

# CRIAÇÃO DE FERRAMENTAS E WORKFLOWS PARA EXECUÇÃO DE MODELOS DE SIMULAÇÃO POR MEIO DA PLATAFORMA GALAXY

**Ramão Tiago Tiburski**

Acadêmico do curso de Ciência da Computação da Universidade de Passo Fundo  
111977@upf.br

**Willingthon Pavan**

Professor do curso de Ciência da Computação da Universidade de Passo Fundo  
pavan@upf.br

**José Maurício Cunha Fernandes**

Pesquisador da Embrapa Trigo – Empresa Brasileira de Pesquisa Agropecuária  
mauricio.fernandes@embrapa.br

**Carlos Amaral Holbig**

Professor do curso de Ciência da Computação da Universidade de Passo Fundo  
holbig@upf.br

**Rafael Rieder**

Professor do curso de Ciência da Computação da Universidade de Passo Fundo  
rieder@upf.br

***Resumo.** Este artigo apresenta um estudo sobre a plataforma Galaxy e modelos de simulação, analisando a possibilidade de uso desta plataforma na agricultura, com o objetivo de facilitar a execução de modelos de simulação através de ferramentas e workflows que podem ser desenvolvidos e incorporados ao Galaxy.*

***Palavras-chave:** Galaxy, Ferramenta, Workflow.*

## 1. INTRODUÇÃO

A sustentabilidade alimentar e as mudanças climáticas têm se tornado assuntos relevantes nos dias atuais. Nesse sentido, prever possíveis cenários pode ser um fator importante para a tomada de decisões. No entanto, a previsão não é uma tarefa simples, pois exige conhecimento e dados de diversas áreas, tais como: agricultura e informática. Porém, o progresso nessas áreas é dificultado pela falta de dados compartilhados, ferramentas e uma linguagem comum para a comunicação. Percebe-se a necessidade da organização, integração e compartilhamento das

informações para que possam ser manipuladas, reusadas e reexecutadas de uma forma dinâmica. Neste sentido, os laboratórios das Universidades de Penn State e Emory desenvolveram o Galaxy, uma plataforma web aberta que possibilita a criação e a execução de workflows, a seleção e integração de ferramentas e o compartilhamento de dados (TEAM, 2013). A funcionalidade central do Galaxy é um sistema interativo baseado na web para o projeto de aplicações e workflows. Um workflow é criado com base em ferramentas e dados de entrada e caracteriza-se como um fluxo de tarefas para ser executado. Outra vantagem do uso do Galaxy é que ele permite que novas ferramentas sejam desenvolvidas, fazendo com que se adequem ao que é necessário.

Este artigo apresenta um estudo sobre a plataforma Galaxy, com o objetivo de desenvolver ferramentas e workflows que possam simplificar o processo de execução de modelos de simulação relacionadas à agricultura. Este processo, quando realizado manualmente, exige do usuário uma série de configurações e um alto conhecimento computacional, pois o usuário necessita ter o

conhecimento sobre os programas envolvidos em sua simulação. Além disso, a simulação não é realizada apenas por um programa e sim por um conjunto de programas invocados sequencialmente.

## 2. MATERIAIS E MÉTODOS

### 2.1 Modelos de Simulação

Modelos de simulação buscam repetir com fidelidade, por intermédio do uso do computador, características, comportamento e condições do sistema do mundo real (CHWIF & MEDINA, 2006).

A grande maioria dos modelos de simulação de culturas combinam dados sobre clima, física do solo, química do solo, fisiologia vegetal, genótipo e manejo cultural em uma superfície homogênea. Eles trabalham com arquivos que possuem dados de entrada e, após o processamento, os resultados são armazenados em arquivos com dados de saídas. Porém, para serem utilizados, estes conjuntos de dados devem possuir uma arquitetura que permita a sua manipulação, de forma a facilitar a obtenção de informações e posterior análise. Para tal nível de evolução, torna-se necessário construir colaborações entre diversos utilizadores potenciais da arquitetura proposta, dentre eles, o meio acadêmico, a indústria e o governo. Neste contexto, é importante também que diferentes áreas sejam abrangidas, tais como geociências, economia, tecnologia e ciências agrícolas.

