço encontrado na busca. São apresentados todos os métodos do serviço encontrado, sendo que cada um dos métodos pode ser anotado como um serviço atômico. Ao clicar sobre o botão “Generate OWL-S”, o serviço é adicionado ao repositório virtual como um serviço atômico. O desenvolvedor pode adicionar dados que permitam, posteriormente, selecionar serviços relevantes baseando-se em aspectos sintáticos, semânticos e pragmáticos. Estes dados são armazenados em um arquivo XML associado a cada serviço e registrados no Repositório Virtual. Isto é realizado através da plataforma E-SECO. Estes dados são armazenados em um arquivo XML associado a cada serviço e registrados no Repositório Virtual.

A anotação semântica é importante para auxiliar o desenvolvedor na análise das funcionalidades de um serviço. A partir do arquivo de anotação semântica são gerados os grafos de composições, portanto, a anotação semântica é necessária para permitir a visualização

Name	Inputs		Outputs		Service Annotation
schema	WSDL Parameter	WSDL Type	WSDL Parameter	WSDL Type	Generate OWL-S
	input	xsd:string	result	xsd:string	
rest	WSDL Parameter	WSDL Type	WSDL Parameter	WSDL Type	Generate OWL-S
	gene	xsd:string	result	npd:queryBu	
	subsection	xsd:string			
search	WSDL Parameter	WSDL Type	WSDL Parameter	WSDL Type	Generate OWL-S
	input	xsd:string	result	npd:SQLque	
tables	WSDL Parameter	WSDL Type	WSDL Parameter	WSDL Type	Generate OWL-S
	input	xsd:string	result	npd:SQLque	
SQLquery	WSDL Parameter	WSDL Type	WSDL Parameter	WSDL Type	Generate OWL-S
	input	xsd:string	result	npd:SQLque	
geneID	WSDL Parameter	WSDL Type	WSDL Parameter	WSDL Type	Generate OWL-S
	input	xsd:string	result	npd:geneIDF	

Figura 3.22: Anotação Semântica Automática de Métodos do Serviço Selecionado

das composições. A próxima etapa consiste em buscar os serviços relevantes registrados na plataforma. A Figura 3.23 apresenta o formulário para busca de serviços relevantes entre os registrados na plataforma (anotados semanticamente). O desenvolvedor pode, então, preencher dados sobre parâmetros de entrada e saída dos serviços (separados por vírgula) para realizar a busca. Neste exemplo, consideramos que o desenvolvedor informa os parâmetros “input” (para entrada) e “result” (para saída). Utilizar sinônimos para a descoberta de serviços, nesta etapa, também retornaria os serviços corretamente, além de considerar os graus de alinhamento entre serviços, apresentados anteriormente. Além disso, o desenvolvedor poderia preencher dados para utilizar o serviço PRIME, através do qual é possível listar os serviços mais relevantes para a composição, considerando a interoperabilidade entre eles.

A partir da busca, são retornados serviços considerados relevantes para a composição, de acordo com os parâmetros informados. A Figura 3.24 apresenta os resultados retornados para a busca realizada.

As visualizações podem então ser realizadas. Baseando-se nos serviços encontrados, o desenvolvedor pode visualizar as relações entre os serviços e identificar as relações de

Relevant Services Search

Enter the Input Parameters Name (Separated by Commas)

Enter the Output Parameters Name (Separated by Commas)

Enter with keywords (Separated by Commas)

▼ Prime Search - Syntactic

Return:

▶ Prime Search - Semantic

▶ Prime Search - Pragmatic

Search

Figura 3.23: Busca Por Serviços Relevantes

Web Services Ranking

3 (1 of 3) 1 2 3

Task Name	Inputs		Outputs		Details	Dependencies Analysis
	Parameter	Type	Parameter	Type		
CompGenEid	input	{http://www.w3.org/2001/XMLSchema#string}	result	{http://www.w3.org/2001/XMLSchema#string}	More Details	Dependencies
getGenEid	input	{http://www.w3.org/2001/XMLSchema#string}	return	{http://www.w3.org/2001/XMLSchema#string}	More Details	Dependencies
queryBuilder	field	{http://www.w3.org/2001/XMLSchema#string}	result	{urn:npd#queryBuilderRe}	More Details	Dependencies
	select	{http://www.w3.org/2001/XMLSchema#string}				
	table	{http://www.w3.org/2001/XMLSchema#string}				
	value	{http://www.w3.org/2001/XMLSchema#string}				

3 (1 of 3) 1 2 3

Figura 3.24: Serviços Relevantes Encontrados

colaboração utilizando o serviço SCView, descrito anteriormente.

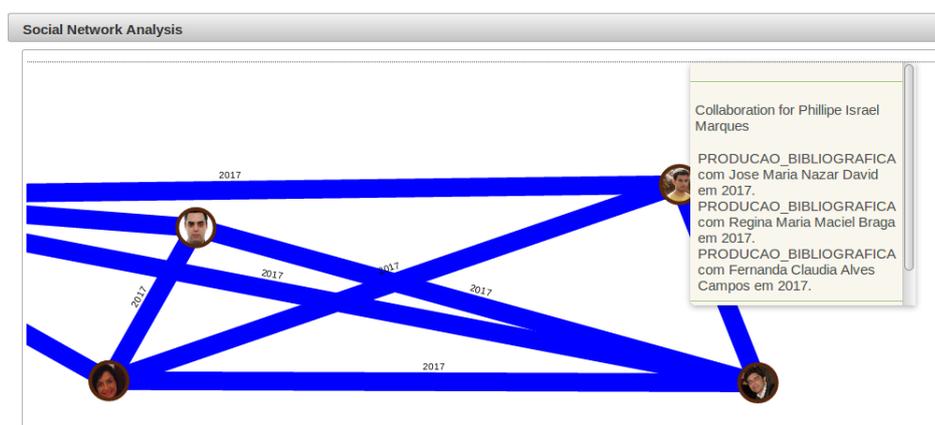


Figura 3.25: Visualização de Rede Social com Relações de Colaboração

A Figura 3.25 apresenta um exemplo de visualização da rede social, com relações de produção bibliográfica ocorridas no ano de 2017. Ao clicar sobre um pesquisador com o botão direito, o desenvolvedor tem a possibilidade de identificar todas as relações de

colaboração na rede. Assim, é possível selecionar quais pesquisadores podem avaliar a composição realizada. Após analisar os serviços e identificar possíveis pesquisadores que podem avaliar as composições, o desenvolvedor pode utilizar um serviço de colaboração disponível na plataforma para se comunicar de maneira síncrona com os pesquisadores.

3.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Este capítulo apresentou um mecanismo que tem como objetivo apoiar a composição de serviços em um ECOSC. Para isso, foram apresentados os processos de desenvolvimento e composição de serviços. O mecanismo utiliza recursos oferecidos para a plataforma que permitem analisar os serviços e colaborações entre pesquisadores.

Além disso, foi apresentado um exemplo de utilização da solução proposta no contexto de um ecossistema de software científico, ressaltando a forma pela qual a composição é apoiada considerando, sobretudo, os elementos de colaboração. No próximo capítulo é apresentada uma avaliação para a solução proposta.

4 AVALIAÇÃO DA PROPOSTA

Este capítulo tem como objetivo apresentar a avaliação da proposta deste trabalho. Para isso, foram realizados dois estudos de caso buscando responder a questão de pesquisa. Desta forma, é possível avaliar a hipótese identificada.

4.1 DEFINIÇÃO DOS ESTUDOS DE CASO

Yin (2015) apresenta um estudo de caso como sendo “uma investigação empírica que investiga um fenômeno contemporâneo dentro de seu contexto da vida real, especialmente quando os limites entre o fenômeno e o contexto não estão claramente definidos”. Especificamente na área de Engenharia de Software, Runeson et al. (2012) apresentam um estudo de caso como sendo “uma investigação empírica que se baseia em múltiplas fontes de evidência para investigar uma instância (ou um pequeno número de instâncias) de um fenômeno de engenharia de software contemporâneo em seu contexto da vida real, especialmente quando os limites entre o fenômeno e contexto não estão claramente definidos”. As avaliações realizadas neste trabalho se baseiam nessas definições.

Visando avaliar os diferentes aspectos que apoiam o processo de composição de serviços na plataforma E-SECO, estudos de caso foram conduzidos. O escopo da avaliação é apresentado com base na estrutura do método GQM (BASILI; WEISS, 1984). Este escopo está definido da seguinte maneira: “**Analisar** a influência de elementos de visualização de serviços, proveniência de dados, anotação semântica e o suporte à colaboração **com o propósito de** apoiar a composição de serviços **sob o ponto de vista de** desenvolvedores e pesquisadores **no contexto de** um ecossistema de software científico”.

Baseando-se no escopo apresentado, derivou-se a seguinte questão de pesquisa: “Como a visualização de serviços, a anotação semântica, a proveniência de dados e o suporte à colaboração potencializam a composição de serviços em um Ecossistema de Software Científico?”. A hipótese nula foi definida como: (H0) A visualização, anotação semântica, proveniência de dados e colaboração não potencializam a composição de serviços em um ecossistema de software científico. A hipótese alternativa, por sua vez, foi definida como: (H1) A visualização, anotação semântica, proveniência de dados e colaboração potencializam a composição de serviços em um ecossistema de software científico. Para responder

a questão de pesquisa, estudos de caso foram realizados.

4.2 PLANEJAMENTO DOS ESTUDOS DE CASO

A preparação dos estudos de caso considerou as habilidades do pesquisador, desejadas para sua realização e os seguintes tópicos apresentados por Yin (2015): (i) treinamento para o estudo, (ii) desenvolvimento de um protocolo e (iii) condução de um estudo de caso piloto.

O treinamento para estudo de caso envolveu a apresentação do processo de composição de serviços na plataforma E-SECO, fornecendo uma documentação que continha casos de uso de utilização da plataforma e um cenário que apresentava a tarefa que deveria ser realizada. Como fontes de coleta de dados foram utilizadas: (i) questionário de avaliação, (ii) entrevistas informais e (iii) observação direta. Os questionários de avaliação utilizados são apresentados no Apêndice C e no Apêndice D.

4.2.1 DEFINIÇÃO DOS PARTICIPANTES

Participaram deste estudo pesquisadores (mestrandos) da UFJF, com diferentes níveis de conhecimento em relação ao desenvolvimento em um ecossistema de software científico. Todos os participantes assinaram um termo de consentimento, conforme apresentado no Apêndice A.

Este estudo de caso envolve os diferentes papéis de **stakeholders** que atuam no ecossistema, assim os participantes foram definidos separadamente desempenhando papéis de desenvolvedor e pesquisador. A atuação dos participantes no estudo ocorreu da seguinte forma: os participantes que possuem o papel de **desenvolvedor** são responsáveis por entender a plataforma, criando composições e utilizando os mecanismos de visualização para analisar a interoperabilidade e dependências funcionais de serviços, bem como para identificar potenciais pesquisadores que podem avaliar a composição realizada. Para tanto, os participantes que possuem o papel de **pesquisador** são responsáveis por avaliar as composições realizadas de acordo com a especificação solicitada previamente ao desenvolvedor. Com o objetivo de identificar os participantes, foi utilizado um questionário de caracterização, conforme apresentado no Apêndice B. Este questionário teve como objetivo identificar o conhecimento de cada participante em relação aos seguintes itens: (i)

ecossistema de software, (ii) desenvolvimento de *web services*, (iii) anotação semântica, e (iv) proveniência de dados.

