



## Gerência de Dados de Processos de Genotipagem no GeneMaisLab

*Eliseu Germano<sup>1</sup>, Luana Rodrigues<sup>1</sup>, Thiago Souza<sup>1</sup>, Tereza Borba<sup>1</sup>, Marcelo Narciso<sup>1</sup>*

<sup>1</sup>Embrapa Arroz e Feijão

Santo Antônio de Goiás, Goiás, Brasil

eliseusilva@inf.ufg.br, {luana.rodrigues, thiago.souza, tereza, marcelo.narciso}@embrapa.br

### RESUMO

Um ensaio de genotipagem consiste em um conjunto de informações de amostras e marcadores moleculares para realização de experimentos destinados ao melhoramento genético. O desafio de gerenciar os dados dos ensaios em uma base de dados estruturada e permitir a exploração deles a partir de uma interface acessível via serviços web, é o principal objetivo do GeneMaisLab. Um modelo arquitetural foi elaborado e uma implementação foi realizada. Neste artigo são apresentados os resultados do projeto GeneMais em termos de artefatos produzidos e validação da proposta de desenvolvimento de um software de gestão de dados de processos de genotipagem.

**PALAVRAS-CHAVE:** Gestão de dados genotípicos, Sistemas de Informação, Arquitetura Orientada a Serviços.

### ABSTRACT

A genotyping assay consists of a set of molecular sample and marker information for experiments aimed at genetic improvement. The challenge of managing test data in a structured database and enabling it to be exploited from an accessible interface via web services is the main goal of GeneMaisLab. An architectural model was developed and an implementation was carried out. In this paper we present the results of the GeneMais project in terms of artifacts produced and validation of the proposal of development of data management software for genotyping processes.

**KEYWORDS:** Genotype data management, Information Systems, Service Oriented Architecture.

### INTRODUÇÃO

Um processo de genotipagem consiste em um mecanismo no qual podem ser identificadas as variantes genéticas em indivíduos de uma espécie (KORDROSTAMI; RAHIMI, 2015). Esse mecanismo pode ser representado a partir de um ensaio de genotipagem, que consiste de um conjunto

de amostras e marcadores moleculares para realização de um experimento destinados ao melhoramento genético (HOSBINO et al., 2002). Assim, gerenciar dados de um ensaio consiste em representar os processos envolvidos desde uma etapa de elaboração de um experimento até o momento em que ocorre a obtenção dos resultados do mesmo.

Embora existam demandas para realizar sistematizações de processos de genotipagem, devido a diversos fatores, há uma complexidade em representar a estrutura desses processos. Uma das principais dificuldades em fazer essa representação está relacionada a dinamicidade dos elementos que podem compor um ensaio, como, marcadores moleculares, métodos de extração de DNA, representação dos genes analisados, entre outros.

Nesse sentido, o GeneMaisLab surgiu com o propósito de criar formas de padronizações de elementos de um ensaio que possuem uma menor variabilidade. De forma sintetizada, o GeneMaisLab consiste de um *software* para elaboração e acompanhamento de ensaios, em que os dados de genotipagem são divididos em subconjuntos de dados que possuem uma variabilidade mais baixa (e.g., plataformas de genotipagem). Assim, esses subconjuntos de dados podem ser tratados separadamente, e quando agrupados permitem representações customizadas de um ensaio.

Podendo ser acessado a partir da web, o GeneMaisLab oferece de forma centralizada em um servidor web, mecanismos para gerenciar um conjunto de dados do processo de genotipagem que atende a demanda de pesquisadores e analistas da Embrapa. De modo geral, os pesquisadores podem atuar no sistema como solicitantes de serviços que devem ser executados em uma determinada análise, enquanto analistas e colaboradores podem armazenar informações sobre o andamento dessa análise (e.g., o status de execução, os resultados obtidos, o local em que os experimentos estão sendo realizados).

