

## Gene + : uma Solução Computacional Distribuída para Gerenciar Ensaio de Genotipagem e Marcadores Moleculares

Eliseu Germano<sup>1</sup>  
Higor de Oliveira Silva<sup>2</sup>  
Edgard Henrique dos Santos<sup>3</sup>  
Marcelo Gonçalves Narciso<sup>4</sup>

### Resumo

Um ensaio de genotipagem é constituído de 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 Gene+. Um modelo arquitetural foi elaborado e uma implementação está sendo realizada. Nessa implementação é utilizado um Application Programming Interface (API) Gateway com a finalidade de compor os serviços web e integrá-los com outros sistemas da Empresa Brasileira de Pesquisa Agropecuária (Embrapa). Como forma de validar o modelo proposto, foi desenvolvido uma aplicação web usando o Google Web Toolkit (GWT) para consumir os recursos do sistema.

### Introdução

Devido aos diversos avanços metodológicos relacionados à área de genética molecular, tornou-se necessário o desenvolvimento de ferramentas computacionais que buscam a automatização e sistematização de alguns processos envolvidos nas análises laboratoriais (CAETANO, 2009). A necessidade de gerenciar grandes volumes de dados gerados por informações genotípicas, durante os ensaios de genotipagem, é um dos desafios desta área. Essa gerência pode demandar desde um processo de síntese das informações genotípicas em uma base de dados estruturada até o processo de consulta e rastreamento desses dados de forma eficiente e segura.

<sup>1</sup> Bacharel em Ciências da Computação, bolsista da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO.

<sup>2</sup> Bacharel em Sistemas de Informação, bolsista da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO.

<sup>3</sup> Analista de sistemas, analista na Embrapa Informática Agropecuária, Campinas, SP.

<sup>4</sup> Engenheiro de Telecomunicações, doutorado em Computação Aplicada, pesquisador na Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO.

Uma das dificuldades de construir um sistema computacional para gerenciar dados moleculares oriundos de ensaios de genotipagem está relacionada com o problema de modelar uma Interface de Programação de Aplicação (do inglês, Application Programming Interface – API) capaz de interoperar com outros sistemas e ferramentas usados no processo de genotipagem. Isso ocorre tanto pela falta de padronização de comunicação com os softwares que disponibilizam esses dados, quanto pelo elevado volume que eles possuem.

O propósito deste comunicado técnico é apresentar um conjunto de procedimentos computacionais empregados na construção de um sistema chamado Gene+. Esse sistema foi desenvolvido para permitir o gerenciamento de ensaios de genotipagem e marcadores moleculares para Empresa Brasileira de Pesquisa Agropecuária (Embrapa). Nele, foram utilizados alguns padrões de integração de software (e.g., API Gateway) a partir de uma abordagem com RESTful Web Services. Esses padrões permitiram a construção de um conjunto de serviços web para o acompanhamento de cada etapa do processo de análise dos ensaios, desde a preparação até o registro dos resultados obtidos.

Diante de uma arquitetura de software baseada em serviços web, foi implementada uma solução computacional com componentes de software interoperáveis com sistemas e ferramentas da Embrapa. Essa implementação enfatiza a necessidade de padronização entre os componentes do Gene+ com os demais sistemas que ele possa ser integrado. Isso se deve ao fato de que, a longo prazo, quando houver um aumento significativo do volume de dados moleculares e da complexidade envolvida nas análises dos ensaios, é natural o surgimento de novos serviços e/ou componentes de software para manter o desempenho e a escalabilidade do sistema.

## Materiais e Métodos

Em genética molecular, os marcadores são regiões sequenciais no Deoxyribonucleic acid (DNA) que geralmente são utilizadas em laboratórios para distinguir células, indivíduos, populações ou espécies (AVISE, 2013; KORDROSTAMI et al., 2015). O processo utilizado para identificar essas

regiões é denominado de genotipagem, sendo ele uma etapa do procedimento de melhoramento genético (EMBRAPA RECURSOS GENÉTICOS E BIOTECNOLOGIA, 2016; HOSHINO et al., 2002; NASCIMENTO et al., 2012). Dessa forma, os marcadores moleculares podem ser utilizados, por exemplo, para detecção de doenças genéticas ou de determinadas características de interesse em uma amostra (CAETANO, 2009).

Um dos desafios de gerenciar um ensaio de genotipagem está associado à quantidade de informações com ele relacionadas, tais como suas características (e.g., a natureza da amostra utilizada, o método de extração de DNA, a plataforma de genotipagem), informações dos marcadores moleculares (e.g., tipo do marcador e identificadores), informações das placas de genotipagem (e.g., o tamanho, as informações necessárias em cada poço/célula), informações genotípicas (e.g., cultura, indivíduos, controle), entre várias outras como os dados dos genes, o status do ensaio, o registro das observações analisadas durante e após a execução do ensaio e os dados gerados por ferramentas de software para este fim (e.g., GeneMapper). Dessa forma, cruzar estes dados, filtrar o que é relevante em uma determinada etapa do ensaio, consultá-los e relacioná-los com dados de outros ensaios, é uma tarefa que, se feita manualmente, demanda um esforço operacional demasiado.

A ausência de um software de gestão de informações que permita uma organização dos ensaios de genotipagem realizados e um acompanhamento dos que estão em execução, ocasiona vários problemas, 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. Diante disso, foi proposto um sistema computacional para cuidar da gestão de dados genotípicos e fenotípicos das diversas culturas (e.g., arroz e feijão) trabalhadas na Embrapa.