O modelo proposto por Kim & Beresford (2012) realiza um estudo sobre um modelo de simulação que detecta o risco da ocorrência do Cancro Europeu (*Nectria*) em uma determinada região com base em dados climáticos. Essa doença atinge principalmente a maçã e é causada pelo fungo *Neonectria galligena*. Este modelo foi selecionado para demonstrar a possibilidade de integração entre o Galaxy e modelos de simulação.

### 2.2 Galaxy

O Galaxy é uma plataforma que agrega um conjunto de componentes reutilizáveis que podem ser integrados em outras aplicações. Estes componentes encapsulam funcionalidades de forma que o usuário tenha acesso somente a interfaces de ferramentas computacionais, manipulando facilmente conjuntos de dados e suas relações. Uma aplicação Galaxy é uma solução que disponibiliza o acesso a determinadas ferramentas por meio de uma interface web, a qual possui opções de execução para uma pesquisa computacional. O servidor Galaxy é uma execução dessa aplicação com determinadas ferramentas disponíveis (GOECKS; NEKRUTENKO; TAYLOR, 2010).

Cabe destacar que a plataforma Galaxy é um software livre utilizado em pesquisas sobre bioinformática e biomedicina. Porém, neste trabalho o objetivo é a utilização de sua estrutura para o desenvolvimento de ferramentas e workflows que facilitem a execução de modelos de simulação. Para tanto, o método utilizado prevê o desenvolvimento de novas ferramentas que podem incorporar o modelo dentro do Galaxy. Assim workflows podem ser criados para o processo de execução e reexecução de modelos.

#### 2.2.1 Ferramentas e Workflows

O Galaxy permite o desenvolvimento de novas ferramentas ou que sejam feitas modificações em suas próprias ferramentas, de maneira flexível, para que atenda todas as necessidades (GOECKS; NEKRUTENKO; TAYLOR, 2010).

O desenvolvimento de ferramentas no Galaxy se baseia em arquivos no padrão XML e seguem uma série de regras quanto às *tags*, que são utilizadas para criar o formulário da ferramenta no Galaxy, o qual interpreta o XML. Campos são gerados e estes interagem com a ferramenta, gerando uma página web dentro do framework (TEAM, 2013).

Um workflow pode ser descrito como as definições de execução de uma ferramenta,

considerando os valores para os seus parâmetros e o conjunto de dados utilizado. O usuário pode configurar um workflow para uma ferramenta, a fim de ser executado quantas vezes for necessário, além de seus parâmetros de entrada e conjunto de dados que irá atuar.

Além disso, o workflow pode ser configurado para execução com qualquer ferramenta, inclusive interligando resultados de saída de um fluxo para a entrada de outro. Nestes casos, é necessária uma verificação de compatibilidade entre as ferramentas (DEFELICIBUS, 2012).

Na Figura 1 pode ser visto um exemplo de workflow, o qual foi criado por meio da ferramenta de operações aritméticas, já existente no Galaxy, cujo objetivo é demonstrar a vantagem do uso de workflows.

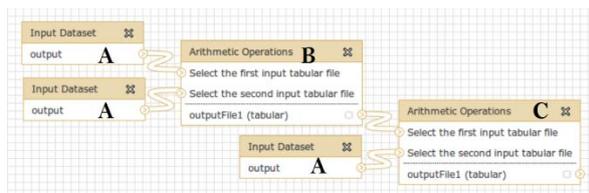


Figura 1. Exemplo de workflow

Neste exemplo, os valores de entrada (arquivos) podem ser vistos na seção A da Figura 1, os quais serão utilizados no processo que será realizado. Na seção B (Figura 1B), é realizada a soma dos arquivos de entrada, já na seção C (Figura 1C), é realizada a multiplicação entre o resultado obtido por B e um terceiro arquivo. Ao final, um único resultado é gerado. A vantagem de criar workflows é que após a escolha dos arquivos ou dos dados de entrada, todo o fluxo definido em sua criação é executado e o resultado final é gerado, simplificando e automatizando o processo.

## 2.4 Outros Recursos

Para o desenvolvimento dessas novas ferramentas, além do Galaxy, foram utilizados o sistema operacional Ubuntu Linux 12.04 32 bits, o sistema gerenciador

de banco de dados PostgreSQL, de onde são buscados os dados necessários para a geração dos resultados; e o RStudio, para a implementação do modelo e execução sem o uso do Galaxy.