Neste artigo são apresentadas as principais características do projeto Genemais sob uma perspectiva das funcionalidades do sistema GeneMaisLab e dos benefícios do sistema após o seu desenvolvimento. Dessa forma, são abordadas algumas tomadas de decisões que ocorreram no projeto tanto do ponto de vista computacional, quanto do ponto de vista do domínio no qual ele faz parte. Também são apresentados os resultados em termos de integração de dados do GeneMaisLab com outros sistemas da Embrapa (e.g., Alelo).

O artigo está organizado em outras três seções. A segunda seção apresenta traz os principais conceitos relacionados ao domínio do projeto Genemais. A terceira seção apresenta os resultados do projeto Genemais, incluindo a sua estrutura e funcionamento do GeneMaisLab. Por fim, a quarta seção apresenta a conclusão e trabalhos futuros.

## MATERIAL E MÉTODOS

O GeneMaisLab é um sistema desenvolvido para viabilizar o gerenciamento de ensaios de genotipagem e marcadores moleculares na Embrapa. A partir dele é possível realizar uma melhor gestão das atividades relacionadas aos ensaios de genotipagem por meio da centralização dos

dados do processo de genotipagem, do armazenamento adequado desses dados em um servidor e da padronização de alguns desses dados diante de um mecanismo de gestão de padrões do sistema. Nesta seção é feita uma descrição do sistema (com ênfase nas motivações de desenvolvimento do projeto), os seus principais componentes de *software* produzidos e um conjunto de tecnologias envolvidas nessa produção.

### ***Descrição do GeneMaisLab***

Diante de um volume grande de informações genotípicas que são geradas continuamente por pesquisadores da Embrapa, surgiu a necessidade do desenvolvimento de um sistema de gestão de ensaios de genotipagem para auxiliar na gerência dessas informações. A ausência de um *software* específico para este fim, propiciava alguns impactos negativos nas atividades de pesquisa, tais como, perdas de dados por serem armazenados de forma inadequada, provocando retrabalho ou duplicidade de informações, perdas de dados por afastamento de pesquisadores, além da falta de padronização das informações coletadas nas análises laboratoriais.

Para lidar com esse cenário, foi desenvolvido um sistema chamado GeneMaisLab, no escopo do projeto chamado Genemais. Entre os vários objetivos desse sistema, o principal deles consiste em fazer a gestão dos dados de processos de genotipagem em um local único, centralizado, de forma estruturada e segura. Isso é feito no sistema por meio do gerenciamento dos dados de execução dos ensaios desde o planejamento até as etapas em que são feitas as considerações finais. Também é proposto no projeto que por meio do sistema seja viável o registro de eventos e o rastreamento das informações durante a execução dos ensaios.

Com o objetivo de sistematizar todo processo de gestão de dados relacionados aos ensaios de genotipagem e aos marcadores moleculares, o GeneMaisLab oferece uma interface *web* personalizada de acordo com as funções de cada perfil de funcionário ou colaborador da Embrapa. Os perfis de usuários definidos no sistema são:

- Solicitante – pesquisadores da Embrapa que são responsáveis por planejar o ensaio e as condições adequadas para a sua execução;
- Executor – analistas da Embrapa que são responsáveis por implantar, conduzir e registrar os resultados dos processos de execução dos ensaios;
- Colaborador – estagiários ou membros externos da empresa;
- Administrador – responsável gerenciar as funcionalidades do sistema e realizar o controle de acesso usuários ao mesmo.

### ***Definição do Fluxo de Solicitação de Atividades dos Ensaios***

O fluxo de atividades relacionados aos processos de genotipagem no GeneMaisLab variam de acordo com as funções de cada perfil de usuário no sistema. A Figura 1 apresenta o principal fluxo de atividades relacionado a execução de um ensaio no sistema. Esse fluxo consiste em quatro etapas que caracterizam os possíveis estados de um ensaio, sendo elas, planejamento, aceitação, em análise e executado.