Com base no objetivo do estudo de caso, as seguintes medidas foram utilizadas para avaliar o estudo de caso:

- **Percepção:** compreensão das atividades pelos participantes (Qualitativa) e execução correta das atividades (Qualitativa);
- **Satisfação do Usuário:** busca avaliar se o usuário se sentiu satisfeito ao utilizar os mecanismos solicitados e considera que os mecanismos possam ser úteis no futuro para apoiar a composição dos serviços na plataforma E-SECO (Qualitativa);
- **Cooperação:** como a interação entre o desenvolvedor (coordenador) e um pesquisador no espaço de trabalho compartilhado, permitiu avaliar uma composição (Qualitativa);
- **Comunicação:** diz respeito à quantidade e à qualidade de mensagens na avaliação da composição pelos pesquisadores (Quantitativa/Qualitativa);
- **Coordenação:** diz respeito à utilização do mecanismo de ARS para apoiar o coordenador no processo de formação da equipe que avaliará a composição (Qualitativa).

A Figura 4.1 apresenta como as dimensões de cooperação, coordenação e comunicação são apresentadas como apoio ao trabalho colaborativo na plataforma. A Figura 4.1-A destaca como a coordenação (através da identificação e seleção de pesquisadores na ARS) permite apoiar a cooperação entre os participantes. Assim, os participantes selecionados podem utilizar os elementos de visualização para avaliar as composições. A cooperação está relacionada às operações conjuntas com o objetivo de avaliar as composições a partir da visualização oferecida pela plataforma, através de um espaço de trabalho compartilhado. A Figura 4.1-B, por sua vez, apresenta como a coordenação pode apoiar a comunicação, através da identificação e seleção de participantes. A Figura 4.1-C destaca como a comunicação entre os participantes pode apoiar a cooperação. A partir da interação entre os pesquisadores, é possível realizar discussões sobre as composições realizadas, visando tomar decisões sobre o resultado obtido.

As medidas de percepção e satisfação do usuário são avaliadas de maneira qualitativa através de um questionário. A medida de comunicação, por sua vez, considera a qualidade

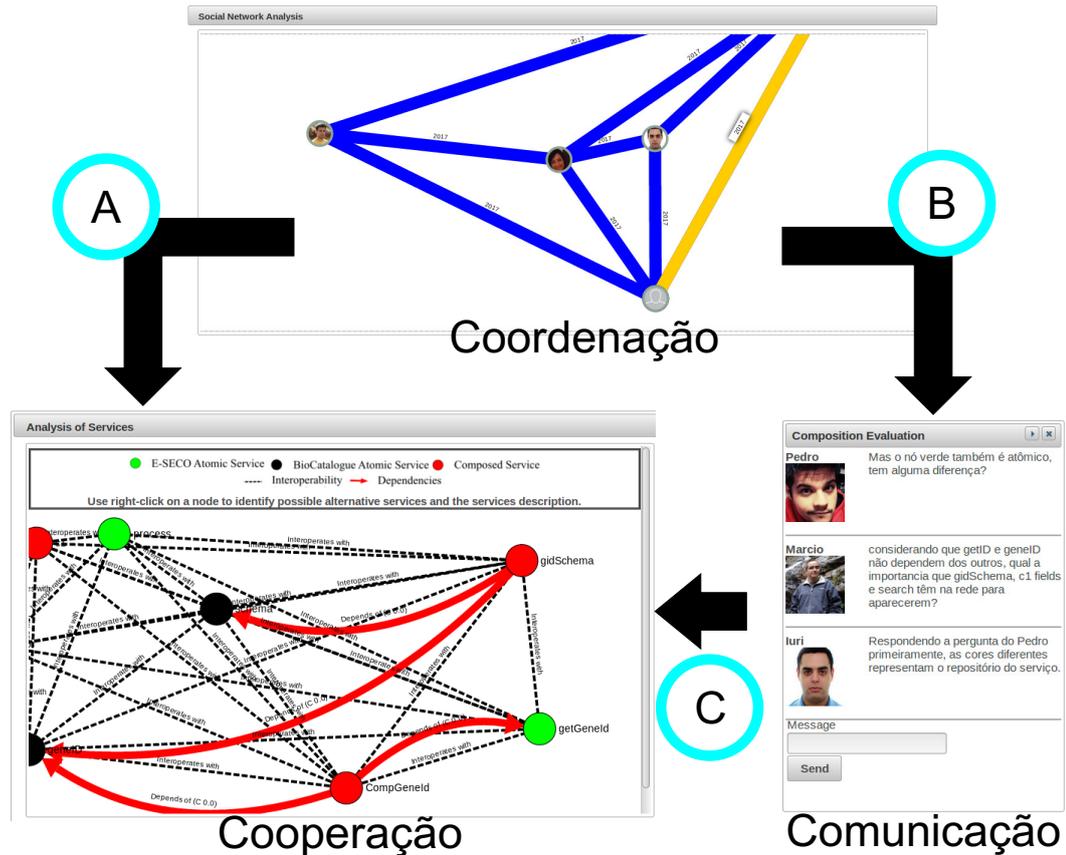


Figura 4.1: Dimensões relacionadas à colaboração no apoio à composição de serviços

das mensagens trocadas entre os participantes, avaliada de maneira qualitativa. A quantidade de mensagens trocadas por cada participante em relação ao total de mensagens trocadas, também é avaliada. A medida de cooperação considera a percepção dos participantes, tendo como objetivo comum a interação no espaço de trabalho compartilhado. Busca-se entender como os elementos de visualização permitem apoiar a cooperação. Assim, ela é avaliada de maneira qualitativa a partir de um questionário. A medida de coordenação, por sua vez, busca avaliar como a análise de redes sociais permitiu ao coordenador (desenvolvedor) identificar os pesquisadores a partir da visualização de uma rede social científica. Essa medida também é avaliada de maneira qualitativa a partir de um questionário.

Um estudo piloto foi realizado inicialmente. Um estudo deste tipo é realizado como o objetivo de avaliar incoerências no planejamento inicial, além de identificar e realizar ajustes nas fontes de coleta de dados, por exemplo (YIN, 2015).

4.3 ESTUDO PILOTO

O estudo piloto envolveu três participantes. Buscou-se participantes que possuíam conhecimento sobre questões que envolvem desenvolvimento em ECOS. Para desempenhar o papel de desenvolvedor no estudo de caso, o participante necessariamente deveria conhecer a abordagem de um ECOS e ter experiência prévia de desenvolvimento nesta área. A partir da caracterização dos participantes, identificada pelo questionário, um dos participantes desempenhou o papel de desenvolvedor enquanto os outros desempenharam o papel de pesquisador. A primeira etapa, caracterizada como um estudo de caso, considerou um único participante, responsável pelo desenvolvimento. A segunda etapa (estudo de viabilidade) considerou os três participantes (coordenador e dois pesquisadores) e teve como objetivo avaliar a interação entre os participantes para avaliar uma composição. Para auxiliar o participante que realizou a etapa de desenvolvimento, dois casos de uso foram elaborados visando apresentar o processo de composição de serviços. Estes casos de uso são apresentados na especificação para o estudo de caso, disponível no Apêndice E.

Para a realização da primeira etapa foi solicitado ao participante (desenvolvedor) que realizasse uma busca por 5 serviços pré-estabelecidos, nos repositórios integrados à plataforma. Foi determinado que o desenvolvedor deveria encontrar um serviço que atendesse a uma especificação. A especificação foi descrita como “buscar um serviço que permita obter um identificador de um gene em uma base de dados a partir do nome de um gene”. Este identificador deve ser informado no formato de uma única cadeia de caracteres, para isto é necessária uma composição. O objetivo era identificar um serviço que tenha sido utilizado anteriormente em algum experimento, através dos dados de proveniência. Para isto, o desenvolvedor deveria analisar os dados de proveniência para os métodos de cada serviço. Assim, é possível conhecer o contexto de utilização do serviço encontrado e reutilizá-lo. Em seguida, foi solicitada uma anotação semântica do serviço encontrado. Uma anotação semântica e registro de um serviço também foram solicitados para os novos serviços (atômico e composto) que fossem criados. Após a realização da composição, o desenvolvedor deveria realizar uma busca entre os serviços registrados a partir dos parâmetros do serviço composto. A partir dos serviços listados, o desenvolvedor deveria selecionar o serviço composto desenvolvido e analisar suas dependências. Por fim, o pesquisador realizou uma análise de uma rede social científica. Foi solicitado que o desenvolvedor identificasse

as relações de colaboração de um pesquisador pertencente a UFJF no ano de 2017. Foi solicitado ao desenvolvedor a análise do perfil deste pesquisador. Assim, seria possível identificar outros pesquisadores com perfil semelhante para a avaliação de composições.

A primeira etapa teve como objetivo avaliar o processo de composição. Porém, com o foco na forma pela qual a interação dos participantes permite avaliar uma composição, foi realizada uma outra etapa, envolvendo os três participantes (desenvolvedor e dois pesquisadores). Esta etapa considerou a utilização do serviço de avaliação de composições. Os participantes que desempenharam o papel de pesquisador desconheciam a especificação de composição solicitada ao participante que desempenhou o papel de desenvolvedor. Foi solicitado aos participantes que, a partir da interação entre eles e da visualização dos serviços, realizassem discussões para avaliar se a composição realizada atendia à especificação. A especificação do serviço composto desenvolvido é apresentada no Apêndice E. A especificação do estudo de viabilidade é apresentada no Apêndice F.

4.3.1 RESULTADOS OBTIDOS DO ESTUDO PILOTO

Os resultados do estudo piloto são analisados em duas etapas. Na primeira etapa é considerada uma análise observacional, com dados obtidos através da observação direta. Em seguida, é apresentada uma avaliação do processo, de acordo com dados qualitativos obtidos através de pela utilização de questionário. A segunda etapa é apresentada como um estudo de viabilidade. Esta etapa tem como objetivo analisar como a interação entre um desenvolvedor e pesquisadores, utilizando elementos de colaboração, apoia a composição.

4.3.1.1 Etapa 1

- Análise Observacional

O participante realizou a busca corretamente e conseguiu identificar metadados associados a cada um dos serviços, assim como os dados de proveniência. Dados como o experimento e o *workflow* nos quais o serviço foi utilizado anteriormente foram identificados corretamente. Porém, no início da análise observou-se uma dificuldade inicial do participante. Esta dificuldade se refere à incompreensão de que cada método de um serviço poderia ser anotado como um serviço atômico, o que fez o participante realizar a anotação semântica de vários serviços atômicos. Porém, após revisar a descrição do

escopo especificado, o participante entendeu que deveria utilizar apenas um dos métodos deste serviço, e este método deveria ter sido utilizado em um experimento anterior. Além disso, foi observado que o participante não identificou o que significava o termo “Task”. Este termo foi utilizado para descrever métodos de um serviço específico que poderiam ser anotados semanticamente como serviços atômicos. Este termo é utilizado para identificar cada serviço atômico executado em um *workflow*. Neste caso, foi identificada a necessidade de apresentar esta descrição ao desenvolvedor. Desde a realização da busca por serviços até a realização da anotação semântica, o desenvolvedor realizou as atividades propostas no tempo de 09:19 minutos.