## Descrição do Gene+

Visando atender os problemas apresentados, o Gene+ tem como propósito gerenciar os dados dos ensaios desde o planejamento até as etapas

em que são feitas as considerações finais. A partir da centralização dos dados, o sistema permite o registro de eventos durante a execução dos ensaios, viabiliza o rastreamento das informações registradas e possibilita consultas aos resultados após o encerramento. Outra necessidade do sistema está relacionada com a comunicação com outros sistemas da Embrapa, como o sistema de login corporativo, para realizar a autenticação dos usuários, e outros sistemas que de alguma forma, disponibilizam informações e serviços que convergem com o que é necessário no Gene+, tais como, o Alelo, Ideare, SIExp e o KEGG.

A necessidade de integração com outros sistemas motivou a análise de alguns padrões arquiteturais de software que visam facilitar essa comunicação. Considerando que o Gene+ foi projetado para funcionar na web e que no futuro algumas de suas funcionalidades podem ser exploradas por outros sistemas, uma arquitetura baseada em serviços web foi adotada. Dessa forma, modelar serviços visando à construção de uma API para troca de mensagens na web são os principais pontos aqui discutidos e que é abordado a seguir. Além disso, também é apresentado alguns detalhes das tecnologias utilizadas na implementação da aplicação web consumidora dos serviços do Gene+.

### **Composição de serviços web a partir de um API Gateway**

Uma arquitetura orientada a serviços é um estilo de arquitetura de software onde as funcionalidades das aplicações são disponibilizadas na forma de serviços (KRAFZIG et al., 2005). Nela os serviços são disponibilizados a partir de um barramento de serviços corporativo (do inglês, Enterprise Service Bus – ESB) que possui uma gama de finalidades, sendo uma das principais delas a de criação de uma camada de software para abstração da comunicação entre diferentes tipos de tecnologias (DHARA et al., 2015).

Embora ainda seja uma prática comum o uso de um ESB para integração de sistemas, principalmente

quando se trata de integração com sistemas legados, dependendo do cenário essa pode não ser a melhor escolha. Um ESB possui uma série de funcionalidades (ENDREI, et al., 2004) que dependendo das entidades comunicantes, são desnecessárias. Em um cenário onde os sistemas interagem via web e possuem uma baixa heterogeneidade na forma de comunicação, o uso de outros padrões de integração, como API Gateway, pode ser mais adequado. Este último está sendo utilizado no Gene+ para composição dos serviços e integração com outros sistemas da Embrapa.

Um API Gateway é uma camada de software entre uma aplicação (front-end) e suas APIs (back-end) (ALAGARASAN, 2015), como é mostrado na Figura 1. A topologia desse tipo de Gateway em uma arquitetura distribuída permite que ele desempenhe algumas funcionalidades relacionadas à composição de serviços web, tais como, autenticação/autorização, controle de tráfego, análise e monitoramento dos serviços, transformações de dados entre entidades comunicantes e caching de dados (FAGERBERG, 2015). Um comparativo dessas funcionalidades entre diferentes soluções de API Gateway amplamente utilizadas pode ser verificado em (GÁMEZ-DÍAZ et al., 2015).

No cenário atual, onde existe um crescimento constante do número de serviços web, é comum o surgimento de novas especificações para padronizar a comunicação, tais como, OAuth<sup>5</sup>, PubSubHubbub<sup>6</sup> e Odata<sup>7</sup>. O OAuth, que atualmente encontra-se na versão 2, é uma especificação que permite a autenticação de aplicações utilizando apenas tokens para acessar um determinado recurso, ou seja, sem a necessidade do compartilhamento de credenciais. Um API Gateway que tenha suporte a esta tecnologia, por exemplo, facilita todo o controle de acesso dos usuários aos recursos disponibilizados por uma API.

<sup>5</sup> Disponível em: <<https://oauth.net/2/>>.

<sup>6</sup> Disponível em: <<https://pubsubhubbub.appspot.com/>>.

<sup>7</sup> Disponível em: <<http://www.odata.org/>>.

