

Um Sistema de Informação para Gerência de Projetos Científicos baseados em Simulações Computacionais

Alternative Title: An IS for Managing Scientific Projects

Leonardo S. Ramos
Instituto de Computação/Universidade
Federal Fluminense – IC/UFF
Niterói – Brasil
leonardosl@id.uff.br

Kary A. C. S. Ocaña
Laboratório Nacional de Computação
Científica - LNCC
Petrópolis - Brasil
karyann@lncc.br

Daniel de Oliveira
Instituto de Computação/Universidade
Federal Fluminense – IC/UFF
Niterói - Brasil
danielcmo@ic.uff.br

RESUMO

Gerenciar projetos científicos é uma tarefa complexa. Um mesmo projeto pode estar associado a vários experimentos científicos diferentes que por sua vez requerem execuções de diferentes simulações computacionais (*i.e. workflows* científicos). Essa gerência ainda se torna mais complexa se considerarmos que as tarefas do projeto devem estar associadas a execução de tais simulações (que podem demorar dias ou semanas) e que a equipe do projeto pode se encontrar geograficamente dispersa. Esse artigo apresenta o sistema de informação SciManager que tem como objetivo auxiliar os cientistas a gerenciar seus projetos científicos. O SciManager apoia a gerência do projeto, seus experimentos associados e *workflows* em uma única ferramenta, fazendo com que todas as informações referentes ao projeto estejam associadas. O SciManager é baseado em uma arquitetura de nuvem, o que faz com que esteja disponível para membros do projeto. Foi realizada uma avaliação experimental e aproximadamente 88% dos usuários concordaram que o SciManager se mostra útil e fácil de usar no dia-a-dia do cientista.

Palavras-Chave

Projeto científico, experimento científico, *workflow* científico.

ABSTRACT

Managing scientific projects is a complex task. The project may be associated with several different scientific experiments that in turn require implementations of different computer simulations (scientific *workflows*). This management becomes even more complex if we consider that the project tasks should be associated with the specification and execution of these simulations (which can take days or weeks to finish) and that the project's team can be geographically dispersed. This paper presents the SciManager information system that aims at helping scientists managing scientific projects. The SciManager is able to manage the project, its associated experiments and *workflows* in a single tool, making all information related. The SciManager is based on a cloud architecture, which means it is easily available for project members. An experimental evaluation was held out and approximately 88% of users agreed that SciManager is useful and easy to use in scientist's daily duties.

Categories and Subject Descriptors

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SBSI 2016, May 17–20, 2016, Florianópolis, Santa Catarina, Brazil.
Copyright SBC 2016.

K.4.3 [Organizational Impacts]: Automation, Computer supported collaborative work.

General Terms

Management, Experimentation, Verification.

Keywords

Scientific Project, scientific experiment, scientific workflow

1. INTRODUÇÃO

Nas últimas décadas a utilização de simulações computacionais para apoiar experimentos em diversos domínios da ciência se tornou uma realidade [1]. Tais experimentos seguem um ciclo de vida bem definido e são comumente compostos pela invocação de diversos programas de forma encadeada, de acordo com a sua produção e o seu consumo de dados, formando um fluxo coerente. Modelar, executar e analisar o resultado de um experimento científico não é uma tarefa trivial de ser realizada, especialmente em experimentos computacionalmente intensivos ou que manipulem um grande volume de dados, complexos e não estruturados, (*i.e. big data*) [2]. Neste cenário, experimentos científicos podem ser apoiados pelos *workflows* científicos [1]. Um *workflow* é uma abstração que modela o encadeamento de atividades (invocações de programas) e dados e são executados por Sistemas de Gerência de *Workflow* Científicos (SGWfC) [1].

Existem diversos SGWfC [1], cada um deles com suas próprias características, vantagens e desvantagens. Os SGWfC possuem um motor de execução acoplado, responsável pela gerência e execução das atividades do *workflow*. Estes mesmos SGWfCs também realizam a coleta de dados de proveniência [3] do experimento, *i.e.* desde a captura de tempos de execução até informações mais complexas como as de domínio, específicas de uma área de pesquisa. Os dados de proveniência descrevem o histórico da execução de um experimento e são fundamentais para a reprodução e/ou a análise dos resultados.

Apesar do valioso auxílio provido pelos SGWfC, a gerência de um experimento vai além da execução de um único *workflow* científico de forma isolada. Experimentos científicos complexos e de larga escala são compostos por mais de um tipo de análise, o que faz com que seja necessário especificar mais de um *workflow* em um mesmo experimento. Além disso, diferentemente de *workflows* de negócio, a estrutura dos *workflows* científicos não é “rígida e fixa” e pode ser modificada ao longo do experimento, *i.e.* um mesmo *workflow* pode ser remodelado e reexecutado *n* vezes ao longo do curso de execução de um experimento.

Para tornar o cenário ainda mais complexo, experimentos de larga escala são parte integrante de projetos científicos que podem ser

conduzidos por equipes geograficamente dispersas. Um projeto científico inclui não somente a modelagem, execução e a análise dos dados de simulação, mas também toda a parte “burocrática” que envolve relatórios de andamento, análise de estatísticas, atribuição de responsabilidades, entre outros. Tomemos como exemplo o *Leishmaniasis Virtual Laboratory*¹ parte do projeto *EU Brazil Cloud Connect*. Esse projeto conta com pesquisadores da FIOCRUZ e PUC-Rio do Brasil, da *University of Newcastle* na Inglaterra, da *Universidad de Valencia* na Espanha, entre outras instituições. Neste laboratório virtual, o foco é a pesquisa molecular da *Leishmania* e seus vetores, o que pode envolver vários experimentos diferentes desde análise filogenética até análise molecular, cada qual com seus respectivos *workflows* associados. No contexto do projeto, as tarefas relativas ao desenvolvimento, monitoramento, análise de resultados, documentação e auditoria devem ser distribuídas entre os participantes que atuam como “membros” do projeto. Já os chefes de laboratórios e coordenadores do projeto atuam como “gerentes” que devem estar cientes do andamento do projeto como um todo. Desta forma, podemos perceber a complexidade de gerenciar um projeto científico em uma escala global.