Após realizar a anotação semântica do serviço solicitado o desenvolvedor iniciou o desenvolvimento da composição de serviços. Para isso, o desenvolvedor necessitou criar um cliente de *web service* para consumir e testar o serviço obtido de um repositório externo. Um fator importante foi observado nesta etapa: embora o arquivo no formato OWL, que possuía a anotação semântica realizada, permitisse identificar a localização do arquivo WSDL, o participante não percebeu que poderia utilizar este arquivo. Para identificar o WSDL o desenvolvedor realizou uma nova busca na plataforma e clicou sobre o botão “download” para recuperar o arquivo WSDL. Embora esta ação tenha sido correta, observou-se um número excessivo de passos para realizar uma atividade que poderia ser facilitada caso a descrição do serviço fornecesse um *link* para a localização deste arquivo. Considerando que esta informação é disponibilizada pelo repositório externo utilizado, observou-se então a necessidade de melhorar a apresentação de metadados de um serviço com a adição da localização deste arquivo.

Durante a criação de um cliente para o serviço o desenvolvedor teve dificuldades para executar o serviço externo, devido a uma inconsistência na marcação identificada no arquivo WSDL. Porém, por possuir a anotação semântica do serviço, conseguiu executar corretamente o serviço externo. Apesar das dificuldades de desenvolvimento encontradas, o desenvolvedor foi capaz de realizar corretamente a composição de serviços e anotar semanticamente os novos serviços desenvolvidos. O processo foi realizado de maneira apropriada e necessidades de adequações foram observadas. Essas adequações se referem ao fornecimento de maior descrição textual relacionadas aos serviços. Tais adequações foram realizadas após a realização do estudo de caso piloto. A etapa de desenvolvimento durou um tempo total de 54:11 minutos.

Em seguida, uma análise de dependências do serviço composto foi realizada. Foi especificado que o desenvolvedor deveria realizar uma busca pelo serviço composto gerado e identificar as relações com outros serviços registrados. Como ponto positivo foi observado que o usuário informou corretamente um parâmetro de entrada, porém informou de maneira errada o parâmetro de saída do serviço. Por considerar uma análise que considera diferentes graus entre a interface do serviço e a requisição de serviços, a busca retornou os serviços relevantes corretamente, apesar de um dos parâmetros não ter sido fornecido corretamente. O desenvolvedor conseguiu identificar corretamente as relações de dependências e interoperabilidade em relação ao serviço selecionado.

Por fim, o desenvolvedor (coordenador) necessitou identificar um pesquisador de uma instituição específica e suas relações em determinado ano. O desenvolvedor conseguiu realizar esta tarefa adequadamente, porém não utilizou o filtro disponível para identificar relações em um ano específico. Além disso, não foi informado ao desenvolvedor sobre a possibilidade de selecionar o pesquisador e, com o botão direito, identificar as relações de cooperação e tipos de relações existentes. Embora o desenvolvedor tenha identificado corretamente estas relações, realizou uma análise visual desnecessária, considerando que poderia realizar um número de passos menor se tivesse maior descrição sobre estas possibilidades. O desenvolvedor realizou a busca de serviços relevantes, análise de dependências de serviços e análise de redes sociais em um tempo total de 11:04 minutos.

- Avaliação

Ao final da realização desta etapa, foi avaliado que o participante compreendeu corretamente as atividades solicitadas, apesar de algumas dificuldades de entendimento da especificação terem sido observadas no início do estudo. O participante classificou como “moderado” o nível de dificuldade encontrado para a realização da composição de serviços. Esta classificação relaciona-se principalmente às dificuldades encontradas durante o processo de desenvolvimento. Apesar das dificuldades encontradas, e apresentadas anteriormente, as atividades foram realizadas corretamente. O participante classificou como “muito bom” a sua satisfação e possibilidade de utilização do mecanismo apresentado para apoiar a composição de serviços, em relação à sua utilização no futuro. Avaliou como “bom” a relevância de utilização de um serviço de proveniência de dados para permitir analisar dados históricos. Em relação às métricas utilizadas para analisar dependências de serviços, o usuário considerou que as métricas apresentadas foram suficientes para

identificar as relações. Em relação à análise de redes sociais, o participante também avaliou como “muito bom” a possibilidade de identificar as relações de colaboração e como “bom” a possibilidade de identificar os tipos de relações. Foi observado que o participante identificou corretamente as relações de colaboração.

4.3.1.2 Etapa 2 - Estudo de Viabilidade

Esta etapa permitiu avaliar a interação entre os participantes identificados. Assim, buscou-se avaliar como a colaboração é apoiada na plataforma, através dos elementos de colaboração apresentados (coordenação, cooperação, comunicação e percepção). Durante a avaliação da interação entre os participantes, observou-se que eles entenderam e realizaram corretamente as atividades solicitadas. Porém, em relação ao serviço de colaboração utilizado para apoiar as composições, os participantes responderam a seguinte pergunta: “Você considera que o serviço de colaboração utilizado foi suficiente para apoiar a avaliação da composição de serviços em um ecossistema de software científico? Justifique”. Os participantes fizeram os seguintes comentários:

- *“O serviço é útil para apoiar a avaliação da composição de serviços, pois apresenta visualmente as informações necessárias para o entendimento do funcionamento dos serviços. Para evoluir, apresentação de legendas para as informações apresentadas podem ser úteis”;*
- *“Eu incluiria um meio de descrever o que cada serviço realiza, desta forma tornaria mais claro o entendimento do papel de cada nó na rede. Uma legenda explicando as cores dos nós iria auxiliar a entender de maneira mais simples também”;*
- *“Não. Apesar de um chat facilitar muito a atividade de comunicação, senti falta de elementos que poderiam facilitar a comunicação, como ser capaz de indicar serviço(s) sobre os quais estou falando, sem ter que ficar digitando os nomes destes no chat”;*

Os comentários sobre esta pergunta permitiram identificar a necessidade de adicionar mais informações descritivas e contextuais sobre o serviço, o que permite melhorar a visualização dos serviços e facilitar a interação entre os usuários. Estas adequações foram realizadas após a realização do estudo de caso piloto.

Tabela 4.1: Colaboração entre Participantes Durante o Estudo de Caso Piloto

Participante	Total de mensagens
Participante 1 (Desenvolvedor)	38 (66,7%)
Participante 2 (Pesquisador)	13 (22,8%)
Participante 3 (Pesquisador)	6 (10,5%)

Os participantes avaliaram como positiva a qualidade das mensagens trocadas. Buscando avaliar a colaboração entre os participantes, a interação foi avaliada também baseando-se na quantidade de mensagens de cada um dos participantes em relação ao total. Ao todo a interação foi realizada durante um período de 41:38 minutos e teve um total de 57 mensagens trocadas. A Tabela 4.1 apresenta a colaboração de cada um dos participantes na interação:

A interação entre os participantes permitiu avaliar a importância do serviço de colaboração para apoiar a composição de serviços na plataforma. Porém, a partir da análise das mensagens entre os participantes foi possível observar que o desenvolvedor interagiu mais em relação aos outros. Isto permitiu identificar uma necessidade de maior descrição textual dos serviços para que o desenvolvedor não necessite detalhar a necessidade do serviço e sua utilização. Este estudo piloto apresentou indícios de que a visualização e a colaboração podem potencializar a composição de serviços em uma plataforma de ECOSC. Porém, novas adaptações necessitaram ser realizadas para a condução de um estudo de caso regular.

4.4 ESTUDO DE CASO

O segundo estudo de caso realizado (regular) buscou responder as mesmas questões do estudo piloto e foi realizado em um mesmo cenário. Novos participantes foram selecionados de acordo com as características observadas. As mesmas etapas do estudo anterior foram realizadas. Porém, pequenas correções foram realizadas baseando-se nos indícios observados no estudo piloto, bem como nos ajustes realizados, tais como:

- Ajustes nos questionários buscando obter resultados mais precisos;

- Ajustes na descrição dos serviços relacionados às dificuldades de localizar informações sobre o serviço;
- Adição de legendas na visualização de serviços, buscando permitir aos participantes identificar corretamente o significado de cores e tipos de arestas do grafo gerado;
- Adaptação do treinamento, buscando informar características da plataforma desconhecidas pelos participantes.

Da mesma forma que no estudo piloto, este estudo contou com três participantes (diferentes do estudo piloto). A partir da caracterização dos participantes, um dos participantes foi selecionado para desempenhar o papel de desenvolvedor enquanto os outros atuaram como pesquisadores. O conhecimento sobre a abordagem de ECOS era diferente entre os participantes, por isso, o papel de desenvolvedor deveria ser realizado por um participante que necessariamente tivesse conhecimento sobre o processo de desenvolvimento de serviços e sobre a abordagem de ECOS. Para a segunda etapa (estudo de viabilidade), o desenvolvedor passou a assumir o papel de coordenador das atividades, nas quais interagiu com os outros participantes (pesquisadores). Isto foi realizado nas duas avaliações.

4.4.1 RESULTADOS OBTIDOS DO ESTUDO DE CASO

Assim como no estudo piloto, os resultados do estudo de caso também são analisados em duas etapas. As mesmas etapas do estudo piloto são realizadas, sendo que a segunda etapa é caracterizada como um estudo de viabilidade, por buscar avaliar como ocorreu a interação entre o desenvolvedor e pesquisadores no estudo, utilizando elementos de colaboração.

4.4.1.1 Etapa 1

- Análise Observacional

Assim como no estudo piloto, o participante realizou corretamente a busca, identificando os metadados associados a cada um dos serviços, assim como dados de proveniência. O participante considerou informações de contexto importantes para identificar os serviços na plataforma. Ao contrário do primeiro estudo de caso, devido às informações adicionais sobre a plataforma fornecidas durante o treinamento de utilização, o participante não teve

dificuldades para realizar a anotação semântica de serviços. Desde a realização da busca por serviços até a realização da anotação semântica, o participante realizou as atividades propostas no tempo de 09:28 minutos. Um tempo muito próximo ao observado no estudo de caso piloto (onde foi gasto um tempo total de 09:19 minutos). A diferença de tempo ocorre principalmente pelo tempo necessário para encontrar os serviços no repositório. O desenvolvedor passou a ter mais descrições textuais e maior facilidade para anotação semântica no estudo regular. Assim, realizou a anotação semântica em menor tempo em comparação ao estudo piloto. No entanto, no início da atividade, necessitou de mais tempo para encontrar um serviço adequado nos repositórios em comparação ao estudo piloto.