## 3. RESULTADOS E DISCUSSÕES

Através dos estudos realizados e baseado no modelo de detecção de risco do Cancro Europeu proposto por Kim & Beresford (2012), foi desenvolvido uma ferramenta utilizando a linguagem R (<http://www.r-project.org>) com o objetivo de executar este modelo por intermédio da plataforma Galaxy.

O modelo implementado é baseado em condições climáticas favoráveis ao surgimento do Cancro Europeu em uma região. Para o cálculo de detecção de risco da doença o modelo utiliza três variáveis de entrada: porcentagem de dias com chuva, a quantidade de chuva e média de horas entre 11 e 16°C/dia em um mês. Após a execução do modelo é gerado um valor entre 0 e 1, sendo que se o resultado obtido for maior que 0.5, há o risco da doença.

Uma das execuções realizadas com o modelo foi em relação ao mês de março de 2013, em que obteve-se o valor 0.53, mostrando que as condições climáticas do mês foram favoráveis ao desenvolvimento da doença. Já, em outra execução do modelo, para o mês de maio de 2013, obteve-se o valor 0.48, mostrando que as condições climáticas desse mês não foram favoráveis ao desenvolvimento da doença.

Em relação ao Galaxy, o mesmo mostra-se uma alternativa muito poderosa, capaz de simplificar o manuseio de ferramentas complexas, possibilitando a reprodução de experimentos e permitindo que um conjunto de recursos seja utilizado durante o desenvolvimento de uma nova ferramenta.

Obteve-se como resultado, uma ferramenta capaz de incorporar o modelo de simulação citado, possibilitando a sua execução por meio da ferramenta e auxiliando na sua utilização e reexecução, conforme a necessidade do usuário.

A ferramenta incorporada ao Galaxy recebe um intervalo de datas utilizado para recuperar os dados necessários para a execução do modelo. Para a criação da ferramenta utilizou-se da estrutura do XML, especificando suas entradas e saídas, e fazendo a referência ao modelo implementado. Deste modo, depois que a ferramenta é incluída no Galaxy, passa a ser um novo item no menu de ferramentas e após a passagem dos parâmetros necessários para sua execução, usa o modelo implementado para fazer a simulação, gerando o resultado. Na Figura 2 pode ser observada a interface da ferramenta criada (A) e o resultado gerado (C).



Figura 2. Ferramenta criada, juntamente com o resultado gerado.

Já o workflow criado, oferece como vantagem a simplificação do trabalho, tendo como ponto positivo o armazenamento, organização de seus dados e a facilidade de reexecução do modelo.

Para este experimento, criou-se um workflow utilizando-se apenas de uma única ferramenta. Porém cabe destacar que a utilidade do workflow está relacionada ao uso de diversas ferramentas interligadas, automatizando o processo de execução.

Futuramente, novos modelos podem ser adicionados ao workflow, conectando a saída de um modelo com a entrada de outro, até que se chegue a um resultado final. Assim, esse processo é automatizado, simplificando o trabalho e facilitando o processo de execução e interligação de modelos.

#### 4. CONSIDERAÇÕES FINAIS

A execução de modelos de simulação por meio do Galaxy facilita e automatiza o processo. O uso de ferramentas

e workflows possibilita a manipulação e análise dos dados, além da ligação destas ferramentas em um único fluxo, respectivamente. Desta forma, um processo trabalhoso como a execução de modelos, pode se resumir a apenas alguns cliques, encurtando o tempo de conclusão de uma análise, melhorando o fluxo de informações e revolucionando os conceitos de execução de modelos na área agrícola.

#### 5. REFERÊNCIAS

- CHWIF, L.; MEDINA, A. Modelagem e Simulação de Eventos Discretos: Teoria e Prática. São Paulo. Ed. dos Autores. 2006. 320p.
- DEFELICIBUS, A. (2012). “Implementação de uma interface de execução do Framework 2PG na Plataforma Galaxy”. Trabalho de Conclusão de Curso – Curso de Ciência da Computação, Centro Universitário Barão de Mauá, 2012.
- GOECKS, J.; NEKRUTENKO, A.; TAYLOR, J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, [S.l.], p.1–13, 2010.
- KIM, K. S., BERESFORD, R. M. (2012). Use of a Climatic Rule and Fuzzy Sets to Model Geographic Distribution of Climatic Risk for European Canker (*Neonectria galligena*) of Apple. *Phytopathology*. 2012 Nov 28.
- TEAM, G. Galaxy Project. 2013. <http://galaxyproject.org/> - Acessado em: 10/05/2013.