Entretanto, gerenciar projetos complexos não é uma exclusividade da área científica. Problemas semelhantes já são enfrentados há anos na área comercial, e a área de Gerência de Projetos [4] tem como objetivo tratar tais problemas. Um exemplo da área comercial, no contexto de projetos de *software*, foi o desenvolvimento do Windows 7 que contou com mais de 2.000 funcionários espalhados por todo o mundo². Na gerência de projetos assume-se a premissa de que a qualidade do produto final é influenciada pela qualidade do processo que levou a geração de tal produto. Desta forma, se a condução de um projeto é bem gerenciada o produto final tem mais chances de ter uma qualidade aceitável para determinado fim. Podemos aplicar a mesma premissa no contexto de projetos científicos. Isto é: “se um projeto possui um processo devidamente especificado e controlado; for executado seguindo procedimentos definidos; a análise dos dados segue um protocolo; e a documentação foi realizada então o produto final (as conclusões acerca da hipótese científica) terá uma garantia de qualidade maior.”

Porém, mesmo que o processo esteja bem definido, a gerência de um projeto científico possui algumas especificidades e desafios. Entre elas: (i) as tarefas associadas a um experimento do projeto e atribuídas aos membros do projeto devem seguir o ciclo de vida do experimento científico definido por Mattoso *et al.* [5]; (ii) as tarefas geradas no contexto do projeto devem poder ser associadas às tarefas de *workflows* a serem executadas por SGWfCs; (iii) a dificuldade de executar e analisar experimentos com equipes dispersas geograficamente; (iv) a falta de experiência de pesquisadores, das diversas ciências, sobre como gerenciar um projeto seguindo as técnicas da área de gerência de projetos; (v) a dificuldade de monitorar tarefas que podem durar semanas/meses; e (vi) a dificuldade de lidar com um grande volume de dados produzidos e/ou disperso geograficamente.

Conforme ao exposto, faz-se necessária uma ferramenta que auxilie o cientista na gerência do projeto científico como um todo e não só particularmente na execução de um determinado

experimento ou *workflow* (para tanto já existem os SGWfCs). Por tanto, se torna interessante que possamos nos beneficiar de todo o arcabouço de gerência de projeto já existente na Engenharia de *Software* e em outras áreas. Entretanto e especificamente para a gerência de projetos científicos, ainda existe uma escassez de abordagens que gerenciem esse tipo de projeto baseado em técnicas de gerência de projetos, o que representa um desafio em aberto. Desta forma, esse artigo propõe o sistema de informação chamado SciManager que visa apoiar a gerência de projetos científicos em larga escala com a colaboração entre equipes geograficamente dispersas. O SciManager é baseado em técnicas e ferramentas existentes de gerência de projetos e possui uma arquitetura baseada no conceito de computação em nuvem. A plataforma de nuvem escolhida foi a Amazon EC2.

O SciManager apoia as etapas do ciclo de vida do experimento científico [5] da seguinte forma: (i) na composição: possibilita aos “gerentes do projeto” atribuir tarefas de especificação de *workflows* do experimento e aos “membros do projeto” realizar esboços e *uploads* de arquivos de configuração/especificação na ferramenta; (ii) na execução: possibilita a geração, execução e monitoramento de tarefas de *workflows* científicos pelo SGWfC acoplado ao SciManager (nesse artigo utilizamos o SciCumulus [6]) e: (iii) na análise: possibilita importar dados de proveniência do *workflow* e do domínio no apoio à análise de dados de um experimento científico.

O restante desse artigo se encontra organizado em 5 seções além desta introdução. Na Seção 2 é apresentado o referencial teórico. Na Seção 3 discutimos e apresentamos o SciManager. A Seção 4 apresenta a avaliação experimental. A Seção 5 discute os trabalhos relacionados e finalmente a Seção 6 conclui esse artigo.

2. REFERENCIAL TEÓRICO

2.1 Projeto, Experimento e *Workflow* Científicos

No contexto desse artigo existem conceitos importantes a serem explicados, porém três deles são fundamentais: projetos científicos, experimentos científicos e *workflows* científicos. Um projeto científico é a unidade de trabalho de mais alto nível e tem como objetivo principal funcionar como um roteiro de trabalho ou instrumento de planejamento, além de ser o elemento direcionador da pesquisa. Desta forma, um projeto científico deve definir qual(is) o(s) objetivo(s) da pesquisa e como a mesma será conduzida. Cada projeto conta com um grupo de pessoas para as quais as tarefas e responsabilidades relativas ao projeto serão atribuídas. Dependendo da quantidade de objetivos a ser alcançada, um projeto científico pode ser composto por um ou mais experimentos científicos.

Podemos definir um experimento científico como “um teste executado sob condições controladas, que é realizado para demonstrar uma verdade conhecida, examinar a validade de uma hipótese, ou determinar a eficácia de algo previamente não explorado” [7]. Um experimento científico está estritamente associado a um conjunto de ações controladas (*i.e.* um protocolo) com um objetivo bem definido. Essas ações incluem variações de testes, e seus resultados são geralmente comparados entre si para aceitar ou refutar uma hipótese científica [5]. Existem diversos tipos de experimentos científicos, a saber: *in vivo*, *in vitro*, *in virtuo* e *in silico* [8]. Neste artigo temos como foco apoiar os experimentos *in silico*, *i.e.* aqueles que são baseados em simulações computacionais.

¹ <http://eubrazilcloudconnect.eu/content/leishmaniasis-virtual-laboratory>

² <http://www.computerworld.com/article/2532600/operating-systems/microsoft-may-have-2-000-developers-working-on-windows-7.html>

Geralmente esses experimentos científicos representam o encadeamento e execução de diferentes programas, que podem consumir múltiplas combinações de parâmetros e grandes quantidades de dados. Assim, experimentos científicos podem ser modelados como *workflows* científicos. *Workflows* representam esse encadeamento por meio de artefatos executáveis e que representam uma das possíveis variações do experimento.