O participante também conseguiu realizar corretamente a composição e anotação de serviços. A etapa de desenvolvimento durou um período total de 32:10 minutos. Tempo inferior quando comparado ao estudo piloto (onde foi gasto um total de 54:11 minutos para esta atividade). Isto permitiu observar uma menor dificuldade encontrada para realizar a composição. O participante do estudo regular demonstrou menor dificuldade para compor os serviços solicitados (conforme informado através de um questionário) em comparação ao participante do estudo piloto. Acredita-se que a redução do esforço para a composição resultou de uma melhor apresentação dos recursos da plataforma (através de um treinamento) e de uma maior descrição dos serviços.

O participante deveria realizar uma análise das dependências funcionais de um serviço. Como foi indicado durante o treinamento que informar os parâmetros era opcional, o participante optou por não informar nenhum parâmetro. Desta forma foram retornados todos os serviços registrados. O desenvolvedor conseguiu identificar o serviço solicitado. Assim como no primeiro estudo, o desenvolvedor conseguiu, também, identificar corretamente as relações de dependências e interoperabilidade em relação ao serviço selecionado.

Por último, o desenvolvedor (coordenador) identificou um pesquisador na rede social, conforme solicitado. Os filtros foram utilizados corretamente e foi possível identificar as relações de colaboração. O desenvolvedor realizou a busca de serviços relevantes, análise de dependências de serviços e análise de redes sociais em um tempo total de 05:26 minutos. Esse tempo foi inferior à metade do tempo gasto no estudo piloto (onde foram gastos 11:04 minutos). Através da adição de uma legenda que tinha como objetivo descrever recursos adicionais no grafo, o desenvolvedor conseguiu realizar essa atividade com maior

facilidade. Desta forma, ele buscou um único pesquisador e analisou suas colaborações ao clicar com o botão direito sobre o pesquisador. Assim, não precisou realizar análises manuais e desnecessárias no grafo.

- Avaliação

Ao final da realização desta etapa, foi avaliado que o participante compreendeu corretamente as atividades. O desenvolvedor classificou como “baixo” o nível de dificuldade para realizar a composição solicitada. Considerou como “bom” a utilização de um serviço para análise de dados de proveniência, considerando a importância de informações contextuais. O participante também classificou como “bom” em relação à compreensão das atividades e em relação a sua satisfação e possibilidade de utilização do mecanismo apresentado para apoiar a composição de serviços, em relação a sua utilização no futuro. Em relação às métricas utilizadas para analisar dependências de serviços, o usuário considerou que as métricas apresentadas foram suficientes para identificar as relações. O participante classificou como “muito bom” a capacidade de identificar as dependências entre serviços. Em relação à análise de redes sociais, o participante avaliou como “bom” a possibilidade de identificar as relações de colaboração e de identificar os tipos de relações. Assim como no estudo piloto, foi observado que o participante identificou corretamente e sem problemas as relações de colaboração.

4.4.1.2 Etapa 2 - Estudo de Viabilidade

Esta etapa permitiu avaliar a interação entre os participantes identificados. Novas questões foram adicionadas ao questionário, buscando avaliar as dificuldades encontradas e permitindo analisar os mecanismos de percepção disponíveis em relação à coordenação, comunicação e cooperação. A Figura 4.2 apresenta uma comparação entre respostas obtidas no estudo de caso piloto e no estudo de caso regular. Essas respostas se relacionam à colaboração para apoiar a composição de serviços em uma plataforma de ECOSC. O gráfico apresenta a avaliação dos usuários em relação à capacidade de identificar as composições realizadas. Uma segunda questão refere-se a avaliação dos usuários ao apoio oferecido pelos serviços de colaboração (comunicação, coordenação, cooperação e percepção) utilizados para avaliar as composições realizadas. A terceira questão avaliada refere-se à satisfação dos usuários ao interagir na plataforma e visualizar as composições realizadas,

em relação à utilidade dos serviços no futuro para apoiar a composição de serviços.

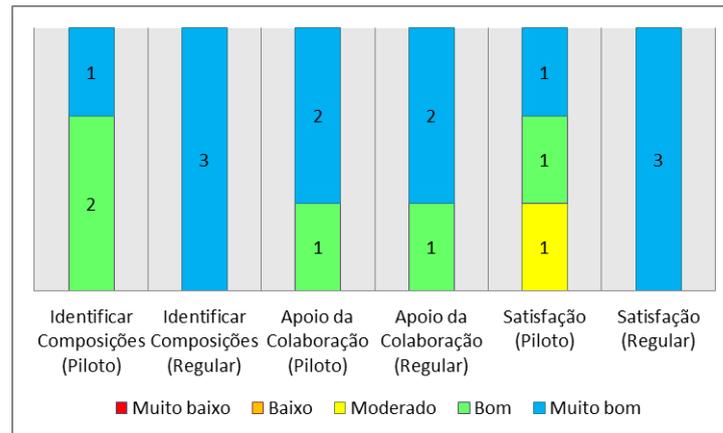


Figura 4.2: Análise Comparativa do Apoio Oferecido Pela Colaboração

Quando comparado ao estudo piloto, foi possível observar que no estudo de caso regular os participantes tiveram menos dificuldades de interação na plataforma. A Figura 4.2 permite analisar uma maior facilidade para identificar as composições e maior satisfação dos usuários ao interagir na plataforma. Provavelmente, isto ocorreu porque o desenvolvedor (coordenador) das atividades adquiriu mais conhecimento sobre a visualização durante o treinamento. Essa informação foi obtida através de entrevistas informais com o participante, que afirmou conhecer os recursos disponíveis na plataforma a partir do treinamento. Além disso, com a adição de uma legenda e maior descrição textual sobre o serviço desenvolvido, os participantes não necessitaram fazer perguntas sobre o significado das cores dos serviços no grafo ou o que significava as linhas (arestas) tracejadas ou direcionadas (que indicam relações de dependências funcionais e interoperabilidade).

A Figura 4.3 apresenta o grafo analisado durante a realização deste estudo de caso, através da qual é possível observar a interação entre os participantes e a cooperação através do espaço de trabalho compartilhado. A legenda adicionada também é apresentada acima do grafo gerado. Essa legenda auxiliou na redução do esforço por parte dos pesquisadores para identificar tipos de serviços e relações entre eles.

Ao contrário do estudo piloto, neste estudo os participantes consideraram que o serviço de comunicação utilizado foi suficiente para apoiar a composição. No primeiro estudo os participantes sentiram a necessidade de maior descrição textual dos serviços, o que não ocorreu no segundo.

Buscando avaliar a colaboração entre os participantes, a interação foi avaliada baseando-se na quantidade de mensagens de cada um dos participantes em relação ao total. Ao

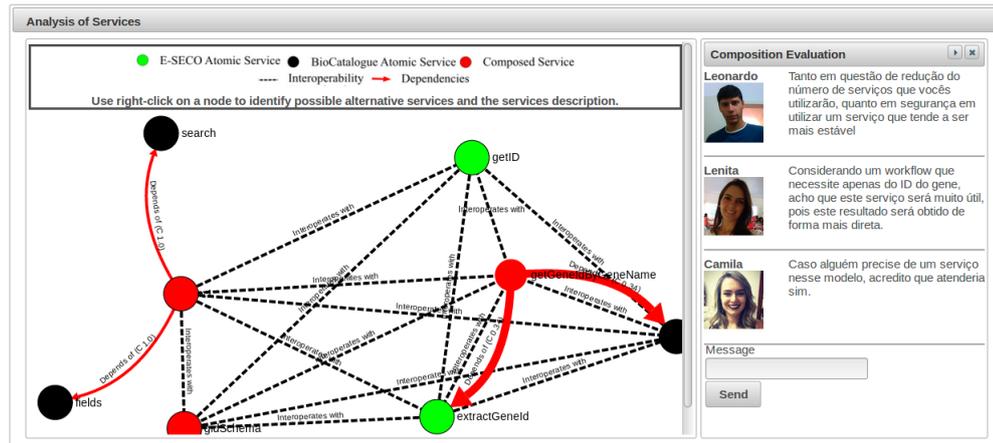


Figura 4.3: Avaliação do Grafo Gerado no Estudo de Caso Regular

Tabela 4.2: Colaboração entre Participantes Durante o Estudo de Caso Regular

Participante	Total de mensagens
Participante 1 (Desenvolvedor)	30 (57,7%)
Participante 2 (Pesquisador)	9 (17,3%)
Participante 3 (Pesquisador)	13 (25%)

todo a interação foi realizada durante um período de 51:49 minutos e teve um total de 52 mensagens trocadas. A Tabela 4.2 apresenta a troca de mensagens entre os participantes neste Estudo de Caso.

O número de mensagens trocadas foi bem próximo ao observado estudo piloto. Porém, no segundo estudo foi observado que o desenvolvedor (coordenador) não teve a necessidade de informar questões que envolvem a cooperação, relacionadas ao grafo. As mensagens trocadas nesse estudo tiveram um objetivo principal de identificar a possibilidade de utilização do serviço (validar a composição) do que tratar questões técnicas, como ocorreu no estudo piloto. Isto pode ser observado pela redução do número de mensagens realizadas pelo desenvolvedor. Os participantes não identificaram problemas técnicos durante a realização do estudo de caso. Problemas relacionados à comunicação, coordenação e cooperação também não foram identificados.

Em relação aos mecanismos de percepção, o coordenador concordou que foi possível identificar os participantes da avaliação. Todos os participantes concordaram que os mecanismos de percepção permitiram identificar os participantes através do *chat*. Além disto, os participantes concordaram que os elementos de percepção disponíveis para a

cooperação permitiram apoiar a interação entre os participantes.

Os resultados obtidos através da condução destes estudos oferecem indícios que elementos de visualização, anotação semântica, proveniência de dados e colaboração podem apoiar a composição de serviços na plataforma. Desta forma não foram encontrados indícios que permitam rejeitar a hipótese (H1): “A visualização, anotação semântica, proveniência de dados e colaboração potencializam a composição de serviços em um ecossistema de software científico”.

4.5 AMEAÇAS À VALIDADE

A condução dos estudos de caso (piloto e regular) apresentam algumas ameaças à validade. A validade de um estudo se relaciona à confiabilidade dos resultados e a avaliar se os resultados não são tendenciosos a partir do ponto de vista subjetivo dos pesquisadores (RUNESON et al., 2012). Para avaliar as ameaças, são considerados quatro aspectos de qualidade: interna, externa, constructo e conclusão (YIN, 2015).

4.5.1 VALIDADE INTERNA

A ameaça à validade interna relaciona-se a uma possível incompreensão das questões técnicas da plataforma E-SECO. Embora um treinamento tenha sido realizado, uma dificuldade de compreensão sobre a plataforma pode ocorrer. Uma incompreensão dos processos de busca e composição de serviços pode ter influenciado os resultados.

O suporte à colaboração também possui ameaças à validade interna, considerando que outros serviços poderiam ser utilizados (fórum de discussões, suporte a áudio, vídeo, entre outros). Para a coordenação de atividades, uma possível ameaça se relaciona à quantidade de informações sobre os pesquisadores selecionados para avaliação. Um número maior de informações sobre os pesquisadores poderia auxiliar na coordenação. Em relação ao processo de desenvolvimento da plataforma, uma ameaça que também deve ser considerada se relaciona à disponibilização de serviços na base e mais informações de proveniência. Uma integração com ferramentas de testes para avaliar a composição também poderia ser realizada.