Figura 1: Principal Fluxo de Atividades no GeneMaisLab



A primeira etapa (planejamento) é realizada por um usuário com perfil *solicitante* e as ações realizadas no sistema são relacionadas a elaboração do ensaio. Dessa forma, o solicitante preenche o formulário de cadastro de ensaios, solicita (caso necessário) o cadastro de novos padrões de dados para um gestor de padrões, importa dados de sistemas externos ao GeneMaisLab (a partir de componentes de *software* integrados a *web services* de outros sistemas), e por fim, cadastra o ensaio como um novo serviço laboratorial a ser realizado.

A segunda etapa (aceitação) é realizada por um usuário com perfil *executor* e as ações realizadas no sistema são relacionadas a avaliação do ensaio. Assim, o executor recebe as informações do ensaio como um novo serviço laboratorial solicitado (e.g., experimentos), avalia as informações do ensaio (e.g., verifica se há recursos disponíveis para realizar os experimentos) e aceita ou rejeita a solicitação. Caso o serviço seja aceito, a terceira etapa (análise) consiste em executar as atividades relacionadas ao ensaio, adicionar observações durante o processo de execução e finalizar o processo anexando os resultados dos experimentos realizados.

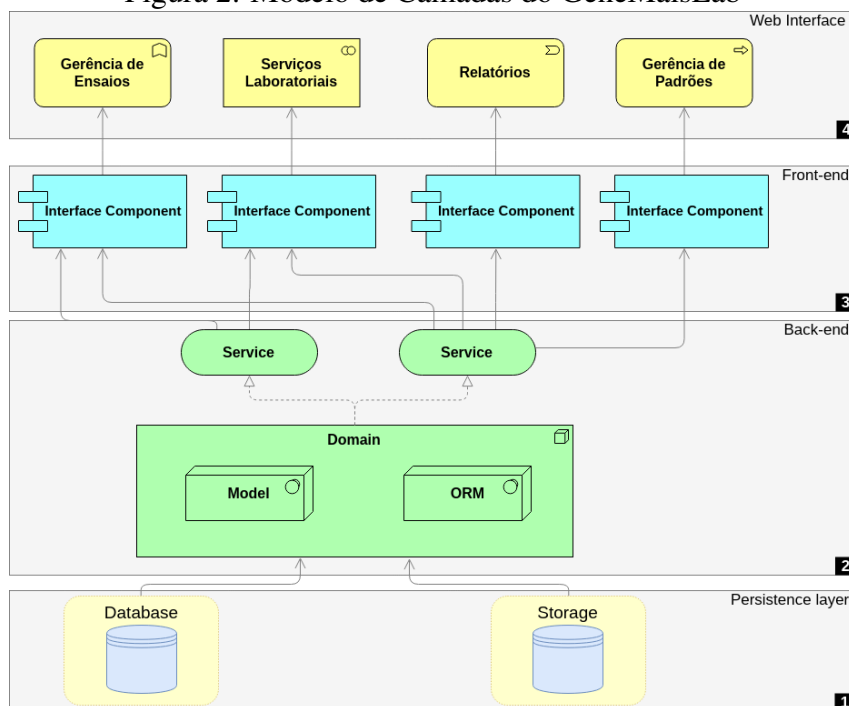
Por fim, a quarta e última etapa do fluxo (executado) está relacionada ao acesso aos ensaios cadastrados no sistema. Todos os ensaios podem ser listados ou consultados por quaisquer usuários do sistema. No entanto, apenas os ensaios públicos podem ser abertos pelos todos usuários (operação que permite acessar os detalhes de um ensaio, como, por exemplo, visualizar como os resultados de uma análise). Alguns ensaios podem ser cadastrados como restritos (nível de visibilidade privado). Esses últimos podem ser abertos apenas por usuários que fazem parte do grupo de pesquisa no qual o ensaio pertence. Tanto a escolha do grupo de pesquisa, quanto o nível de visibilidade do ensaio, é algo definido pelo solicitante durante a etapa de elaboração do ensaio.