Dessa forma, temos a seguinte interconexão entre os conceitos de projeto científico, experimento científico e *workflow*. Um *workflow* científico faz parte de um determinado experimento, que por sua vez se encontra no contexto de um projeto científico. Um *workflow* é a representação concreta de um experimento científico e representa um dos possíveis ensaios do experimento (*i.e. trials*). Já cada experimento segue um ciclo de vida bem definido que envolve as fases definidas por Mattoso *et al.* [5], apresentado na Seção 2.2. Além disso, apresentamos na Seção 2.3 que o próprio projeto possui um ciclo de vida próprio independente do ciclo de vida do experimento.

2.2 Ciclo de Vida do Experimento Científico

O ciclo de vida de um experimento científico em larga escala, segundo Mattoso *et al.* [5] consiste basicamente em múltiplas interações, a serem realizadas por cientistas, durante o curso de um experimento. Uma vez que podem existir diversos experimentos coexistindo dentro de um mesmo projeto científico, podemos ter mais de um ciclo de vida, superpostos ou interagindo entre si. Na Figura 1 apresentamos uma versão simplificada do ciclo proposto por Mattoso *et al.* [5] onde as principais fases podem ser identificadas: a execução, a composição e a análise.

Na fase de composição, os cientistas definem a estrutura do experimento científico *i.e.* a sequência lógica das atividades, os tipos de dados de entrada/saída e parâmetros que devem ser fornecidos. A fase de execução é responsável por executar uma especificação de *workflow* (*trial*) em um determinado SGWFC ou máquina de execução. Finalmente, a fase de análise apoia a interpretação e avaliação dos dados gerados pelas fases de composição e execução. Esta fase é altamente dependente de dados de proveniência [3] do experimento que foram gerados nas fases anteriores. Os dados de proveniência registram o histórico do *workflow*, desde sua especificação, os parâmetros utilizados, até os tempos de execução de cada atividade do mesmo. Com esses dados os cientistas são capazes de auditar execuções e garantir a reprodutibilidade do mesmo.

2.3 Ciclo de Vida do Projeto Científico

Da mesma forma em que o ciclo de vida do experimento apresenta fases e características definidas, o ciclo de vida de um projeto científico (quando baseado nas técnicas e metodologias de gerência de projetos) consiste em um conjunto de fases que são executadas pelos membros do projeto, conforme apresentado na Figura 2. Cada uma das fases apresentadas na Figura 2 possui diversas atividades realizadas por membros do projeto. Essas atividades são consideradas como “finalizadas” quando é feita a entrega dos produtos, comumente chamados de “entregáveis” (do inglês *deliverables*).

Na primeira fase deste ciclo, chamada de “Início do Projeto”, são discutidos os objetivos do projeto, quais metodologias serão utilizadas e quem serão os membros (cientistas) e os gerentes (chefes de laboratório). Todas essas informações são requisitos de mais alto nível do projeto. Como resultado desta fase é produzido um documento que representa um termo de abertura do projeto. O termo de abertura contém explicitamente os objetivos, os componentes da equipe, verba disponível, expectativas de HH do

projeto, *etc.* Até que o termo de abertura do projeto se encontre devidamente finalizado, várias interações podem se fazer necessárias.

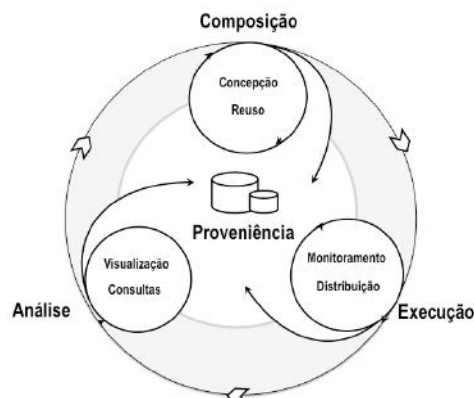


Figura 1 Ciclo de vida do experimento científico adaptado de Mattoso *et al* [5].

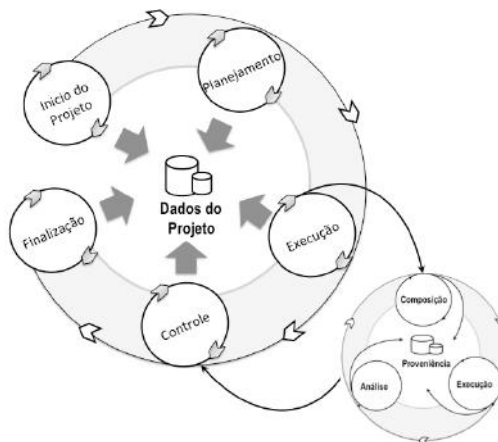


Figura 2 Ciclo de vida de um Projeto Científico

A segunda fase, chamada de “planejamento” do projeto, discute sobre os requisitos de alto nível levantados na fase inicial a fim de se criar requisitos mais específicos que guiam a criação de tarefas. Uma vez definidos os requisitos específicos, uma ou mais tarefas são criadas e atribuídas aos membros do projeto. Por exemplo, em um projeto de bioinformática um requisito de alto nível é “Identificar fármacos para combater a Malária”. Esse objetivo de alto nível pode ser detalhado em objetivos específicos como “Execução de Análise Filogenética” e “Estudo de Modelagem Molecular”. Cada um desses objetivos mais específicos deve estar associado a um experimento e uma ou mais tarefas podem ser criadas associadas ao objetivo. Por exemplo, para “Execução de Análise Filogenética” podemos criar tarefas intituladas “Estudo do Programa ModelGenerator”, “Criação do Esboço do *Workflow* de Análise Filogenética”, “Seleção dos Dados de Entrada”, *etc.* Cada uma dessas tarefas está associada a um experimento e um *workflow* e deve ser atribuída a um (ou mais) membro(s) do projeto.