4.5.2 VALIDADE EXTERNA

A proposta realizada está inserida em um domínio específico. Por esta questão, em relação à validade externa, este estudo está limitado ao contexto de uma plataforma que utiliza a abordagem de um ECOSC.

4.5.3 VALIDADE DE CONSTRUCTO

Durante a avaliação, é possível que algum participante não tenha compreendido corretamente alguma das questões apresentadas no questionário. Caso isto tenha ocorrido, existe uma ameaça à validade do projeto. Outra ameaça à validade de constructo se relaciona à utilização e ao correto entendimento das métricas utilizadas para analisar dependências funcionais entre serviços e a importância de serviços na rede. Essa mesma ameaça existe na utilização de métricas para a análise de redes sociais científicas. Cabe ressaltar também que, apesar dos indícios observados sobre a utilização das métricas, outras métricas poderiam ser utilizadas neste projeto.

4.5.4 VALIDADE DE CONCLUSÃO

Em relação à validade dos resultados, embora os estudos tenham sido realizados em uma plataforma de ECOSC, novos estudos necessitam ser realizados. Os estudos aqui realizados não consideraram problemas reais. O número reduzido de participantes também é considerado uma ameaça à validade dos resultados.

4.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Neste capítulo o apoio à composição de serviços na plataforma E-SECO, através de elementos de visualização, e colaboração, foi avaliado. Além do apoio oferecido pela proveniência de dados e anotação semântica. Foi possível avaliar as etapas que compõem o processo, e identificar como os elementos de colaboração (percepção, coordenação, comunicação e cooperação) podem apoiar o processo de composição. Os resultados apresentam indícios que o processo de composição apresentado neste trabalho pode apoiar a composição. Apesar dos indícios observados, novos estudos devem ser realizados, considerando outros contextos reais de experimentação científica.

5 CONCLUSÕES

Este capítulo tem como objetivo apresentar um resumo do trabalho e suas contribuições, além de suas limitações e trabalhos futuros.

5.1 BREVE RESUMO

Este trabalho apresentou conceitos relacionados a área de *e-Science*, caracterizando o ciclo de vida de um experimento científico. Posteriormente foram apresentados conceitos relacionados a um ECOS e as dificuldades existentes para a composição de serviço neste contexto. As dificuldades observadas se relacionam às dimensões técnicas e sociais de um ECOS. Questões técnicas estão relacionadas, principalmente, à distribuição de recursos e ao aumento do número de componentes. Questões sociais, por sua vez, estão relacionadas ao apoio da colaboração em um ECOS, buscando gerar conhecimento. O suporte à colaboração associado à dimensão social foi visto como uma forma de apoiar a composição de serviços.

A partir das dificuldades observadas, este trabalho buscou apresentar uma solução para tratar a complexidade para compor serviços no contexto de um ECOSC. Esta solução possui como enfoque uma plataforma que utiliza a abordagem de ECOSC. Para isto, foi apresentado um mecanismo que considera o processo de desenvolvimento de serviços da plataforma. O mecanismo buscou associar a proveniência de dados e o suporte à interoperabilidade ao processo de composição de serviços, buscando apoiar, dentre outros requisitos, a flexibilidade para desenvolver serviços. Elementos de visualização e colaboração também foram apresentados para apoiar a atividade de composição de serviços na plataforma.

Estudos foram conduzidos para avaliar a proposta e permitiram obter indícios sobre a viabilidade de utilização do mecanismo proposto no contexto de um ECOSC.

5.2 CONTRIBUIÇÕES

A realização de atividades distribuídas e multidisciplinares é uma prática comum na área de *e-Science*. No contexto de *e-Science* é importante auxiliar os diferentes aspectos rela-

cionados à colaboração. Um suporte à coordenação, comunicação e cooperação é importante para apoiar tomadas de decisão, gerando um conhecimento coletivo. A plataforma E-SECO permite apoiar as diferentes atividades relacionadas à experimentação científica. Essas atividades normalmente envolvem a execução de serviços web em *workflows* científicos.

ECOS possuem diferentes soluções que envolvem o compartilhamento de software e serviços, considerando a interação e colaboração entre atores que utilizam uma determinada plataforma. Por estar em um contexto de ECOSC, questões que envolvem a distribuição de recursos e a extensibilidade da plataforma E-SECO devem ser tratadas. Ao apoiarmos a interação entre especialistas de diferentes áreas estamos também oferecendo suporte aos aspectos técnicos de uma atividade de composição de serviços web. O aperfeiçoamento pode ser alcançado através da interação de especialistas em uma área específica. Como resultado dessa interação, buscamos desenvolver serviços que possam apoiar atividades em experimentos científicos.

Não foram encontrados na literatura trabalhos que consideram a utilização de elementos de visualização e colaboração entre pesquisadores, buscando apoiar a composição de serviços no contexto de uma plataforma de ECOSC. A identificação de *stakeholders* capazes de auxiliar a composição de serviços se mostra importante para o processo de desenvolvimento de serviços na plataforma. Ao mesmo tempo, esta solução considera aspectos técnicos (reutilização, interoperabilidade, dependência funcional, entre outros) que podem influenciar na decisão sobre esta composição. Contribuições para este trabalho foram obtidas inicialmente a partir de comentários e sugestões realizados por pesquisadores da área de Sistemas Colaborativos no “7^o Workshop de Teses e Dissertações em Sistemas Colaborativos” (MARQUES et al., 2016). Posteriormente, um serviço para visualização de composições e redes sociais científicas foi apresentado (MARQUES et al., 2017).

Diante disto, este trabalho apresentou um mecanismo que busca apoiar a composição de serviços na plataforma E-SECO, considerando um suporte à visualização e colaboração, bem como às dimensões técnicas e sociais da plataforma de ECOSC. A solução buscou tratar os seguintes requisitos: (i) busca em diferentes repositórios (externos e interno), (ii) contexto no passado de um serviço (proveniência de dados), (iii) anotação semântica de serviços, (iv) análise semântica de uma requisição, (v) interoperabilidade sintática, semântica e pragmática entre serviços, (vi) visualização de serviços capazes de interoperar

em diferentes níveis e das dependências funcionais para os serviços compostos e (vii) visualização de redes sociais científicas em diferentes níveis.

Para permitir avaliar o mecanismo apresentado, foram realizados estudos de caso, que forneceram indícios sobre sua utilidade na plataforma. A partir dos estudos realizados foi possível responder à questão de pesquisa formulada, onde foram apresentados indícios de que a visualização, a anotação semântica, a proveniência de dados e a colaboração potencializam a composição de serviços em um ECOSC. Como lições aprendidas, foi possível observar a importância de uma maior descrição textual que permita avaliar as visualizações. Estas observações foram feitas durante os estudos de caso. Foi possível identificar também a necessidade de uma melhor apresentação dos recursos oferecidos pela plataforma para os usuários.

5.3 LIMITAÇÕES

Este trabalho está limitado à utilização no contexto de uma plataforma que utiliza a abordagem de um ECOSC. Isso implica que as estratégias aqui apresentadas não podem ser generalizadas. Em outros contextos, essas estratégias podem possuir características distintas.

Além disso, elementos de colaboração precisam ser estudados no contexto de software científico, buscando adequá-los às diferentes atividades de colaboração que são conduzidas no ciclo de vida de um experimento científico. Como resultado, a colaboração neste contexto poderá ser potencializada reduzindo assim, as ameaças à validade do experimento.

5.4 TRABALHOS FUTUROS

Como trabalhos futuros, novas avaliações necessitam ser realizadas, considerando problemas reais de utilização.

Em relação às composições de serviços, novas características de visualizações podem ser adicionadas, de maneira a permitir que o desenvolvedor seja capaz de explorar mais os serviços e composições. Outras redes sociais científicas podem ser integradas à plataforma. Além disto, novos serviços de comunicação podem ser adicionados à plataforma, incluindo recursos como fórum de discussões, suporte a áudio, vídeo, entre outros.

Outros princípios da orientação a serviços, que interferem diretamente na capacidade

de composição de serviços, necessitam ser explorados, tais como: ausência de estado, escalabilidade e portabilidade.

REFERÊNCIAS

- ALTINTAS, I.; BERKLEY, C.; JAEGER, E.; JONES, M.; LUDASCHER, B.; MOCK, S. Kepler: an extensible system for design and execution of scientific workflows. **Scientific and Statistical Database Management, 2004. Proceedings. 16th International Conference on**, IEEE, p. 423–424, 2004.
- AMORIM, S. d. S.; ALMEIDA, E. S. D.; MCGREGOR, J. D. Extensibility in ecosystem architectures: an initial study. **Proceedings of the 2013 International Workshop on Ecosystem Architectures**, ACM, p. 11–15, 2013.
- ANDREWS, T.; CURBERA, F.; DHOLAKIA, H.; GOLAND, Y.; KLEIN, J.; LEYMANN, F.; LIU, K.; ROLLER, D.; SMITH, D.; THATTE, S. et al. **Business process execution language for web services**, 2003.
- ANVAARI, M.; CONRADI, R.; JACCHERI, L. Architectural decision-making in enterprises: preliminary findings from an exploratory study in norwegian electricity industry. **European Conference on Software Architecture**, Springer, p. 162–175, 2013.
- AOYAMA, M. Model and its management architecture of software service supply chains. **Serviceology for Services**, Springer, p. 181–189, 2014.
- ARAKAKI, M.; MARTINS, G.; DAVID, J. M.; BRAGA R.; CAMPOS, F.; NEIVA, F. A process for the development of collaboration services in a scientific software ecosystem. **XIII Brazilian Symposium in Collaborative Systems**, p. 1522–1535, 2016. (In Portuguese).
- BASILI, V. R.; WEISS, D. M. A methodology for collecting valid software engineering data. **IEEE Transactions on software engineering**, IEEE, n. 6, p. 728–738, 1984.
- BECHHOFFER, S. Owl: Web ontology language. **Encyclopedia of Database Systems**, Springer, p. 2008–2009, 2009.
- BELHAJJAME, K.; ZHAO, J.; GARIJO, D.; GAMBLE, M.; HETTNE, K.; PALMA, R.; MINA, E.; CORCHO, O.; GÓMEZ-PÉREZ, J. M.; BECHHOFFER, S. et al. Using

a suite of ontologies for preserving workflow-centric research objects. **Web Semantics: Science, Services and Agents on the World Wide Web**, Elsevier, v. 32, p. 16–42, 2015.

BELLOUM, A.; INDA, M. A.; VASUNIN, D.; KORKHOV, V.; ZHAO, Z.; RAUWERDA, H.; BREIT, T. M.; BUBAK, M.; HERTZBERGER, L. O. Collaborative e-science experiments and scientific workflows. **IEEE Internet Computing**, IEEE, v. 15, n. 4, p. 39–47, 2011.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. et al. The semantic web. **Scientific american**, New York, NY, USA:, v. 284, n. 5, p. 28–37, 2001.

BOSCH, J. From software product lines to software ecosystems. **Proceedings of the 13th international software product line conference**, p. 111–119, 2009.

BOSCH, J.; BOSCH-SIJTSEMA, P. M. Softwares product lines, global development and ecosystems: collaboration in software engineering. **Collaborative Software Engineering. Springer Berlin Heidelberg**, p. 77–92, 2010.