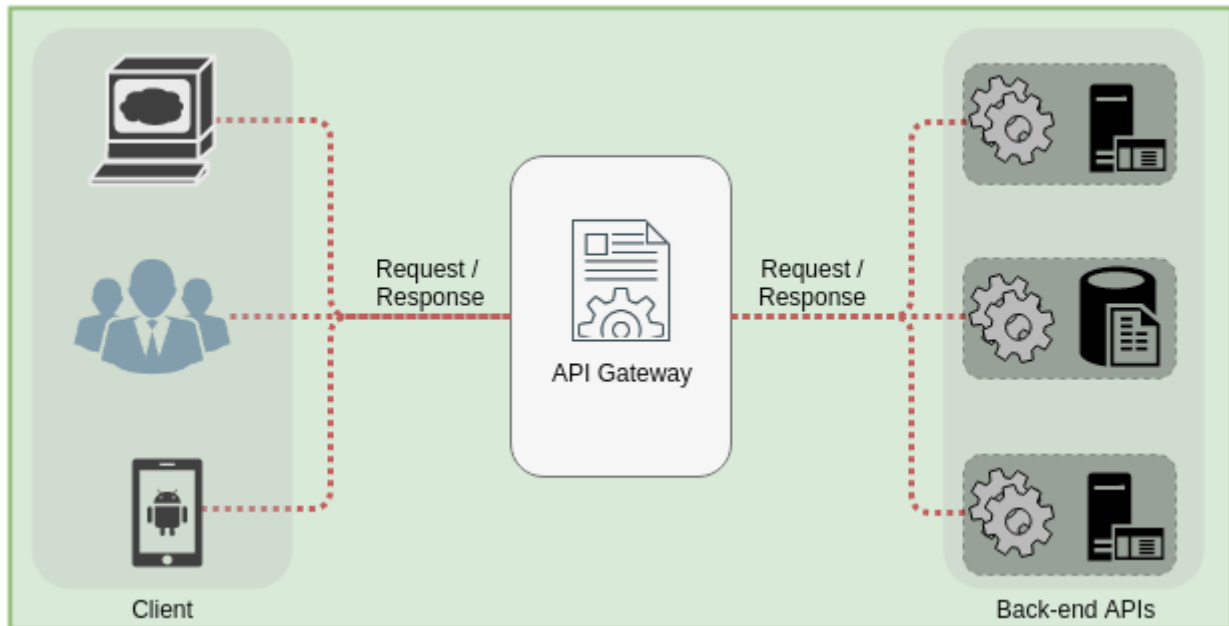


Figura 1. Expondo APIs a partir de um API Gateway.

## Descrição alto nível dos componentes de software do Gene+

O Gene+ é um sistema totalmente escrito em linguagem Java. Todos seus componentes, desde aqueles de mais baixo nível, executados em web servers, até os que geram os componentes de interface para serem executados em web browsers, foram implementados nessa linguagem de programação. Isso foi possível devido à combinação de alguns dos principais frameworks, bibliotecas e ferramentas da linguagem, como, Spring<sup>8</sup>, Hibernate<sup>9</sup>, Google Web Toolkit (GWT)<sup>10</sup>, GWTBootstrap3<sup>11</sup> e Google Web Toolkit Project (GWTP)<sup>12</sup>, integrados pela ferramenta de automação de compilação Maven<sup>13</sup>.

O Maven é utilizado no projeto tanto para gerenciar as dependências externas (tais como as citadas no último parágrafo) quanto as dependências internas, constituídas pelos componentes de software do sistema. Um esquema mostrando os relacionamentos entre essas dependências é apresentado na Figura 2 por meio de um diagrama

de componentes. Nele, a nomenclatura utilizada nos módulos é a mesma utilizada nos namespaces da implementação dos respectivos componentes de software, sendo eles:

- Genemais-shared – constituído de um conjunto de entidades de software que são comuns aos módulos genemais-domain e genemais-gwt.
- Genemais-domain – possui a maior parte das “regras de negócio” do sistema, ou seja, uma parte considerável da implementação das rotinas que representam as funcionalidades e viabilizam a realização das operações no software (e.g., disponibilizar a um método do genemais-service dados retornados de uma consulta ao banco de dados).
- Genemais-service – é a interface de programação entre o genemais-domain e o genemais-gwt, sendo o responsável por redirecionar as requisições aos métodos que disponibilizam os recursos do sistema.

<sup>8</sup> Disponível em: <<https://spring.io/docs/reference>>.

<sup>9</sup> Disponível em: <<http://hibernate.org/orm/documentation/5.1/>>.

<sup>10</sup> Disponível em: <<http://www.gwtproject.org/doc/latest/DevGuide.html>>.

<sup>11</sup> Disponível em: <<http://gwtbootstrap3.github.io/gwtbootstrap3-demo/apidocs/index.html>>.

<sup>12</sup> Disponível em: <<http://dev.arcbees.com/gwtp/>>.

<sup>13</sup> Disponível em: <<http://maven.apache.org/guides/>>.