A terceira fase é a “execução” do projeto. Nessa fase, as tarefas especificadas na fase anterior são efetivamente executadas; os *workflows* são implementados/testados/executados, e os dados são analisados, a fim de gerar as conclusões relativas do experimento sendo executado. A quarta fase é a de controle e monitoramento, nela as tarefas executadas na fase de execução são verificadas e

auditadas pelos gerentes do projeto. Uma vez que as tarefas tenham sido executadas, o projeto passa para a última quinta fase de finalização. Nessa fase é obtido um documento final que apresenta tanto os resultados gerados como algumas estatísticas do projeto.

Entre uma fase e outra do projeto são realizadas reuniões entre os membros do projeto, com o objetivo de verificar o andamento e/ou conclusão do mesmo. Essas reuniões determinam se o projeto deve continuar para sua próxima fase e ajudam os cientistas a detectar e/ou corrigir erros que afetem o andamento do projeto.

3. O SCIMANAGER

Nessa seção apresentamos a ferramenta proposta SciManager. O SciManager é de código aberto e pode ser obtido diretamente no repositório do projeto no BitBucket³.

3.1 Arquitetura do Sistema

O SciManager foi desenvolvido na linguagem Java e seguindo o padrão arquitetural MVC (*Model-View-Controller*) por meio do arcabouço Spring Boot/Spring MVC que já oferece apoio para desenvolvimento de aplicações nesse padrão. Para o desenvolvimento das interfaces, *templates* e telas do SciManager foi utilizado o arcabouço Thymeleaf. A vantagem de utilizar o Thymeleaf é que ele implementa o conceito de modelos naturais, *i.e.* arquivos de modelo que podem ser carregados diretamente em navegadores *Web*, substituindo dessa forma o JSP (*Java Server Pages*) como páginas da *Web*. Foi utilizado o arcabouço Spring *Security* para autenticação e segurança dos dados.

Como o SciManager é uma aplicação *Web*, o mesmo deve ser hospedado em um servidor de aplicação (parte 2 na Figura 3). O servidor escolhido foi o Apache Tomcat e o mesmo foi instalado e configurado em um servidor no ambiente de nuvem Amazon EC2. Uma máquina virtual do tipo *m4.large* foi criada para hospedar a aplicação. Dessa forma, como o provedor de nuvem garante disponibilidade de 99,9%, garantimos que a aplicação esteja sempre disponível e possa ser facilmente acessada pela internet por meio de navegadores *Web* comuns como o Chrome e o Internet Explorer (parte 1 na Figura 3).

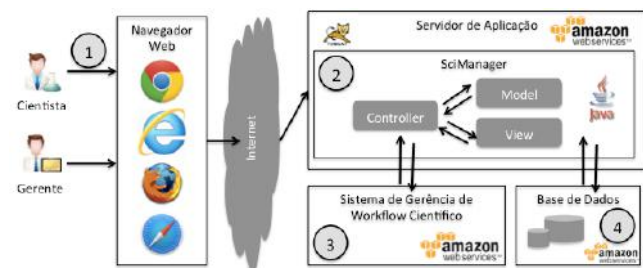


Figura 3 Arquitetura do SciManager.

Na aplicação *Web* é onde os cientistas registram as tarefas do projeto bem como o gerente acompanha o andamento das mesmas. Porém, conforme mencionado anteriormente, as tarefas do projeto (conforme será melhor explicado a seguir) necessitam ter uma interligação com os *workflows* que serão executados nos SGWfCs. Assim, para que mantivéssemos todos os componentes na nuvem, escolhemos o SGWfC SciCumulus que é capaz de executar atividades do *workflow* na nuvem (parte 3 na Figura 3), instanciando um *cluster* de máquinas virtuais. Além disso, o banco de dados que contém os resultados do experimento e os

dados relativos ao acompanhamento do projeto também foi hospedado em um servidor PostgreSQL na nuvem (parte 4 na Figura 3). Dessa forma, o cientista não necessita instalar nenhuma aplicação em seu computador, podendo acessar qualquer uma das funcionalidades remotamente.

3.2 Funcionalidades da Ferramenta

3.2.1 Gerência de Usuários e Grupos de Usuários

O SciManager foi projetado para atender projetos científicos com múltiplos usuários geograficamente dispersos. Assim, se faz necessário um controle desses usuários na ferramenta. Esse controle de usuários é realizado em dois níveis. No nível do “Perfil Pessoal”, cada usuário da ferramenta configura/preenche suas informações pessoais. Em seu perfil, o usuário define informações básicas como nome, *e-mail* de contato, instituição e, opcionalmente, pode realizar o *upload* de uma foto. Ao se visualizar o perfil de um usuário também temos acesso aos seus *workflows* e tarefas associados (Figura 4)



Figura 4 Perfil pessoal do cientista no SciManager

No nível de “Grupos de usuários”, o(s) usuário(s) administrador(es) (com privilégios para criar grupos) procuram, selecionam e agrupam (Figura 5) perfis de usuários existentes, e que a partir da criação do grupo irão partilhar os mesmos privilégios na ferramenta.



Figura 5 Gerência de grupos de usuários.

3.2.2 Gerência de Projetos, Experimentos e Workflows

Após definir os perfis e grupos de usuários, as funcionalidades-chave da ferramenta são habilitadas. Assim, a primeira tarefa que o usuário pode realizar é cadastrar um projeto científico no sistema que é a unidade de trabalho de mais alto nível, conforme apresentado na Figura 6.

Um projeto possui um nome, uma descrição, uma fase (que segue o ciclo de vida do projeto apresentado na Figura 2) e o grupo de

³ https://bitbucket.org/danielcmo/scientific_project_management

usuários que estão associados àquele projeto. Nele, os administradores, além do registrar o projeto no sistema, podem editá-lo e modificar a fase em que um projeto se encontra.

Já os cientistas membros (usuários), tem a capacidade de criar experimentos a serem associados a um determinado projeto. Além disso, eles podem criar/editar/incluir uma documentação levantando informações importantes desse experimento/projeto no próprio sistema. A Figura 7 apresenta a funcionalidade de Wiki no apoio à atualização e documentação do projeto. Por razões de espaço, a tela de cadastro de experimentos não é exibida, porém nela constam áreas para definir um nome do experimento, a hipótese científica associada e a fase em que o mesmo se encontra (que segue o ciclo de vida apresentado na Figura 1).