BUNEMAN, P.; KHANNA, S.; WANG-CHIEW, T. Why and where: A characterization of data provenance. **International conference on database theory**, Springer, p. 316–330, 2001.

CAMPBELL, P. R.; AHMED, F. A three-dimensional view of software ecosystems. **Proceedings of the Fourth European Conference on Software Architecture: Companion Volume**, ACM, p. 81–84, 2010.

CAO, Y.; JONES, C.; CUEVAS-VICENTTIN, V.; JONES, M. B.; LUDÄSCHER, B.; MCPHILLIPS, T.; MISSIER, P.; SCHWALM, C.; SLAUGHTER, P.; VIEGLAIS, D. et al. Provone: extending prov to support the dataone scientific community. 2016.

CHEN, F.; LU, C.; WU, H.; LI, M. A semantic similarity measure integrating multiple conceptual relationships for web service discovery. **Expert Systems with Applications**, Elsevier, v. 67, p. 19–31, 2017.

CHRISTENSEN, E.; CURBERA, F.; MEREDITH, G.; WEERAWARANA, S. et al. **Web services description language (WSDL) 1.1**. 2001.

- CRNKOVIC, I.; STAFFORD, J.; SZYPERSKI, C. Software components beyond programming: From routines to services. **IEEE software**, v. 28, n. 3, p. 22–26, 2011.
- DAVIDSON, S. B.; FREIRE, J. Provenance and scientific workflows: challenges and opportunities. **Proceedings of the 2008 ACM SIGMOD international conference on Management of data**, ACM, p. 1345–1350, 2008.
- DEELMAN, E.; BLYTHE, J.; GIL, Y.; KESSELMAN, C.; MEHTA, G.; PATIL, S.; SU, M.-H.; VAHI, K.; LIVNY, M. Pegasus: Mapping scientific workflows onto the grid. **Grid Computing**, Springer, p. 131–140, 2004.
- DEELMAN, E.; GANNON, D.; SHIELDS, M.; TAYLOR, I. Workflows and e-science: An overview of workflow system features and capabilities. **Future generation computer systems**, Elsevier, v. 25, n. 5, p. 528–540, 2009.
- DELICATO, F. C.; PIRES, P. F.; BATISTA, T.; CAVALCANTE, E.; COSTA, B.; BARROS, T. Towards an IoT ecosystem. **Proceedings of the First International Workshop on Software Engineering for Systems-of-Systems**, p. 25–28, 2013.
- EDMONDS, J.; KARP, R. M. Theoretical improvements in algorithmic efficiency for network flow problems. **Journal of the ACM (JACM)**, ACM, v. 19, n. 2, p. 248–264, 1972.
- ELLIS, C. A.; GIBBS, S. J.; REIN, G. Groupware: some issues and experiences. **Communications of the ACM**, ACM, v. 34, n. 1, p. 39–58, 1991.
- ERL, T. Soa: Principles of service design. Prentice Hall; 1st edition, v. 1, 2008.
- ERL, T.; MERSON, P.; STOFFERS, R. Service-oriented architecture: Analysis and design for services and microservices. Prentice Hall; 2 edition, 2017.
- FINLAYSON, M. A. Java libraries for accessing the princeton wordnet: Comparison and evaluation. **Proceedings of the 7th Global Wordnet Conference, Tartu, Estonia**, v. 137, 2014.
- FONTÃO, A. D. L.; SANTOS, R. P.; FILHO, J. F.; DIAS-NETO, A. C. MSECO-DEV: Application development process in mobile software ecosystems. **28th International**

Conference on Software Engineering and Knowledge Engineering, USA, p. 317–322, 2016.

FOUAD, H.; GILLIAM, A.; GULEYUPOGLU, S.; RUSSELL, S. M. Automated evaluation of service oriented architecture systems: a case study. **SPIE Defense+ Security**, International Society for Optics and Photonics, 2017.

FREIRE, J.; KOOP, D.; SANTOS, E.; SILVA, C. T. Provenance for computational tasks: A survey. **Computing in Science & Engineering**, IEEE, v. 10, n. 3, 2008.

FREIRE, J.; SILVA, C. T.; CALLAHAN, S. P.; SANTOS, E.; SCHEIDEGGER, C. E.; VO, H. T. Managing rapidly-evolving scientific workflows. In: SPRINGER. **International Provenance and Annotation Workshop**, 2006. p. 10–18.

FREITAS, V.; DAVID, J. M. N.; BRAGA, R.; CAMPOS, F. An architecture for scientific software ecosystem. **Workshop on Distributed Software Development, Software Ecosystems and Systems-of-Systems (WDES)**, p. 41–48, 2015.

FUKS, H.; RAPOSO, A.; GEROSA, M. A.; PIMENTEL, M.; LUCENA, C. J. The 3c collaboration model. **The Encyclopedia of E-Collaboration, Ned Kock (org)**, p. 637–644, 2007.

FUKS, H.; RAPOSO, A. B.; GEROSA, M. A.; LUCENA, C. J. P. Do modelo de colaboração 3c à engenharia de groupware. **Simpósio Brasileiro de Sistemas Multimídia e Web-Webmidia**, p. 0–8, 2003.

GOBLE, C. A.; BHAGAT, J.; ALEKSEJEVS, S.; CRUICKSHANK, D.; MICHAELIDES, D.; NEWMAN, D.; BORKUM, M.; BECHHOFFER, S.; ROOS, M.; LI, P. et al. myexperiment: a repository and social network for the sharing of bioinformatics workflows. **Nucleic acids research**, Oxford University Press, v. 38, n. suppl_2, p. W677–W682, 2010.

GOECKS, J.; NEKRUTENKO, A.; TAYLOR, J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. **Genome biology**, BioMed Central, v. 11, n. 8, p. R86, 2010.

HAN, J.; PEI, J.; KAMBER, M. **Data mining: concepts and techniques**, 2011.

- HEARST, M. A. Automatic acquisition of hyponyms from large text corpora. **Proceedings of the 14th conference on Computational linguistics-Volume 2**, Association for Computational Linguistics, p. 539–545, 1992.
- HEY, A. J.; TREFETHEN, A. E. The data deluge: An e-science perspective. Wiley and Sons, 2003.
- IDE, N.; PUSTEJOVSKY, J. What does interoperability mean, anyway? toward an operational definition of interoperability for language technology. In: **Proceedings of the Second International Conference on Global Interoperability for Language Resources. Hong Kong, China**, 2010.
- JANSEN, S.; CUSUMANO, M. Defining software ecosystems: A survey of software platforms and business network governance. **Proceedings of IWSECO**, p. 41, 2012.
- JANSEN, S.; CUSUMANO, M. A.; BRINKKEMPER, S. **Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry**, 2013.
- JANSEN, S.; FINKELSTEIN, A.; BRINKKEMPER, S. A sense of community: A research agenda for software ecosystems. In: IEEE. **Software Engineering-Companion Volume, 2009. ICSE-Companion 2009. 31st International Conference on**, 2009. p. 187–190.
- JIMÉNEZ, B. G.; WILKINSON, M. D. Automatic annotation of bioinformatics workflows with biomedical ontologies. **International Symposium On Leveraging Applications of Formal Methods, Verification and Validation. Springer**, p. 464–478, 2014.
- KIM, Y.; STANTON, J. M. Institutional and individual influences on scientists' data sharing practices. **Journal of Computational Science Education**, v. 3, n. 1, p. 47–56, 2012.
- LAFOURCADE, M. Guessing hierarchies and symbols for word meanings through hyperonyms and conceptual vectors. **Advances in Object-Oriented Information Systems**, Springer, p. 31–35, 2002.
- LAKSHMI, M. D.; DHAS, J. P. M. A hybrid approach for discovery of owl-s services based on functional and non-functional properties. **WSEAS Trans. Comput**, v. 14, p. 62–71, 2015.

- LIM, C.; LU, S.; CHEBOTKO, A.; FOTOUHI, F. Prospective and retrospective provenance collection in scientific workflow environments. **Services Computing (SCC), 2010 IEEE International Conference on**, IEEE, p. 449–456, 2010.
- LUNGU, M.; LANZA, M.; GÎRBA, T.; ROBBES, R. The small project observatory: Visualizing software ecosystems. **Science of Computer Programming**, Elsevier, v. 75, n. 4, p. 264–275, 2010.
- MANIKAS, K. Revisiting software ecosystems research: A longitudinal literature study. **Journal of Systems and Software**, Elsevier, v. 117, p. 84–103, 2016.
- MANIKAS, K.; HANSEN, K. M. Software ecosystems—a systematic literature review. **Journal of Systems and Software**, Elsevier, v. 86, n. 5, p. 1294–1306, 2013.
- MARIN, M.; DEURSEN, A. V.; MOONEN, L. Identifying aspects using fan-in analysis. In: IEEE. **Reverse Engineering, 2004. Proceedings. 11th Working Conference on**, 2004. p. 132–141.
- MARINHO, A. S. **PROVMANAGER: Uma Abordagem para Gerenciamento de Proveniência de Experimentos Científicos**. Tese (Doutorado) — dissertação de mestrado, Engenharia de Sistemas e Computação, UFRJ/COPPE, Rio de Janeiro, RJ, Brasil, 2011.
- MARQUES, P.; DAVID, J. M. N.; BRAGA, R. Busca e recuperação de componentes em um ecossistema de software científico. **7^o Workshop de Teses e Dissertações em Sistemas Colaborativos (WTD-SC)**, 2016.
- MARQUES, P.; DAVID, J. M. N.; STRÖELE, V.; BRAGA, R.; CAMPOS, F.; ARAÚJO, M. A. P. Apoiando a composição de serviços em ecossistemas de software científico. **14^o Simpósio Brasileiro de Sistemas Colaborativos**, 2017.
- MARSHALL, S.; JACKSON, K.; ANSLOW, C.; BIDDLE, R. Aspects to visualising reusable components. In: AUSTRALIAN COMPUTER SOCIETY, INC. **Proceedings of the Asia-Pacific symposium on Information visualisation-Volume 24**, 2003. p. 81–88.