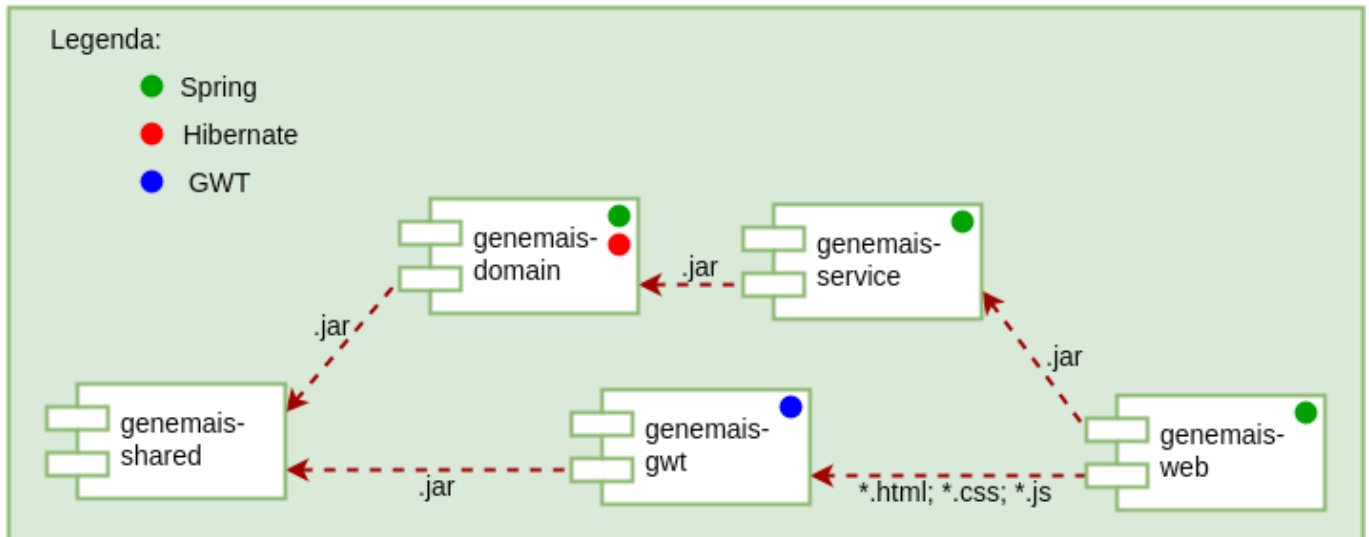


Figura 2. Dependências Internas e Externas do Gene+.

- Genemais-gwt – responsável pela geração de códigos que constituem a implementação dos componentes de interface do software (aplicação cliente). Essa geração de código é necessária pelo fato de que esses componentes são totalmente escritos em linguagem Java, sobre frameworks e bibliotecas do GWT, que realizam o parser do código para os componentes para linguagem de marcação HTML e de programação JavaScript.
- Genemais-web – têm a função de inicializar todas as classes de configuração dos outros módulos do sistema e de empacotar todos artefatos gerados por eles em um pacote war (do inglês, Web application ARchive).

### Análise do Gene+ sob uma perspectiva dos serviços web

A possibilidade de padronizar a troca de mensagens entre os componentes de software em um ambiente de computação distribuída que faz uso do Hypertext Transfer Protocol (HTTP), instigam o uso de uma arquitetura RESTful. Essa arquitetura é derivada do estilo arquitetural, para sistemas hipermídia distribuídos, conhecido como Representational State Transfer (REST) [FIELDING 2000]. A adesão à arquitetura traz uma série de benefícios para o sistema, como, o uso de uma notação comum para transferência de dados (e.g., JSON e XML); a exposição de recursos dos serviços a partir de uma estrutura conhecida como Uniform Resource Identifier (URI), deixando transparente para quem os consome toda a complexidade envolvida na lógica implementada;

facilidade de reutilização dos recursos a partir do uso de métodos (verbos) do protocolo HTTP. Todas essas características de padronização de troca de mensagens foram utilizadas na comunicação entre os componentes do Gene+.

A Figura 2, que apresenta visão do software em nível de componentes/módulos, é importante para o entendimento de algumas de suas funcionalidades e de tecnologias empregadas no seu desenvolvimento. No entanto, ela não apresenta uma perspectiva topológica desses componentes em uma rede de computadores. Dessa forma, a Figura 3 traz uma outra visão do sistema, a partir de um diagrama de implantação da Unified Modeling Language (UML). Nesse diagrama é enfatizado a comunicação das aplicações do Gene+ com seus respectivos serviços e com os serviços pertencentes a outros sistemas da Embrapa (e.g., ALELO).

Outro ponto importante na Figura 3 está na camada de software que possibilita a integração do sistema, constituída de um conjunto de funcionalidades que representa um padrão chamado API Gateway. Essa camada é um middleware entre as aplicações e os serviços RESTful do Gene+. As funcionalidades do API Gateway utilizado no Gene+ são:

- Autenticação e segurança - impede o acesso não autorizado aos serviços do sistema, permitindo que antes de cada requisição seja verificado as credenciais (por meio de access tokens) de quem há realizou.