#	Projeto científico	Grupos administradores
#	Análise Filogenética do Plasmodium Falciparum	Análise Filogenética
#	Análise Filogenética do Plasmodium Vivax	Análise Filogenética
#	Genômica Comparativa (Malária)	Genômica Comparativa
#	Modelagem Molecular Cisteína Protease	Genômica Comparativa

Figura 6 Gerência dos projetos no SciManager

Uma vez que o experimento esteja devidamente definido no projeto no SciManager, os usuários estão aptos a cadastrar os workflows associados ao experimentos (Figura 8). No cadastro, cada workflow deve possuir um nome, o SGWfC que o executa, o projeto e experimento aos quais está associado e o usuário/grupo responsável (Figura 8). Como o workflow é o artefato central de um projeto científico, mais funcionalidades foram definidas em torno do mesmo. Além de registrar um workflow, o sistema possibilita que o usuário especifique um workflow a ser executado em um SGWfC existente conforme apresentado na Figura 9. É importante ressaltar que um membro do projeto só está apto a especificar um workflow uma vez que uma tarefa tenha sido atribuída a ele (conforme apresentado na seção 3.3.3).

Figura 7 Wiki para Documentação do Projeto.

#	Nome do workflow	SGWfC	Projeto científico
#	SciHm	SciCumulus	Genômica Comparativa (Malária)
#	SciPhy - Falciparum	SciCumulus	Análise Filogenética do Plasmodium Falciparum
#	SciPhy - Vivax	SciCumulus	Análise Filogenética do Plasmodium Vivax
#	SciPPgX	SciCumulus	Modelagem Molecular Cisteína Protease

Figura 8 Gerência dos workflows no SciManager.

Após o cadastro, o usuário se encontra apto a especificar um workflow a partir de um modelo “em branco” (Figura 9) ou importar workflows já anteriormente cadastrados (Figura 10).

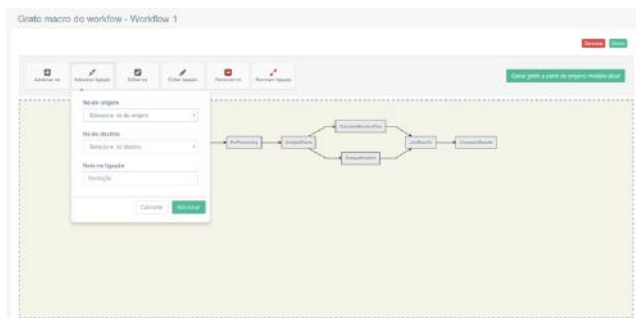


Figura 9 Especificação do Workflow

No caso de realizar a importação, o sistema gera o grafo associado para visualização e exibe um diff sintático da especificação submetida no momento com a última versão da especificação do workflow que se encontra registrada no sistema (Figura 10). O sistema apresenta uma listagem de todas as versões dos workflows especificadas e cadastradas previamente no sistema. Somente uma das versões pode ser considerada “ativa” no sistema, as demais são categorizadas como “inativas”, porém continuam sendo registradas para caso o usuário sinta a necessidade de retornar a uma versão anterior, como ocorre em sistemas de controle de versão como o Git e o SVN.

Figura 10 Importação de Workflow e diff sintático.

3.2.3 Gerência de Tarefas no Contexto do Projeto

Uma vez que o projeto, o experimento e os workflows foram definidos, os gerentes do projeto estão aptos a criar as tarefas do projeto e atribuí-las aos membros de cada projeto. A Figura 11

apresenta a tela de criação de tarefas no SciManager. Cada tarefa possui um título, uma descrição, o número de dias estimados para a sua execução, uma data limite para a tarefa, a qual fase do projeto a mesma está associada (*i.e.* a tarefa só estará habilitada para execução quando o projeto atingir tal fase), o *workflow* ao qual a tarefa está associada (opcional), quem é(são) o(s) responsável(is) e uma *tag* que pode ser usada para auxiliar nos filtros por tarefas.

Figura 11 Criação de uma Tarefa de um Projeto no SciManager

Uma vez que todas as tarefas são cadastradas, o sistema gera um quadro Kanban (Figura 12), onde cartões (*i.e.* *cards*) que representam as tarefas associadas a um projeto seguem um fluxo pré-estabelecido de estágios (*a fazer, Em progresso e feito*). Na medida em que o estado das tarefas vai evoluindo, os cartões mudam de estágio no quadro, e sempre que uma nova tarefa é identificada, um novo cartão é criado. Nesse quadro, os membros do projeto tem como filtrar quais tarefas pertencem a um determinado experimento ou *workflow*. Além disso, uma barra com colorações diferentes indica a qual fase do ciclo de vida do experimento e do projeto a tarefa está associada. Por meio do quadro Kanban, os gerentes estão aptos a ter uma visão geral do projeto e tomar ações corretivas e/ou evolutivas. Na Seção 4 é apresentada a avaliação realizada com o sistema proposto.