- MARTIN, D.; BURSTEIN, M.; HOBBS, J.; LASSILA, O.; MCDERMOTT, D.; MCILRAITH, S.; NARAYANAN, S.; PAOLUCCI, M.; PARSIA, B.; PAYNE, T. et al. Owl-s: Semantic markup for web services. **W3C member submission**, v. 22, 2004.
- MARTIN, D.; PAOLUCCI, M.; MCILRAITH, S.; BURSTEIN, M.; MCDERMOTT, D.; MCGUINNESS, D.; PARSIA, B.; PAYNE, T.; SABOU, M.; SOLANKI, M. et al. Bringing semantics to web services: The owl-s approach. **International Workshop on Semantic Web Services and Web Process Composition**, Springer, p. 26–42, 2004.
- MCGOVERN, J.; TYAGI, S.; STEVENS, M.; MATHEW, S. **Java web services architecture**, 2003.
- MCILRAITH, S. A.; SON, T. C.; ZENG, H. Semantic web services. **IEEE intelligent systems**, IEEE, v. 16, n. 2, p. 46–53, 2001.
- MESSERSCHMITT, D. G.; SZYPERSKI, C. et al. Software ecosystem: understanding an indispensable technology and industry. **MIT Press Books**, The MIT Press, v. 1, 2005.
- MIER, P. R.; PEDRINACI, C.; LAMA, M.; M., M. An integrated semantic Web service discovery and composition framework. **IEEE Transactions on Services Computing**, 2015.
- MILLER, G. A. Wordnet: a lexical database for english. **Communications of the ACM**, ACM, v. 38, n. 11, p. 39–41, 1995.
- MILLER, G. A.; BECKWITH, R.; FELLBAUM, C.; GROSS, D.; MILLER, K. J. Introduction to wordnet: An on-line lexical database. **International journal of lexicography**, Oxford University Press, v. 3, n. 4, p. 235–244, 1990.
- MOGHADDAM, M.; DAVIS, J. G. Service selection in web service composition: A comparative review of existing approaches. **Web Services Foundations**, Springer, p. 321–346, 2014.
- MOREAU, L.; CLIFFORD, B.; FREIRE, J.; FUTRELLE, J.; GIL, Y.; GROTH, P.; KWASNIKOWSKA, N.; MILES, S.; MISSIER, P.; MYERS, J. et al. The open provenance model core specification (v1. 1). **Future generation computer systems**, Elsevier, v. 27, n. 6, p. 743–756, 2011.

- MOREAU, L.; MISSIER, P. Prov-dm: The prov data model. World Wide Web Consortium, 2013.
- MORRIS, C. **Foundations of the theory of signs**, 1938. (International encyclopedia of unified science). Disponível em: <<https://books.google.com.br/books?id=QmXvAAAAIAAJ>>.
- MOURA, A. D. C.; PORTO, F.; POLTOSI, M.; PALAZZI, D. C.; MAGALHÃES, R. P.; VIDAL, V. Integrating ecological data using linked data principles. **Joint V Seminar on Ontology Research in Brazil (ONTOBRAS)**, p. 156–167, 2012.
- NASERI, M.; LUDWIG, S. A. Automatic service composition using pomdp and provenance data. **Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on**, IEEE, p. 246–253, 2013.
- NEIVA, F. W.; DAVID, J. M. N.; BRAGA, R.; CAMPOS, F.; FREITAS, V. Prime: Pragmatic interoperability architecture to support collaborative development of scientific workflows. **Components, Architectures and Reuse Software (SBCARS), 2015 IX Brazilian Symposium on**, IEEE, p. 50–59, 2015.
- NORTHROP, L. M. Sei’s software product line tenets. **IEEE software**, IEEE, v. 19, n. 4, p. 32–40, 2002.
- OINN, T.; LI, P.; KELL, D. B.; GOBLE, C.; GODERIS, A.; GREENWOOD, M.; HULL, D.; STEVENS, R.; TURI, D.; ZHAO, J. Taverna/my grid: aligning a workflow system with the life sciences community. **Workflows for e-Science**, Springer, p. 300–319, 2007.
- OLSON, G. M. The next generation of science collaboratories. In: IEEE. **Collaborative Technologies and Systems, 2009. CTS’09. International Symposium on**, 2009. p. 15–16.
- PAOLUCCI, M.; KAWAMURA, T.; PAYNE, T.; SYCARA, K. Semantic matching of web services capabilities. **International Semantic Web Conference. Springer Berlin Heidelberg**, p. 333–347, 2002.
- PESSOA, R. M.; SILVA, E.; SINDEREN, M. V.; QUARTEL, D. A.; PIRES, L. F. Enterprise interoperability with soa: a survey of service composition approaches. **Enterprise**

Distributed Object Computing Conference Workshops, 2008 12th, IEEE, p. 238–251, 2008.

PIMENTEL, M.; FUKS, H. **Sistemas colaborativos**, 2012.

POKRAEV, S. V. **Model-driven semantic integration of service-oriented applications**. Tese (Doutorado) — Novay, 2009.

RODRIGUES, S. A.; SILVA, T. S.; PIVOTTO, C.; GIRAO, A.; SOUZA, J. M. de; ROCHA, E. Integrating web and mobile knowledge management tools to improve negotiations. **Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on**, IEEE, p. 3035–3040, 2011.

RUNESON, P.; HOST, M.; RAINER, A.; REGNELL, B. **Case study research in software engineering: Guidelines and examples**, 2012.

SADI, M. H.; YU, E. Analyzing the evolution of software development: from creative chaos to software ecosystems. **Research Challenges in Information Science (RCIS), 2014 IEEE Eighth International Conference on**, IEEE, p. 1–11, 2014.

SANTOS, R.; WERNER, C.; TOSTES, L.; LIMA, T.; BARBOSA, G. Supporting negotiation and socialization for component markets in software ecosystems context. **Computing Conference (CLEI), 2016 XLII Latin American**, IEEE, p. 1–12, 2016.

SANTOS, R. P.; TOSTES, L. R.; WERNER, C. M. L. A brechó-ecosys extension to support negotiation in the software ecosystems context. **Information Reuse and Integration (IRI). IEEE 14th International Conference**, p. 578–585, 2013.

SANTOS, R. P. dos; WERNER, C. M. L. A proposal for software ecosystems engineering. **IWSECO@ ICSOB**, p. 40–51, 2011.

SANTOS, R. P. dos; WERNER, C. M. L. Treating social dimension in software ecosystems through reuseecos approach. **Digital Ecosystems Technologies (DEST), 2012 6th IEEE International Conference on**, IEEE, p. 1–6, 2012.

SCHOTS, M. On the use of visualization for supporting software reuse. In: **Companion Proceedings of the 36th International Conference on Software Engineering**,

2014. (ICSE Companion 2014), p. 694–697. ISBN 978-1-4503-2768-8. Disponível em: <<http://doi.acm.org/10.1145/2591062.2591095>>.

SILVA, R. A. C.; SOARES, L. S.; BRAGA, J. L. Workflow aplicado a engenharia de software baseada em processos: uma visão geral. **INFOCOMP Journal of Computer Science**, v. 5, n. 3, p. 76–84, 2006.

SIRQUEIRA, T. F.; DALPRA, H. L.; BRAGA, R.; ARAÚJO, M. A. P.; DAVID, J. M. N.; CAMPOS, F. E-seco proversion: An approach for scientific workflows maintenance and evolution. **Procedia Computer Science**, Elsevier, v. 100, p. 547–556, 2016.

STRÖELE, V.; CAMPOS, F.; DAVID, J. M. N.; BRAGA, R.; ABDALLA, A.; LANCELLOTTA, P. I.; ZIMBRÃO, G.; SOUZA, J. Data abstraction and centrality measures to scientific social network analysis. **21st International Conference on Computer Supported Cooperative Work in Design**, 2017.

SYCARA, K.; PAOLUCCI, M.; ANKOLEKAR, A.; SRINIVASAN, N. Automated discovery, interaction and composition of semantic web services. **Web Semantics: Science, Services and Agents on the World Wide Web**, v. 1, n. 1, 2011.

TENOPIR, C.; ALLARD, S.; DOUGLASS, K.; AYDINOGLU, A. U.; WU, L.; READ, E.; MANOFF, M.; FRAME, M. Data sharing by scientists: practices and perceptions. **PloS one**, Public Library of Science, v. 6, n. 6, 2011.

TENOPIR, C.; DALTON, E. D.; ALLARD, S.; FRAME, M.; PJESIVAC, I.; BIRCH, B.; POLLOCK, D.; DORSETT, K. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. **PLoS One**, Public Library of Science, v. 10, n. 8, 2015.

TRAVASSOS, G. H. et al. O enfoque de linha de produto para desenvolvimento de software. Sociedade Brasileira da Computação. ; Ingrid Jansch Porto. (Org.), XXI Jornada de Atualização em Informática – Livro Texto, SBC. v.2, p. 1–32, 2002.

USMAN, Z.; YOUNG, R. I. M.; CHUNGOORA, N.; PALMER, C.; CASE, K.; HARDING, J. A manufacturing core concepts ontology for product lifecycle interoperability. In: SPRINGER. **International IFIP Working Conference on Enterprise Interoperability**, 2011. p. 5–18.

WALI, B.; GIBAUD, B. Extending owl-s for the composition of web services generated with a legacy application wrapper. **ICIW2012 The Seventh International Conference on Internet and Web Applications and Services**. Stuttgart, Germany, p. 97, 2012.

WASSERMAN, S.; FAUST, K. **Social network analysis: Methods and applications**, 1994.

WILKINSON, M. D.; DUMONTIER, M.; AALBERSBERG, I. J.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J.-W.; SANTOS, L. B. da S.; BOURNE, P. E. et al. The fair guiding principles for scientific data management and stewardship. **Scientific data**, Nature Publishing Group, v. 3, p. 160018, 2016.

YIN, R. K. Estudo de caso-: Planejamento e métodos. Bookman editora, 5a. Ed., 2015.

Apêndice A - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Condutor do Estudo: Phillipe Israel Marques

Instituição: Universidade Federal de Juiz de Fora

Prezado(a) Senhor(a), a plataforma E-SECO utiliza a abordagem de Ecosistema de Software Científico (ECOSC) e busca auxiliar pesquisadores nas diversas etapas que compõem o ciclo de vida de um experimento científico. A partir disso, este trabalho busca apoiar a composição de serviços científicos, considerando que serviços estão presentes em todas as etapas que envolvem a experimentação. A plataforma E-SECO permite obter serviços de diferentes fontes e, a partir da composição, torna-se possível reutilizar e estender serviços existentes. Este trabalho apresenta um mecanismo que busca apoiar a composição de serviços em um ECOSC. Este estudo de caso busca avaliar como elementos de visualização de serviços e colaboração permitem apoiar a composição de serviços.

A.1 PROCEDIMENTOS

A plataforma E-SECO será utilizada para executar todas as etapas que compõem o processo de composição de serviços apresentado neste trabalho. Para participar deste estudo, solicitamos uma colaboração especial para: (i) permitir um estudo dos dados resultantes para a avaliação da proposta apresentada; (ii) participar da entrevista e/ou responder ao questionário. Para todos os coletados, serão retiradas informações pessoais, que não serão utilizadas em nenhum momento durante a análise ou apresentação dos resultados.

A.2 TRATAMENTO DE RISCOS

Todas as providências necessárias serão tomadas durante a coleta de dados, visando garantir a sua privacidade. Os dados coletados durante o estudo destinam-se apenas às atividades relacionadas com a solução proposta.

A.3 BENEFÍCIOS E CUSTOS

Espera-se que este estudo seja capaz de lhe fornecer novos conhecimentos sobre questões relacionadas à composição de serviços, bem como aspectos de visualização e colaboração. Este estudo visa contribuir com a área de Ecossistemas de Software, tratando questões técnicas e sociais.

A participação neste estudo não envolve nenhum gasto ou ônus. O(A) participante também não receberá qualquer espécie de reembolso ou gratificação devido à participação na pesquisa.