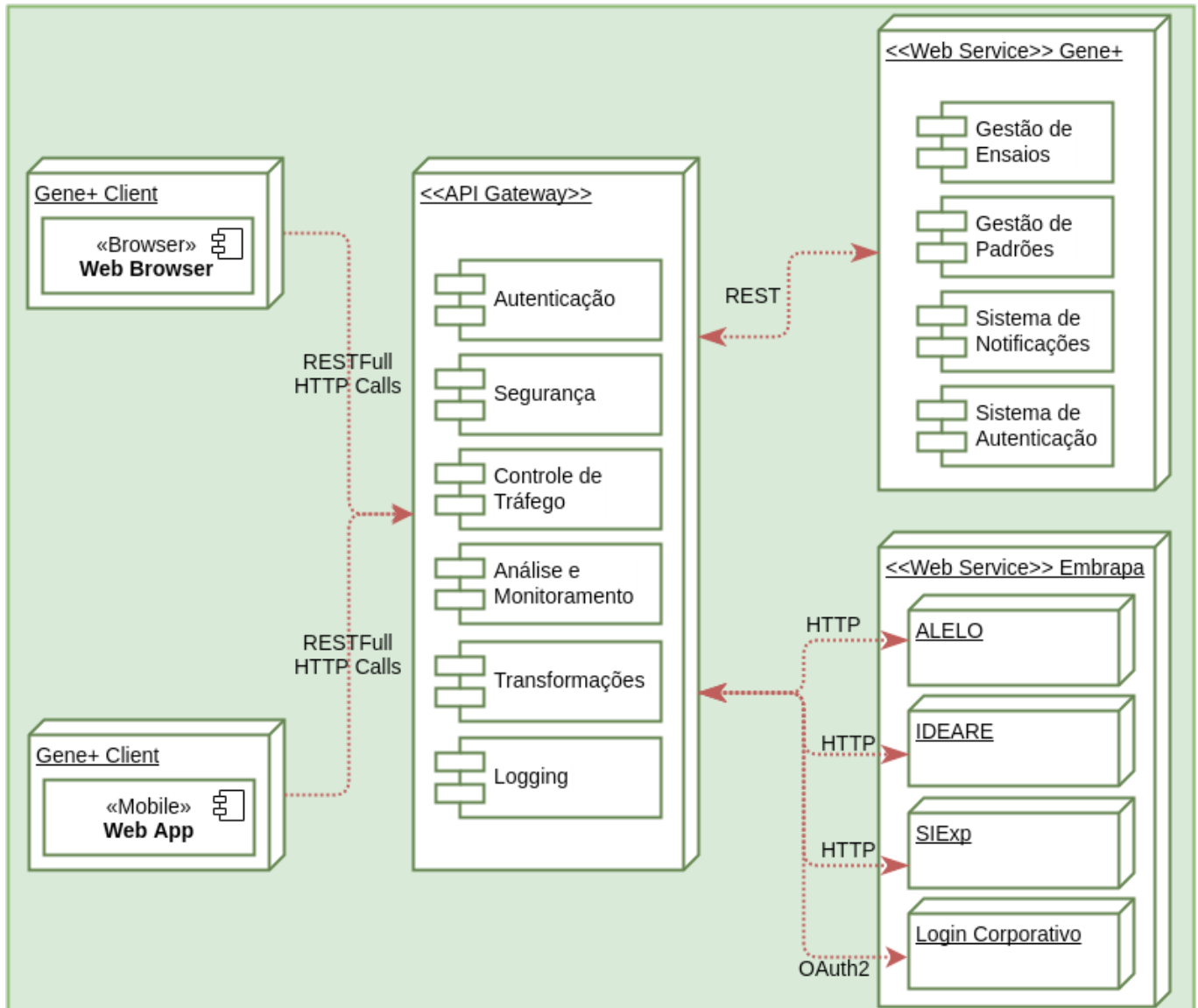


Figura 3. Gene+, seus componentes de software e a comunicação com outros sistemas.

- **Controle de tráfego** - permite o acompanhamento e gerenciamento de informações administrativas de acesso aos serviços de acordo com o tráfego de requisições.
- **Análise e monitoramento** - Permite a verificação de desempenho das APIs a partir de um conjunto de métricas e de ferramentas específicas para esse tipo de avaliação.
- **Transformações** - possibilita a filtragem e a manipulação de dados de parâmetros das requisições e das respostas dos serviços.
- **Logging** - constitui um conjunto de ferramentas para manipulação de logs relacionados a toda infraestrutura de comunicação.

Os componentes de software representados no web service do Gene+, é uma abstração em alto nível das principais funcionalidades deste sistema, sendo elas:

- **Gestão de ensaios** - responsável por disponibilizar um conjunto de serviços para manipulação dos ensaios de genotipagem, assim como das informações que o compõe, desde as formas de 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.
- **Gestão de padrões** - possibilita o gerenciamento dos padrões relacionados aos

ensaios de genotipagem, viabilizando adição e modificação dinâmicas desses padrões, como as culturas, os métodos de extração de DNA, os genótipos, as plataformas de genotipagem, entre outros.

- Sistema de notificações - responsável por informar a equipe que relaciona a um determinado ensaio de genotipagem, sobre cada evento realizado com este.
- Sistema de autenticação - possibilita a autenticação de usuários no sistema e sincroniza as informações dos usuários que utilizam o sistema de login corporativo da Embrapa.

Os componentes de software representados no web service do Gene+, é uma abstração em alto nível das principais funcionalidades deste sistema, sendo elas:

- Gestão de ensaios - responsável por disponibilizar um conjunto de serviços para manipulação dos ensaios de genotipagem, assim como das informações que o compõe, desde as formas de 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.
- Gestão de padrões - possibilita o gerenciamento dos padrões relacionados aos ensaios de genotipagem, viabilizando adição e modificação dinâmicas desses padrões, como as culturas, os métodos de extração de DNA, os genótipos, as plataformas de genotipagem, entre outros.
- Sistema de notificações - responsável por informar a equipe que relaciona a um determinado ensaio de genotipagem, sobre cada evento realizado com este.
- Sistema de autenticação - possibilita a autenticação de usuários no sistema e sincroniza as informações dos usuários que utilizam o sistema de login corporativo da Embrapa.

## **Análise do Gene+ sob a perspectiva da aplicação web**

Analisando a construção do Gene+ do ponto de vista da construção da aplicação (back-end) com seus respectivos componentes de interface, vale a pena apresentar algumas tecnologias utilizadas e a forma em que elas foram inseridas na arquitetura do sistema. Essas tecnologias estão relacionadas com o uso do framework Google Web Toolkit (GWT). Entre várias outras coisas, o GWT se caracteriza por permitir ao desenvolvedor de software a construção dos componentes de interface utilizando a linguagem de programação Java. Uma vez escrito o código destes componentes, quando este é compilado, o framework faz um parser dele para linguagem JavaScript. São vários os benefícios obtidos por essa característica, tais como:

- Evitar que desenvolvedores de software mais familiarizados com a linguagem Java tenham que escrever códigos em HTML, CSS e JavaScript;
- Evitar que um desenvolvedor tenha que lidar com o problema de compatibilidade da aplicação nos diversos navegadores web, uma vez que isso já é tratado pelo próprio framework;
- Gerar códigos ofuscados da aplicação desenvolvida, protegendo assim propriedade intelectual produzida e reduzindo o tamanho dos arquivos JavaScript gerados;
- Gerar código otimizado a partir de técnicas como Code Splitting<sup>14</sup>, onde o desenvolvedor de aplicações pode explorar características do AJAX para viabilizar o carregamento parcial e gradual da aplicação no navegador;
- Permitir o uso de uma arquitetura Model-View-Presenter (MVP) visando aumentar o processamento na aplicação, ou seja, no front-end, objetivando um aumento na escalabilidade do sistema como um todo e a redução no tráfego de dados no servidor.