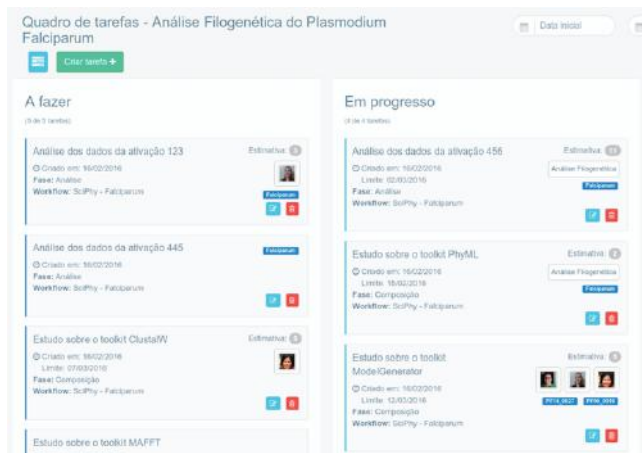


Figura 12 Quadro Kanban de Tarefas do Projeto

4. AVALIAÇÃO EXPERIMENTAL

De forma a avaliar a abordagem proposta utilizamos o modelo de avaliação denominado TAM (*Technology Acceptance Model*) [9]. Este modelo já foi utilizado em trabalhos semelhantes como proposto por De Souza *et al.* [10]. O experimento proposto por De Souza *et al.* [10] foi a base para a avaliação utilizada nesse artigo. A ideia principal por trás do TAM é avaliar a receptividade/comportamento de um usuário no que se refere a facilidade e utilidade da tecnologia/ferramenta que está sendo proposta, neste caso o SciManager. A *utilidade* refere-se ao quanto o usuário acredita que a abordagem proposta o auxiliará em suas tarefas e a *facilidade* se refere ao quão fácil/simples será utilizar tal abordagem.

Dessa forma, para realizar a avaliação, o TAM sugere a criação de questionários cujas perguntas estejam relacionadas a facilidade e a utilidade da abordagem proposta. Seguindo uma escala de Likert, para cada pergunta do questionário, o usuário pode selecionar uma e somente uma das opções a seguir: (a) Discordo Totalmente, (b) Discordo Parcialmente, (c) Neutro, (d) Concordo Parcialmente e (e) Concordo Totalmente. A Tabela 1 apresenta as questões utilizadas para a avaliação do SciManager. As questões cujo identificador se inicia com a letra T se referem a avaliação do treinamento. As questões cujo identificador se inicia com a letra F se referem a avaliação da facilidade e os demais se referem a avaliação da utilidade.

Tabela 1 Questões utilizadas na avaliação do SciManager

ID	Questão
T1	O treinamento oferecido pelos avaliadores foi completo?
T2	O treinamento me ofereceu subsídios para poder utilizar o SciManager?
T3	O treinamento foi adequado em relação ao tempo e ao detalhamento?
F1	Você gostou de utilizar o SciManager?
F2	O acesso ao SciManager é simples?
F3	Utilizar o SciManager é uma boa ideia?
F4	As funcionalidades no SciManager são simples de serem compreendidas?
F5	É fácil encontrar a informação que desejo no SciManager?
F6	O SciManager possui uma interface atraente e agradável?
F7	Mesmo antes de clicar em um funcionalidade você já conseguia prever o que iria acontecer?
U1	O uso do SciManager agrega valor aos experimentos que eu executo?
U2	Utilizar o SciManager é útil para ensinar novatos a trabalhar em projeto?
U3	Utilizar o SciManager pode melhorar o desempenho do meu projeto científico?
U4	O uso do SciManager pode facilitar meu trabalho?
U5	Eu pretendo integrar o uso do SciManager na minha rotina de trabalho?
U6	Eu recomendaria o SciManager?
U7	Os conceitos de experimentação e projeto científico foram abordados por completo no SciManager?

Para poder avaliar o SciManager, cada um dos usuários escolhidos passou por um treinamento simples e presencial de aproximadamente 1 hora. A avaliação do SciManager contou com a participação de 9 avaliadores, sendo 8 deles estudantes de graduação, mestrado e doutorado da Universidade Federal Fluminense (UFF) e 1 pesquisador do LNCC com experiência na área de bioinformática. Do total de alunos avaliadores, 4 eram alunos de graduação e 4 de pós-graduação. Os alunos de pós-

graduação já se encontram familiarizados com o ambiente de pesquisa e o funcionamento de um projeto científico. Os alunos de graduação apesar de não terem tanta experiência cursaram a disciplina de *e-Science* oferecida como eletiva para a graduação e se encontram familiarizados com os conceitos envolvidos no SciManager. Os projetos e experimentos utilizados na avaliação são alguns já existentes e cujos avaliadores já se encontravam familiarizados com o conteúdo, a saber o experimento de análise filogenética executado pelo *workflow* SciPhy [11].

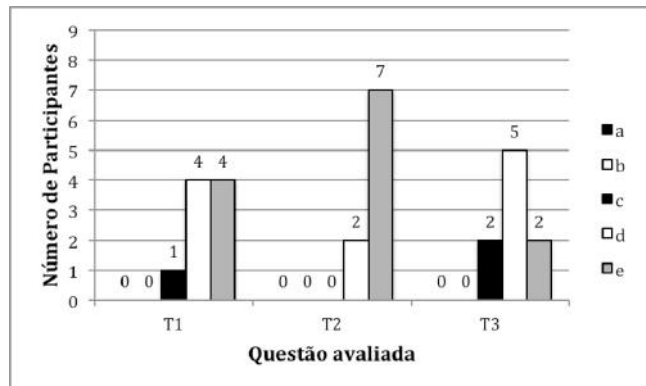


Figura 13 Resultado da avaliação do treinamento

A Figura 13 apresenta os resultados relativos ao treinamento oferecido. Em relação a completude do treinamento, 88% dos avaliadores consideraram o treinamento completo. Apenas um avaliador queixou-se de que nem todos os conceitos de experimentação científica foram explicados. Infelizmente devido ao curto espaço de tempo da sessão de treinamento, alguns conceitos não puderam ser explicados com detalhes. Em relação ao subsídio necessário para se utilizar a ferramenta, todos os avaliadores consideraram que o treinamento ofereceu tais subsídios de forma satisfatória. Em relação a duração do treinamento, 77% dos avaliadores consideraram o tempo satisfatório e dois deles mencionaram que 1 hora não é suficiente para apresentar todas as funcionalidades da ferramenta. Apesar das críticas, consideramos que o treinamento foi satisfatório e procuraremos melhorá-lo em próximas apresentações do SciManager.