### ***Componentes do Sistema***

Com acesso a partir de uma interface web, o GeneMaisLab possui um conjunto de elementos computacionais orientados a serviços que seguem o estilo arquitetural REST (FIELDING; TAYLOR, 2000). Esses elementos estão divididos em camadas de software de acordo com o que é apresentado na Figura 2. Essas camadas representam os níveis de abstração dos componentes.

Ainda na Figura 2, no nível mais baixo (nível 1) estão os componentes responsáveis pela persistência de dados e arquivos. No nível acima (nível 2) estão os componentes implementam as rotinas de *back-end* do sistema, ou seja, que fazem o gerenciamento dos recursos do nível

Figura 2: Modelo de Camadas do GeneMaisLab



1 e oferecem interfaces de alto nível para serem utilizadas, por exemplo, na implementação das regras de negócio dos serviços. Uma situação em que isso ocorre são nos componentes do nível 2 são os que fazem Mapeamento objeto-relacional (do inglês *Object-relational mapping* – ORM). Enquanto a camada de *back-end* provê os recursos implementados na forma de serviços, a camada de *front-end* (nível 3) consome esses recursos e apresenta os por meio de componentes de interface. Esses componentes são responsáveis por renderizar os dados do sistema em elementos visuais, que por sua vez, são apresentados no nível 1 a partir interfaces *web*.

Devido a demanda do GeneMaisLab por dados de outros sistemas corporativos da Embrapa, foi necessário realizar algumas integrações entre esses sistemas. A Figura 3 representa a partir de uma diagrama da UML as principais integrações realizadas. As integrações foram realizadas a partir do Barramento de Serviços da Embrapa<sup>1</sup>, exceto a integração realizada com o sistema Alelo Vegetal<sup>2</sup>.

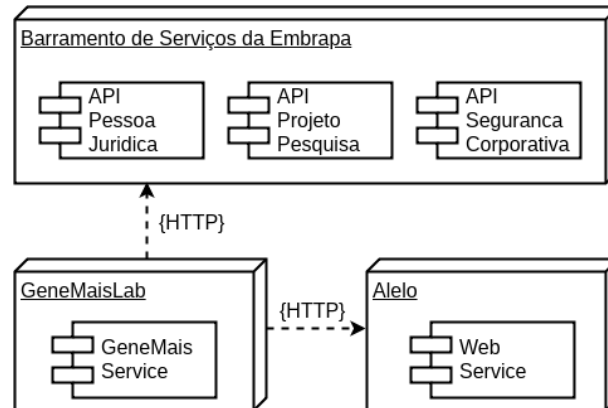
As APIs do Barramento de Serviços da Embrapa consumidas pelo GeneMaisLab foram:

- API Pessoa Jurídica – referente os serviços disponíveis para consumos das informações corporativas referente as unidades centrais e descentralizadas que compõe a estrutura organizacional da empresa, das pessoas físicas e jurídicas com vínculo com a Embrapa.
- API Projeto de Pesquisa – referente os serviços disponíveis para consumos das informações corporativas referente aos projetos de pesquisa juntamente com seus planos de ação e atividades.
- API Segurança Corporativa – refere-se as formas de autenticação dos usuários aos

<sup>1</sup><https://sandbox-api.sede.embrapa.br/store/?tenant=embrapa.sede>

<sup>2</sup><http://alelobag.cenargen.embrapa.br/AleloConsultas/Home/index.do>

Figura 3: GeneMaisLab no Contexto de Sistemas da Embrapa



sistemas corporativos da Embrapa.

Em relação ao Alelo Vegetal, a integração foi realizada com o objetivo de obtenção de dados Genotípicos armazenados e categorizados no Banco ativo de germoplasma (BAG). O BAG consiste de um conjunto de dados (acessos) conservados a curto e médio prazos, geneticamente representativos de um grupo, família, espécie ou gênero, conservados, regenerados, caracterizados e disponibilizados para uso em pesquisas científicas e intercâmbio, sem a prática de descarte. Essas informações ficam armazenadas em um banco de dados relacional, que por sua vez, foi consultado pelo *web service* desenvolvido como um componente do Alelo.