No Gene+ também foi utilizado uma versão do Twitter Bootstrap adaptada para os componentes de interface do GWT, chamada GWTBootstrap3.

---

<sup>14</sup> Disponível em: <<http://www.gwtproject.org/doc/latest/DevGuideOptimizing.html>>.

Usando-a foi possível a construção de layouts responsivos e utilizando dos componentes disponibilizados pelo Bootstrap a partir do GWT. Além disso, também foi utilizado o Google Web Toolkit Project (GWTP) para auxiliar na padronização e modularização da aplicação a partir de uma arquitetura MVP. O GWTP também disponibiliza uma API REST-Dispatch para padronização da forma de estabelecer a comunicação da aplicação com os serviços utilizando Java API for RESTful Web Services (JAX-RS)<sup>15</sup>. Essa API REST-Dispatch foi utilizada para que a aplicação web do Gene+ consuma os serviços REST do mesmo.

## Resultados e discussão

As decisões de um projeto de software incluem a determinação de diversos aspectos relacionados ao produto de software a ser construído, tais como, o tipo da aplicação, a forma de distribuição em redes de computadores e os possíveis estilos arquiteturais de comunicação que podem ser utilizados. Essas decisões geralmente têm um impacto significativo no resultado do software, e, grande parte delas precisam ser definidas em etapas que antecedem o desenvolvimento. A seguir é apresentada uma breve discussão de alguns pontos que foram considerados nas etapas de projeto do Gene+, na forma em que eles foram implementados e nas suas possíveis melhorias.

Na arquitetura do Gene+ podemos encontrar dois pontos críticos que influenciam diretamente no funcionamento do sistema. O primeiro deles é o API Gateway, que funciona como um portal de acesso aos serviços. O principal problema dele é ser um ponto único de falhas, pois, como todas as requisições aos serviços passam por ele, a sua indisponibilidade torna o sistema todo indisponível. Outro ponto crítico está relacionado com um possível crescimento exponencial dos dados de ensaios de genotipagem. Considerando que o número de marcadores moleculares aumenta constantemente, assim como os ensaios de genotipagem, que por sua vez possuem um subconjunto desses marcadores, a longo prazo a manipulação desses dados pode se tornar impraticável.

Para os dois pontos críticos da arquitetura do Gene+, existem soluções que resolvem os respectivos problemas. Uma forma de lidar com o problema do API Gateway ser um único ponto de falhas é executando-o em um ambiente distribuído, pois, dessa forma, sempre que uma instância dele se torne indisponível, a outra assume imediatamente a comunicação. Outro cuidado em relação ao uso de API Gateway é no sentido de que ele deve ser escalável, ou seja, não deve introduzir uma latência ou ser um gargalo na comunicação. Uma forma de lidar com isso é introduzindo um balanceador de cargas entre as requisições das aplicações e as instâncias dos Gateways, como é mostrado na Figura 4.

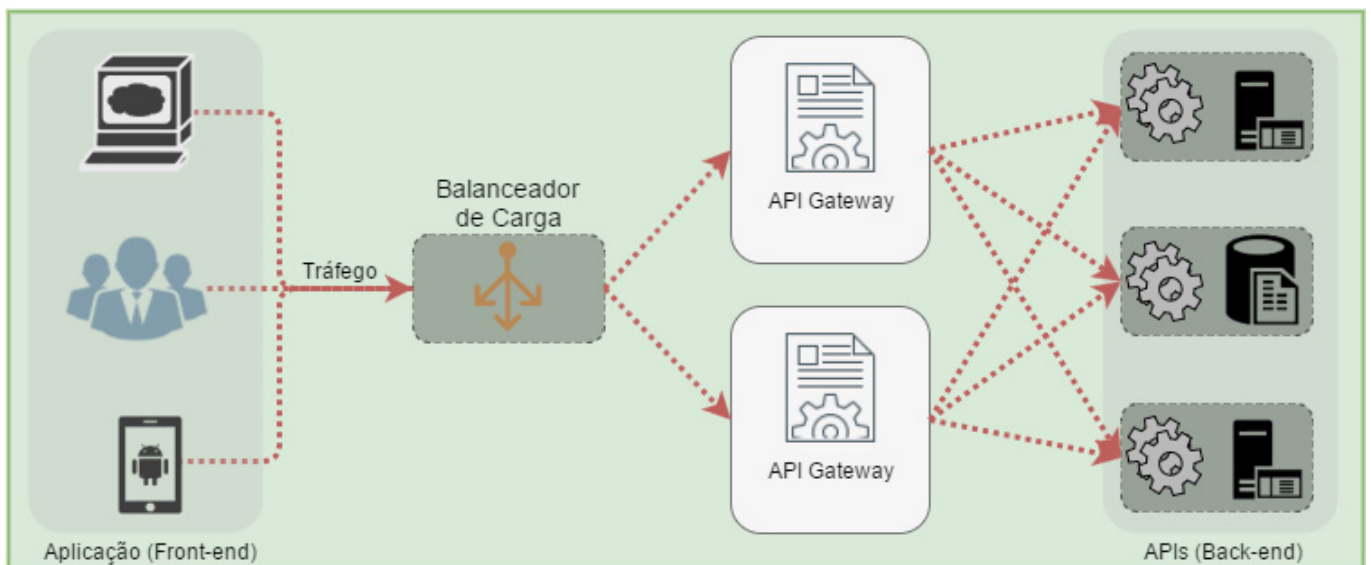


Figura 4. API Gateway distribuído.