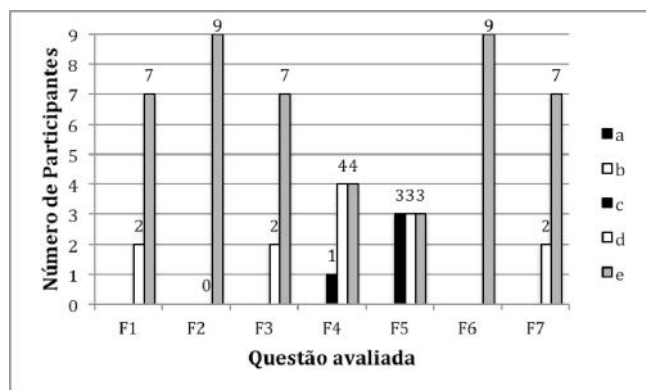


Figura 14 Resultado da avaliação da facilidade de uso

A Figura 14 apresenta os resultados relativos a facilidade de uso percebida com o SciManager. Em relação a facilidade de acesso e de uso do SciManager, 100% dos avaliadores concordaram (totalmente ou parcialmente) que o sistema é de fácil acesso e fácil de ser utilizado. O mesmo aconteceu na avaliação da interface em que 100% dos avaliadores concordaram que a interface é amigável e atraente. As únicas perguntas em alguns

dos avaliadores se mantiveram neutros foram as perguntas F4 e F5, onde avaliamos a simplicidade de compreensão das funcionalidades e a facilidade de se encontrar informação, respectivamente. Na pergunta F4, um usuário se manteve neutro e alegou que alguns títulos de menus não estavam transparentes, o que já foi modificado na ferramenta e será liberado em versões futuras. Em relação a pergunta F5, 33% dos avaliadores alegaram que mais filtros se fazem necessários no sistema, pois quando a quantidade de informações crescer se tornará inviável buscar *workflows* e tarefas nas listas geradas.

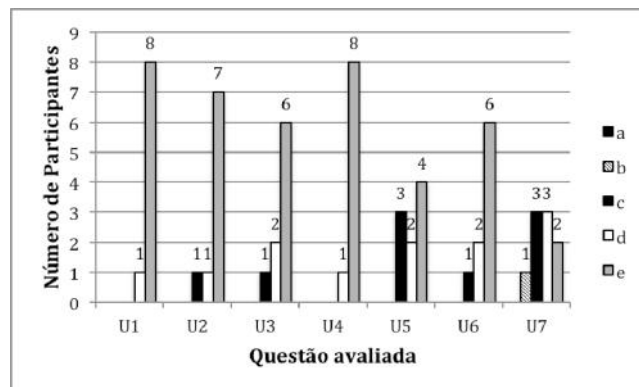


Figura 15 Resultado da avaliação da utilidade

A Figura 15 apresenta os resultados relativos a utilidade do SciManager. Nas perguntas em que se verificava se o uso do SciManager agrega valor ao trabalho e se o uso do SciManager facilita o trabalho, 100% dos avaliadores concordaram. Em relação ao uso do SciManager com novatos, 88% dos avaliadores concordaram que utilizar a ferramenta ajuda membros novatos do projeto a entender o processo. Apenas um avaliador mencionou que pode se tornar um empecilho se o treinamento não for bem feito, o que corrobora a crítica em relação ao tempo de treinamento. Em relação a verificação se o uso do SciManager pode melhorar o desempenho do serviço, também 88% dos avaliadores concordaram e um dos avaliadores mencionou que mensurar a melhora do desempenho pode ser complexo, pois dependendo do problema, as tarefas são bastante demoradas, o que faz com que a contribuição do SciManager seja difícil de ser mensurada. Na questão U5 que trata da incorporação do SciManager no trabalho, 33% dos avaliadores se mantiveram neutros, pois alegaram que inserir uma ferramenta de gerência em projetos em andamento seria demasiadamente complexo, porém consideraram o uso em projetos futuros. Na questão U6, 88% dos avaliadores recomendariam o SciManager e o avaliador que se manteve neutro alegou que precisaria de mais tempo para estudar e experimentar a tecnologia. Na questão U7, 22% dos avaliadores se mantiveram neutros e 11% discordaram parcialmente. Os avaliadores alegaram que existem conceitos específicos de domínio que não foram tratados na ferramenta. Entretanto, essa customização é complicada e em um primeiro momento, optamos por manter a ferramenta o mais genérica possível. Todas as considerações foram anotadas e serão consideradas nas próximas versões do SciManager.

Apesar de os resultados serem promissores, novas avaliações devem ser realizadas no SciManager a fim de confirmar sua utilidade para a comunidade científica. Além disso, segundo Travassos *et al.* [12] devemos explicitar quais são as ameaças a validade do experimento realizado. Uma das ameaças é o uso de alunos como avaliadores. Tentou-se diminuir o impacto dessa ameaça selecionando alunos que já conhecessem os conceitos de

experimentação científica e já trabalhassem em projetos científicos reais (seja no nível de IC, Mestrado ou Doutorado). Além disso, Svahnberg *et al.* [13] defendem que o uso de alunos em experimentos é válido desde que o experimento tenha sido conduzido da forma correta. Outra ameaça a validade foi o curto espaço de tempo para se realizar o experimento e o treinamento. Os avaliadores tiveram cerca de uma hora de treinamento e duas horas cada um para realizar a análise. Tal item foi questionado na avaliação e um dos avaliadores mencionou que necessitava de mais tempo para avaliar a ferramenta de forma satisfatória. Uma nova avaliação está planejada com um grupo maior de avaliadores.

5. TRABALHOS RELACIONADOS

Existem algumas iniciativas que tem como objetivo oferecer apoio aos cientistas na gerência de projetos científicos. O LabGuru⁴ tem como objetivo auxiliar o cientista na gerência do seu laboratório. Apesar de possuir funcionalidades para distribuição de tarefas, o LabGuru tem como objetivo maior alocação de recursos (pessoas e equipamentos) nas tarefas do dia-a-dia e a compra de equipamentos para que o laboratório possa continuar funcionando. Não há uma preocupação com a execução do experimento e com a análise dos dados no LabGuru. A importância de se ter métodos e técnicas que ajudem a gerenciar projetos científicos foi também levantada por Portny e Austin [14]. Nesse mesmo artigo, Portny e Austin discutem sobre a falta de ferramentas para apoiar tal gerência. Pinheiro *et al.* [15] propõem uma metodologia de gestão de projetos de pesquisa, que se baseia no uso de técnicas de gerência de projetos na pesquisa para a obtenção de produtos. Apesar de não apresentarem nenhuma ferramenta, os autores discutem a importância de tais ferramentas para o processo. Além disso, ferramentas comerciais já existentes como o Trello⁵ e o Tasker⁶ podem ser utilizadas para gerência de projetos científicos. Entretanto, como essas ferramentas não tem como objetivo atender tal categoria de projetos, não há associação com o conceito de experimento científico e nem mecanismos para especificar um *workflow* ou executá-lo em um SGWfC existente.