As tecnologias utilizadas para implementar esses componentes de software estão relacionadas com a linguagem Java e com alguns de seus principais frameworks e bibliotecas. Dentre os utilizados estão o Spring Boot<sup>3</sup>, Hibernate<sup>4</sup> e Google Web Toolkit Project (GWTP)<sup>5</sup>. Todos os componentes de foram integrados pela ferramenta de automação de compilação e gerência de dependências Maven<sup>6</sup> e são executados em *containers* da plataforma Docker<sup>7</sup>. Os detalhes computacionais relacionados ao desenvolvimento do GeneMaisLab podem ser encontrados em (GERMANO et al., 2017), (GERMANO; NARCISO, 2018b) e (GERMANO; NARCISO, 2018a).

## RESULTADOS E DISCUSSÃO

Neste seção são apresentados os resultados da construção dos principais componentes do GeneMaisLab, sendo eles: Gestão de Ensaios, Gestão de Padrões, Gestão Laboratorial e Gestão do Conhecimento.

### *Resultados da Implementação da Gestão de ensaios*

Componente responsável por disponibilizar um conjunto de serviços para manipulação dos ensaios de genotipagem, assim como das informações que o compõe, desde as formas de

<sup>3</sup><https://spring.io/projects/spring-boot>

<sup>4</sup><https://hibernate.org/>

<sup>5</sup><http://dev.arcbees.com/gwtp/>

<sup>6</sup><http://maven.apache.org>

<sup>7</sup><https://www.docker.com/>

identificação do ensaio até suas características mais peculiares, como os dados da equipe de genotipagem, locais de execução, marcadores moleculares utilizados, os mapas de placas com seus conjuntos de amostras, entre outras. A Figura 4 apresenta a tela de elaboração dos ensaios.

Figura 4: Gestão de Ensaios no GeneMaisLab.

Durante a gestão de um ensaio, podem ser definidas informações de identificação de um ensaio (ex., funcionário responsável pela solicitação e o projeto de pesquisa relacionado) e informações administrativas (ex., prazo para entrada das amostras e prazo de entrega dos resultados). O software também permite que, durante o cadastro de um ensaio, sejam importadas informações de outro sistemas da Embrapa, como por exemplo, alguns detalhes de informações genotípicas que podem ser obtidos a partir do Sistema Alelo.

### ***Resultados da Implementação da Gestão de Padrões***

Componente que possibilita o gerenciamento dos padrões relacionados aos ensaios de genotipagem (Figura 5), viabilizando adição e modificação dinâmicas desses padrões, como, os métodos de extração de DNA (kit, CTAB, etc.), natureza da amostra (folha, semente, buque, etc.), as plataformas de genotipagem (Dart, Cornel, Ilumina, etc.), cultura (arroz, feijão, milho, etc.), caráter(seca, resistência à doença, etc.) entre outros.

Figura 5: Gestão de Padrões no GeneMaisLab.

## Resultados da Implementação da Gestão de Laboratorial

Uma vez cadastrado, um ensaio é colocado em uma lista de serviços (Figura 6) a ser realizado em um laboratório da Embrapa. Esses serviços podem ser acessados por usuários que realizaram a execução dos experimentos nos laboratórios (ex., analistas e colaboradores), que por sua vez, podem armazenar informações sobre o andamento de um ensaio, como o status de execução, os resultados obtidos e o local em que os experimentos estão sendo realizados.

Figura 6: Gestão Laboratorial no GeneMaisLab.