<sup>15</sup> Disponível em: <<https://jax-rs-spec.java.net/nonav/2.0-rev-a/apidocs/index.html>>.



Em relação ao problema de crescimento do volume de dados de informações moleculares, existem algumas soluções de software para viabilizar buscas em real-time para lidar com esta situação. A solução adotada pelo Gene+ é o Elasticsearch<sup>16</sup>, que é um servidor de busca distribuído que disponibiliza uma API RESTful para realização de consultas. Essa ferramenta será integrada na arquitetura apresentada na Figura 3, como um novo componente do Gene+, visando principalmente o aumento de desempenho nas consultas mais complexas e permitindo a exploração dos dados moleculares registrados no sistema.

Do ponto de vista de andamento do desenvolvimento do Gene+, atualmente a maior parte dos requisitos funcionais, e alguns dos não funcionais, estão implementados e em validação com pesquisadores e analistas da unidade da Embrapa Arroz e Feijão. Em relação à implantação do sistema, inicialmente será realizada visando atender às necessidades dos funcionários dessa unidade, sendo eles, pesquisadores, analistas e estagiários. Posteriormente, será disponibilizado a partir de serviços web, uma API com recursos do sistema que possam ser consumidos por outros serviços/sistemas da Embrapa. A princípio, a política de acesso aos dados do Gene+ se restringe ao uso exclusivo das unidades da Embrapa, não possuindo, assim, uma API pública para colaboradores externos. Isso se deve às próprias características do sistema, que prioriza a gestão das informações dos ensaios de genotipagem preparados e analisados internamente. No caso da realização de serviços externos à empresa, o sistema permite que o funcionário da Embrapa que os solicitou, faça os registros das informações obtidas dos mesmos.

## Conclusões

Diante da necessidade de gerenciar informações de ensaios de genotipagem e dados de marcadores moleculares, foi proposto o desenvolvimento de um sistema computacional que permita a centralização dessas informações. Visando atender às necessidades de pesquisadores e analistas do laboratório de biotecnologia da Embrapa, o Gene+ disponibiliza um conjunto de funcionalidades

que permite o acompanhamento dos ensaios de genotipagem, desde a fase de planejamento até o processo de registro e análises dos resultados obtidos. Além disso, ele possibilita os cruzamentos de dados e a geração de relatórios relacionados aos ensaios cuja execução já foi encerrada. Isso o torna uma ferramenta de gestão de informações genotípicas das diversas culturas trabalhadas na Embrapa.

Embora atualmente a principal forma de acesso ao sistema é por meio de uma aplicação web, sua arquitetura foi projetada visando um suporte, há diversos tipos de aplicações (e.g., aplicações móveis) que possam consumir serviços web a partir da internet. Além disso, devido a importância da gestão das informações registradas e importadas para o Gene+, é natural que outros sistemas corporativos sejam adaptados no sentido de consumir as informações armazenadas nas bases de dados dele. Dessa forma, já foi desenvolvido uma API a partir de uma arquitetura RESTful, onde informações do Gene+ são disponibilizadas.

De acordo com o crescimento do número de ensaios de genotipagem, do número de marcadores moleculares e do número de sistemas que venham utilizar os serviços do Gene+, uma melhoria na forma em que as informações são buscadas deve ser realizada. Uma forma de lidar com essa situação e que está prevista no projeto é a adição da ferramenta de busca Elasticsearch. Essa ferramenta facilitará a realização de buscas complexas na base de dados do sistema.

Outras melhorias que ainda precisam ser realizadas estão relacionadas com os pontos críticos na comunicação, como no API Gateway, onde deve ser evitado que ele seja um ponto único de falha do sistema e que seja um gargalo na comunicação. Atualmente, uma solução viável é a partir da redundância no número de Gateways, garantindo assim que na ausência de um deles, um outro possa ser assumido. Além disso, com o uso de um balanceador de cargas, é possível redirecionar cada requisição para um Gateway específico, sendo esta escolha baseada no tráfego na rede.

---

<sup>16</sup> Disponível em: <<https://www.elastic.co/guide/en/elasticsearch/reference/1.4/index.html>>.