6. CONCLUSÕES

Gerenciar um projeto científico é uma tarefa nada trivial. Um projeto científico pode englobar vários experimentos e cada experimento pode necessitar várias execuções de diferentes *workflows* científicos. Essa gerência ainda se torna mais complexa a medida que a equipe do projeto se encontra distribuída geograficamente. Dessa forma, distribuir tarefas, medir esforços despendidos, e controlar a equipe é uma tarefa muito complexa. A área de gerência de projetos já trata de problemas semelhantes em diversas áreas como no desenvolvimento de *software*. Entretanto, a gerência de um projeto científico possui peculiaridades e características únicas. Dessa forma, ferramentas específicas para a gerência do projeto científico se fazem necessárias.

Nesse artigo apresentamos o SciManager, um sistema de informação cujo foco é apoiar a gerência de um projeto científico, desde a criação de tarefas, modelagem do *workflow*, até o acompanhamento de resultados. O SciManager foi desenvolvido utilizando *software* livre e é hospedado em um ambiente de nuvem para garantir alta disponibilidade de seus serviços. A avaliação experimental mostrou que mais de 88% dos avaliadores

acreditam que o uso do SciManager agrega valor ao seu trabalho e possivelmente melhorará o desempenho do projeto científico. Trabalhos futuros incluem uma avaliação experimental em uma escala maior e com uma maior variedade de avaliadores.

7. AGRADECIMENTOS

Os autores gostariam de agradecer ao CNPq, CAPES e à FAPERJ pelo financiamento parcial deste trabalho.

8. REFERÊNCIAS

- [1] E. Deelman, D. Gannon, M. Shields, and I. Taylor, "Workflows and e-Science: An overview of workflow system features and capabilities," *Future Generation Computer Systems*, vol. 25, no. 5, pp. 528 – 540, 2009.
- [2] H. V. Jagadish, J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J. M. Patel, R. Ramakrishnan, and C. Shahabi, "Big data and its technical challenges," *CACM*, vol. 57, no. 7, pp. 86–94, Jul. 2014.
- [3] J. Freire, D. Koop, E. Santos, and C. T. Silva, "Provenance for Computational Tasks: A Survey," *CSE*, vol. 10, no. 3, pp. 11–21, 2008.
- [4] *A Guide to the Project Management Body of Knowledge: PMBOK(R) Guide*, 5 edition. Project Management Institute, 2013.
- [5] M. Mattoso, C. Werner, G. H. Travassos, V. Braganholo, E. Ogasawara, D. Oliveira, S. Cruz, W. Martinho, and L. Murta, "Towards supporting the life cycle of large scale scientific experiments," *International Journal of Business Process Integration and Management*, vol. 5, no. 1, pp. 79–92, 2010.
- [6] D. Oliveira, E. Ogasawara, F. Baião, and M. Mattoso, "SciCumulus: a lightweight cloud middleware to explore many task computing paradigm in scientific workflows," in *3rd IEEE Cloud, USA*, 2010, pp. 378–385.
- [7] C. Soanes and A. Stevenson, *Oxford Dictionary of English*, 2nd Revised edition. Oxford University Press, 2003.
- [8] G. H. Travassos and M. O. Barros, "Contributions of In Virtuo and In Silico Experiments for the Future of Empirical Studies in Software Engineering," in *2nd Workshop on Empirical Software Engineering*, Rome, 2003, pp. 117–130.
- [9] F. D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Q.*, vol. 13, no. 3, pp. 319–340, Sep. 1989.
- [10] I. Souza, P. Oliveira, E. Bispo Junior, A. C. Inocêncio, and P. A. Parreira Junior, "TESE– Um Sistema de Informação para Gerenciamento de Projetos Experimentais em Engenharia de Software," in *Anais do XI SBSI*, Goiânia, 2015, pp. 555–562.
- [11] K. A. C. S. Ocaña, D. de Oliveira, E. Ogasawara, A. M. R. Dávila, A. A. B. Lima, and M. Mattoso, "SciPhy: A Cloud-Based Workflow for Phylogenetic Analysis of Drug Targets in Protozoan Genomes," in *BSB 2011*. Springer Berlin Heidelberg, 2011, pp. 66–70.
- [12] S. N. Mafra and G. H. Travassos, "Técnicas de Leitura de Software: Uma revisão Sistemática," in *XIX Simpósio Brasileiro de Engenharia de Software*, Uberlândia, Brasil, 2005.
- [13] M. Svahnberg, A. Aurum, and C. Wohlin, "Using students as subjects - an empirical evaluation," in *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*, Kaiserslautern, Germany, 2008, pp. 288–290.
- [14] S. E. Portny, J. A. 12, 2002, and 8:00 Am, "Project Management for Scientists," Available: www.sciencemag.org/careers/2002/07/project-management-scientists. [Acessado: 19-Fev-2016].
- [15] A. A. Pinheiro, A. C. Siani, J. de F. Guilhermino, M. das G. M. de O. Henriques, C. M. Quental, and A. P. B. Pizarro, "Management methodology for product-oriented R&D projects: a proposal," *Revista de Administração Pública*, vol. 40, no. 3, pp. 457–478, Jun. 2006.

⁴ <http://www.labguru.com/features/research-project-management>

⁵ <https://trello.com/>

⁶ <http://www.tasker.com.br/>