The screenshot shows the 'Gerenciador de Marcadores Moleculares' interface. On the left is a sidebar with navigation options: Gestão de Ensaio, Laboratório, Serviços, Mapas de Placas, Relatórios, Tutoriais, Gestão do Conhecimento, and Gestão de Padrões. The main area is titled 'Acompanhamento dos Serviços' and contains a table with the following data:

Código	Título	Responsavel	Status	
201913030	Ensaio 001	ELISEU GERMANO DA SILVA	Planejado	
201913031	Ensaio 002	ELISEU GERMANO DA SILVA	Planejado	
201913032	Ensaio 003	ELISEU GERMANO DA SILVA	Planejado	

Diante da Gestão Laboratorial também é possível ter acesso a diferentes visões das informações de um ensaio, como por meio dos relatórios que sintetizam as informações dos ensaios de acordo com algum tipo de filtro escolhido, ou pela *interface* de um mapa de placas, que consiste em um desenho esquemático das placas de genotipagem de 96 ou 382 poços (Figura 7)). Assim o usuário poderá verificar a qualquer momento a disposição dos indivíduos, os indivíduos controles (testemunhas) e os marcadores que estarão sendo avaliados no ensaio.

Figura 7: Gestão Laboratorial no GeneMaisLab – Mapa de Placas.



The screenshot shows a 'Placa de 96 poços' (96-well plate) interface. The plate is represented as a grid with columns 1-12 and rows A-H. Most wells contain a green circle icon, while wells C12 and D12 contain a red circle icon. A tooltip for cell B7 is displayed, showing the following information:

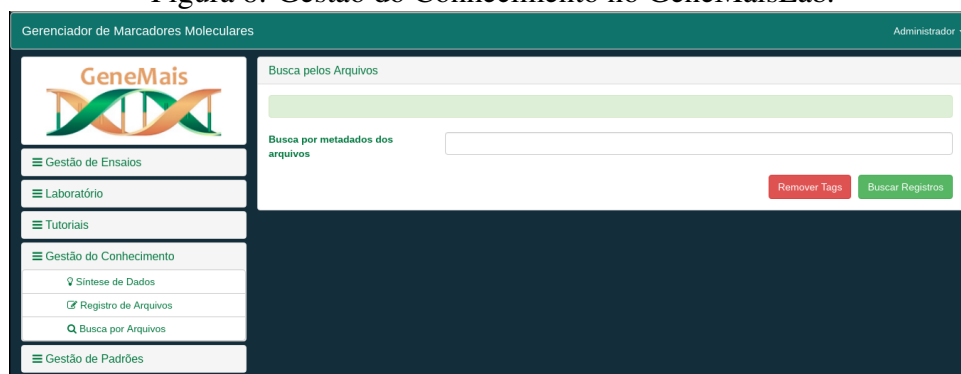
- Código Local (BGA/BGF): BGF11762
- BRA: 00021282-9
- Nome Comum: CF 800015, CHUMBINHO, ROXINHO, ROXINHO E CHUMBINHO
- Nome Científico: *Phaseolus vulgaris* L.
- Forma de Obtenção: Coleta



## Resultados da Implementação da Gestão do conhecimento

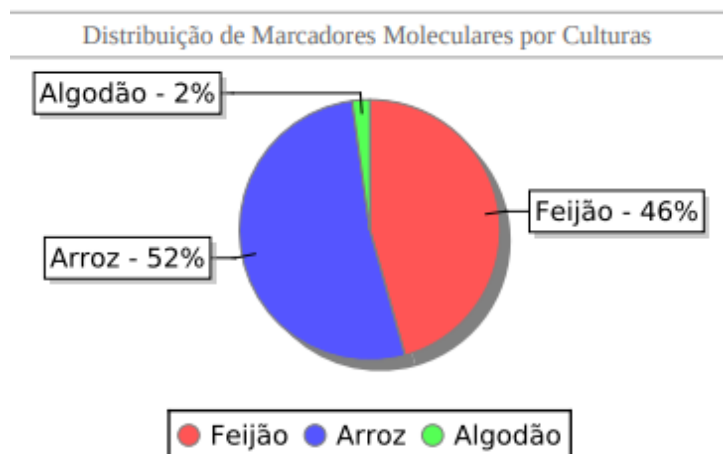
A gestão de conhecimento consiste de recursos do sistema que permite que durante a importação de arquivos no sistema sejam definidos um conjunto de metadados relacionados a esse arquivo. Esse metadados, por sua vez, são utilizados como elementos de busca em um elemento de interface para consulta aos arquivos Figura 8.