## Referências

- ALAGARASAN, V. **API management introduction and principles**. 2015. Disponível em: <<http://servicetechmag.com/189/0315-1>>. Acesso em: 12 nov. 2016.
- AVISE, J. C. **Molecular markers, natural history and evolution**. Sunderland: Sinauer Associates, 2013. 684 p. ill.
- CAETANO, A. R. **Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro**. Revista Brasileira de Zootecnia, v. 38, 8, p. 64-71, 2009.
- DHARA, K. M.; DHARMALA, M.; SHARMA, C. K. **A survey paper on service oriented architecture approach and modern web services**. [S.l.: s.n.], 2015.
- EMBRAPA RECURSOS GENÉTICOS E BIOTECNOLOGIA. **Genotipagem**. Brasília, DF: Embrapa. [2016]. Disponível em: <[http://www.agencia.cnptia.embrapa.br/Agencia45/AG01/arvore/AG01\\_23\\_1211200714127.html](http://www.agencia.cnptia.embrapa.br/Agencia45/AG01/arvore/AG01_23_1211200714127.html)>. Acesso em: 14 nov. 2016.
- ENDREI, M.; ANG, J.; ARSANJANI, A.; CHUA, S.; COMTE, P.; KROGDAHL, P.; LUO, M.; NEWLING, T. **Patterns: service-oriented architecture and web services**. [United States]. International Business Machines Corporation: International Technical Support Organization, 2004. Disponível em: <<http://www.immagic.com/eLibrary/ARCHIVES/GENERAL/IBM/SG246303.pdf>>. Acesso em: 12 nov. 2016.
- FAGERBERG, A. **Optimising clients with API gateways**. 2015. 71 p. (Master's thesis) – Faculty of Engineering, Lund University, Lund.
- FIELDING, R. T. Representational state transfer (REST). In: FIELDING, R. T. **Architectural styles and the design of network-based software architectures**. 2000. p. 76-105. Dissertation (Doctor of Philosophy in Information and Computer Science) – University of California, Irvine. Disponível em: <[https://www.ics.uci.edu/~fielding/pubs/dissertation/rest\\_arch\\_style.htm](https://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm)>. Acesso em: 13 nov. 2016.
- GÁMEZ-DÍAZ, A.; FERNÁNDEZ-MONTES, P.; RUIZ-CORTÉS, A. Towards SLA-Driven API gateways. In: JORNADAS DE CIENCIA E INGENIERIA DE SERVICIOS, 11., 2015, Santander. **Actas...** Cantabria: Universidad de Cantabria, 2015. Disponível em: <[http://www.isa.us.es/sites/default/files/Towards\\_SLA\\_API\\_Gateways.pdf](http://www.isa.us.es/sites/default/files/Towards_SLA_API_Gateways.pdf)>. Acesso em: 13 nov. 2015.
- HOSHINO, A. A.; PALMIERI, D. A.; BRAVO, J. P.; PEREIRA, T. E. B.; LOPES, C. R.; GIMENES, M. A. Marcador microssatélite na conservação de germoplasma vegetal. **Revista de Biotecnologia, Ciência e Desenvolvimento**, n. 29, p.146-150, 2002.
- KORDROSTAMI, M.; RAHIMI, M. **Molecular markers in plants: concepts and applications**. Genetics in the 3rd millennium, v. 13, n. 2, p. 4024-4031, 2015. Disponível em: <[http://g3m.ir/browse.php?a\\_id=953&slc\\_lang=en&sid=1&ftxt=1](http://g3m.ir/browse.php?a_id=953&slc_lang=en&sid=1&ftxt=1)>. Acesso em: 13 nov. 2016.
- KRAFZIG, D.; BANKE, K.; SLAMA, D. **Enterprise SOA: service-oriented architecture best practices**. Indianapolis: Prentice Hall Professional, 2005.
- NASCIMENTO, L. C.; VIDAL R. O.; MONDEGO, J. M. C.; COSTA, G. G. L.; JUNIOR, O. R.; RODRIGUES, F.; NEPOMUCENO, A. L.; MARCELINO-GUIMARÃES, F. C.; ABDELNOOR, R. V.; PEREIRA, G. A. G.; CARAZZOLLE, M. F. Genotipagem de cultivares brasileiros de soja através da detecção de SNPs. In: CONGRESSO BRASILEIRO DE SOJA, 6., 2012, Cuiabá. **Soja: integração nacional e desenvolvimento sustentável: anais**. Brasília, DF: Embrapa, 2012. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/63745/1/8-s340.pdf>>. Acesso em: 13 nov. 2016.

**Comunicado  
Técnico, 126**

**Embrapa Informática Agropecuária**  
Endereço: Av. Dr. André Tosello, 209 - Cidade  
Universitária, Campinas - SP  
Fone: (19) 3211-5700  
<https://www.embrapa.br/informatica-agropecuaria>

1ª edição publicação digital - 2017



MINISTÉRIO DA  
AGRICULTURA, PECUÁRIA  
E ABASTECIMENTO



**Comitê de  
publicações**

**Presidente:** *Giampaolo Queiroz Pellegrino*  
**Secretário-Executivo:** *Carla Cristiane Osawa*  
**Membros:** *Adriana Farah Gonzales, Carla Geovana do Nascimento Macário, Flávia Bussaglia Fiorini, Ivo Pierozzi Júnior, Kleber X. Sampaio de Souza, Luiz Antonio Falaguasta Barbosa, Maria Goretti G. Praxedes, Paula Regina K. Falcão, Ricardo Augusto Dante, Sônia Ternes*

**Expediente**

**Supervisão editorial:** *Kleber X. Sampaio de Souza*  
**Revisão de texto:** *Adriana Farah Gonzales*  
**Normalização bibliográfica:** *Maria Goretti G. Praxedes*  
**Editoração eletrônica:** *Tuíra Santana Favarin, sob supervisão de Flávia Bussaglia Fiorini.*