A.4 CONFIDENCIALIDADE DA PESQUISA

A informação coletada neste estudo é confidencial. As respostas informadas, bem como seus dados pessoais, tais como número de documentos e e-mail, não serão identificados de modo algum. Solicitamos apenas uma autorização para utilização de imagens geradas durante a realização do estudo de caso. Estas imagens podem incluir o nome e a fotografia do participante no espaço de trabalho compartilhado, o que pode permitir identificar os participantes. Para isto, solicitamos uma autorização para esse fim.

A.5 PARTICIPAÇÃO

Sua participação neste estudo é muito importante e voluntária. Você tem o direito de não querer participar ou de sair deste estudo a qualquer momento, sem penalidades. Em caso de você decidir se retirar do estudo, favor notificar um pesquisador responsável.

Os pesquisadores responsáveis por este estudo irão fornecer explicações sobre o estudo. Para isto, entre em contato com um dos pesquisadores a partir dos seguintes endereços de e-mails:

Phillipe Israel Marques - [phillipe.marques\(at\)gmail.com](mailto:phillipe.marques@gmail.com)

José Maria Nazar David - [jmndavid\(at\)gmail.com](mailto:jmndavid(at)gmail.com)

A.6 DECLARAÇÃO DE CONSENTIMENTO

Li as informações contidas neste documento antes de assinar este termo de consentimento. Declaro, para os devidos fins, que toda a linguagem técnica utilizada na descrição deste

estudo de pesquisa foi explicada satisfatoriamente e que recebi respostas para minhas dúvidas. Confirmo também que sou livre para me retirar do estudo em qualquer momento, sem qualquer penalidade. Declaro ter mais de 18 anos e dou meu consentimento de livre e espontânea vontade para participar deste estudo.

Local e Data

CPF do Participante

Nome do Participante

Assinatura do Participante

Apêndice B - QUESTIONÁRIO DE CARACTERIZAÇÃO

Dados Pessoais

Informe seus dados pessoais

Nome

E-mail

Atuação:

- Mercado
- Academia

Informações

1. Como você avalia seus conhecimentos em relação a Ecossistemas de Software?

- Muito baixo
- Baixo
- Moderado
- Bom
- Muito bom

2. Como você avalia seus conhecimentos em relação ao desenvolvimento de *web-services*?

- Muito baixo
- Baixo
- Moderado
- Bom
- Muito bom

3. Como você avalia seus conhecimentos em Anotação Semântica?

Muito baixo

Baixo

Moderado

Bom

Muito bom

4. Como você avalia seus conhecimentos em Proveniência de Dados?

Muito baixo

Baixo

Moderado

Bom

Muito bom

Apêndice C - QUESTIONÁRIO - ETAPA 1

Nome do Participante

Tendo como base a realização de uma busca por serviços e a composição de um serviço especificado previamente, responda às seguintes questões:

1. Caso você tenha identificado os dados de proveniência relacionados ao serviço, em qual experimento o serviço foi utilizado? Ele foi utilizado em algum *workflow*?

2. Em relação à utilização do serviço para análise de proveniência de dados, qual a relevância deste serviço ao analisar dados históricos para realizar novos experimentos?

- Muito baixo
- Baixo
- Moderado
- Bom
- Muito bom

3. Como você avalia o nível de dificuldade encontrado para realizar a composição do serviço especificado?

- Muito baixo
- Baixo
- Moderado
- Alto
- Muito alto

4. Como você avalia seu nível de compreensão em relação às atividades solicitadas?

- Muito baixo
- Baixo
- Moderado
- Alto
- Muito alto

5. Como você avalia sua satisfação ao utilizar o mecanismo solicitado na plataforma em relação à sua utilidade no futuro para apoiar a composição de serviços?

- Muito baixo
- Baixo
- Moderado
- Bom
- Muito bom

Em relação aos elementos de visualização e colaboração, responda às seguintes questões:

1. Com base no grafo de composições gerado, como você avalia a capacidade de identificar dependências entre serviços?

- Muito baixo
- Baixo
- Moderado
- Alta
- Muito alta

2. Você considera que as métricas utilizadas para identificar os serviços foram suficientes? Caso não as considere suficientes, poderia sugerir alguma?

3. Em relação à visualização de redes sociais, como você avalia a possibilidade de identificar relações de colaboração para um pesquisador específico?

- Muito baixo
- Baixo
- Moderado
- Alta
- Muito alta

4. Como você avalia a capacidade de identificar os tipos de relações de colaboração entre pesquisadores?

- Muito baixo

- Baixo
- Moderado
- Alta
- Muito alta

5. Como você avalia a capacidade de analisar a o nível de colaboração de um pesquisador específico na rede analisada?

- Muito baixo
- Baixo
- Moderado
- Bom
- Muito bom

Apêndice D - QUESTIONÁRIO - ETAPA 2

Nome do Participante

Em relação às interações na plataforma, responda às seguintes questões:

1. Em relação à visualização de serviços, como você avalia a capacidade de identificar as composições realizadas?

- Muito baixo
- Baixo
- Moderado
- Bom
- Muito bom

2. Como você avalia o apoio oferecido pelo serviço de colaboração utilizado para avaliar as composições realizadas?

- Muito baixo
- Baixo
- Moderado
- Bom
- Muito bom

3. Como você avalia seu nível de compreensão em relação às atividades solicitadas?

- Muito baixo
- Baixo
- Moderado
- Bom
- Muito bom

4. A partir dos serviços utilizados, como você avalia sua satisfação ao interagir na plataforma e visualizar as composições realizadas, em relação à utilidade dos serviços no futuro para apoiar a composição de serviços?

- Muito baixo

- Baixo
- Moderado
- Bom
- Muito bom

5. Você considera que serviço de colaboração utilizado foi suficiente para apoiar a avaliação da composição de serviços em um ecossistema de software científico? Justifique.

Com relação às dificuldades encontradas durante a realização do trabalho, responda:

1. Que problemas técnicos ocorreram durante a realização do trabalho?

2. Que problemas de comunicação ocorreram durante a realização do trabalho?

3. Que problemas de coordenação ocorreram durante a realização do trabalho? (Restrita ao coordenador)

4. Que problemas de cooperação ocorreram durante a realização do trabalho?

Em relação aos mecanismos de percepção, avalie as seguintes afirmações:

1. Os mecanismos de percepção disponíveis para a coordenação permitiram identificar os participantes para a avaliação (Restrita ao coordenador).

- Discordo Fortemente
- Discordo Parcialmente
- Indiferente
- Concordo Parcialmente
- Concordo Totalmente

2. Os mecanismos de percepção disponíveis para a comunicação permitiram identificar

os participantes através do chat.

- Discordo Fortemente
- Discordo Parcialmente
- Indiferente
- Concordo Parcialmente
- Concordo Totalmente

3. Os mecanismos de percepção disponíveis para a cooperação (tais como espessura das arestas, cor dos nós, tipos de arestas, informações nos nós) permitiram apoiar a interação entre os participantes.

- Discordo Fortemente
- Discordo Parcialmente
- Indiferente
- Concordo Parcialmente
- Concordo Totalmente

Apêndice E - ESPECIFICAÇÃO PARA O ESTUDO DE CASO

Caso de Uso 1:

Atores: desenvolvedor

Fluxo Básico:

Descrição: este caso se inicia quando um desenvolvedor realiza login na plataforma e seleciona a opção “Search Components”.

- Desenvolvedor realiza login na plataforma;
- Na página inicial, desenvolvedor clica em “Search” e em seguida, em “Search Components” para realizar a busca serviços em diferentes repositórios;
- Desenvolvedor informa dados no formulário de busca e clica no botão “Search” para listar os serviços encontrados;
- Ao identificar um serviço, o desenvolvedor clica em “More Details” para visualizar metadados sobre o serviço e cada uma de suas operações;
- Desenvolvedor clica em “Semantic Service Annotation” para listar operações do serviço;
- Desenvolvedor clica em “Generate OWL-S” para anotar semanticamente o serviço escolhido;
- Desenvolvedor realiza uma composição de serviços;
- Desenvolvedor seleciona “Search” e em seguida, seleciona “Service Manager” para realizar anotação semântica do serviço desenvolvido (**A1**).

Fluxo alternativo:

- [**A1**] Adicionar dados extras, se necessário: em “Search” e em seguida, “Additional Prime Data”, desenvolvedor adiciona outros dados que permitam realizar buscas por serviços em diferentes níveis de interoperabilidade.

Caso de Uso 2:**Atores:** desenvolvedor**Fluxo Básico:**

Descrição: este caso se inicia quando um desenvolvedor realiza login na plataforma e seleciona a opção “Service Composition”.

- Desenvolvedor realiza login na plataforma;
- Desenvolvedor seleciona botão “Search” e em seguida, “Service Composition”;
- Desenvolvedor busca serviços registrados na plataforma;
- Desenvolvedor informa parâmetros para busca e seleciona botão “Search”;
- Ao identificar o serviço composto criado, o desenvolvedor seleciona o botão “Dependencies” para visualizar serviços;
- Desenvolvedor seleciona aba “Social Network” e preenche dados solicitados para identificar as relações de um pesquisador específico.

Descrição das atividades

Tendo como base o Caso de Uso 1, realize uma busca por serviços e identifique os serviços que foram utilizados em experimentos anteriores. Faça a anotação semântica para um serviço que já foi utilizado anteriormente. Este serviço deve permitir que, a partir do nome de um gene, seja retornado seu identificador na base de dados. Este identificador deve ser informado no formato de uma única cadeia de caracteres (String), para isto é necessário realizar uma composição. Realize uma anotação semântica para o serviço composto desenvolvido também. Tendo como base o Caso de Uso 2, realize uma busca pelo serviço composto desenvolvido. Analise as dependências do serviço. Por fim, tente identificar as relações de colaboração de um pesquisador da universidade UFJF no ano de 2017.

Apêndice F - ESPECIFICAÇÃO PARA ESTUDO DE VIABILIDADE

Caso de Uso 1:

Atores: desenvolvedor, pesquisadores

Fluxo Básico:

Descrição: este caso se inicia quando um desenvolvedor identifica pesquisadores na rede social e seleciona a aba “Composition Evaluation”.

- Desenvolvedor preenche dados sobre a composição (nome e descrição), seleciona pesquisadores e digita uma mensagem aos pesquisadores. Em seguida clica sobre o botão “Create Discussion”;
- Desenvolvedor seleciona a aba “Services” e seleciona a discussão criada;
- Pesquisadores digitam mensagens e clicam sobre o botão “Send”;
- Desenvolvedor interage de maneira síncrona com pesquisadores descrevendo as funcionalidades do serviço;
- Pesquisadores avaliam se serviço desenvolvido atende à especificação.

Descrição das atividades

Desenvolvedor interage com pesquisadores, descrevendo a funcionalidade do serviço composto. Além disto, apresenta o ambiente disponível aos pesquisadores, utilizando a visualização para permitir identificar serviços capazes de substituir um serviço que possa falhar. A partir do espaço de trabalho compartilhado e do serviço de comunicação, os participantes interagem analisando como os serviços se relacionam. A partir das mensagens trocadas, os pesquisadores devem avaliar se o serviço composto atende a especificação apresentada pelo desenvolvedor, além de avaliar também a possibilidade de identificar relações de dependência entre serviços.