Figura 8: Gestão do Conhecimento no GeneMaisLab.



Outra funcionalidade da gestão de conhecimentos são os relatórios gerados a partir de macro visões dos dados registrados no banco de dados do GeneMaisLab. Na Figura 9, por exemplo, é representada a distribuição de marcadores moleculares por culturas. Nesse exemplo em específico, é possível visualizar de forma rápida alguns comportamentos dos dados registrados relacionado aos marcadores, como o fato de que a proporção de marcadores de Arroz e Feijão é consideravelmente maior que dos marcadores de Algodão.

Figura 9: Distribuição de Marcadores Moleculares por Culturas.



## CONCLUSÕES

O GeneMaisLab é um sistema que foi desenvolvido para viabilizar o gerenciamento de ensaios de genotipagem e marcadores moleculares na Embrapa. Por meio dele é possível realizar

uma melhor gestão das atividades relacionadas aos ensaios de genotipagem viabilizada pela da centralização dos dados do processo de genotipagem, do armazenamento adequado desses dados em um servidor e da padronização de alguns desses dados em função de um mecanismo de gestão de padrões.

Os benefícios que o Genemaislab trará aos usuários estão relacionados a possibilidade de ter um maior controle sobre os dados produzidos durante os processos de Genotipagem. Esse controle pode ser obtido a partir do registro de eventos durante a execução dos ensaios, do rastreamento das informações que são produzidas durante a execução de um ensaio e da possibilidade de consultar os resultados produzidos após o encerramento de todas atividades de um ensaio. Dessa forma, os usuários envolvidos na realização de um ensaio (e nos serviços em torno dele) podem ter um melhor acompanhamento do andamento das atividades a partir do sistema e das notificações (emitidas por e-mail, por exemplo) que são encaminhadas. Além disso, o sistema viabiliza uma melhor comunicação entre pesquisadores e analistas (executores dos ensaios) a partir da troca de mensagens que envolvem os serviços que são realizados durante o processo de genotipagem.

Para a Embrapa o sistema oferece a partir de suas interfaces a possibilidade de acompanhamento das atividades que são realizadas entre pesquisadores e analistas. Dessa forma, é possível que seja verificado o andamento das tarefas, a execução delas diante dos prazos estabelecidos além do acompanhamento de custos que são gerados em cada um dos ensaios de acordo com os recursos que são necessários para execução dos mesmos.

## REFERÊNCIAS

- FIELDING, R. T.; TAYLOR, R. N. *Architectural styles and the design of network-based software architectures*. [S.l.]: University of California, Irvine Doctoral dissertation, 2000.
- GERMANO, E.; NARCISO, M. Uma abordagem baseada em websocket para comunicação em tempo real no genemaislab. In: *IX EATI–Encontro Anual de Tecnologia da Informação*. [S.l.: s.n.], 2018. p. 110.
- GERMANO, E.; NARCISO, M. G. Encapsulando um sistema gerenciador de ensaios de genotipagem e de marcadores moleculares em containers docker. 2018.
- GERMANO, E. et al. Gene+: uma solução computacional distribuída para gerenciar ensaios de genotipagem e marcadores moleculares. *Embrapa Informática Agropecuária-Comunicado Técnico (INFOTECA-E)*, Campinas: Embrapa Informática Agropecuária, 2017., 2017.
- HOSBINO, A. et al. Marcador microssatélite na conservação de germoplasma vegetal. *Biotec Ciência Desenvol*, v. 29, p. 146–150, 2002.
- KORDROSTAMI, M.; RAHIMI, M. Molecular markers in plants: concepts and applications. *G3M*, v. 13, p. 4024–4031, 2015.