

# Construção de um repositório de dados do programa de melhoramento de arroz da Embrapa

Ana Carolina Simões Silva<sup>1</sup>, Flávio Breseghello<sup>2</sup>, Orlando Peixoto de Moraes<sup>3</sup>, Sérgio Lopes Junior<sup>4</sup>

## Introdução

Em programas de melhoramento de plantas, os dados de ensaios de campo são essenciais para a escolha das linhagens que apresentam características agronômicas mais favoráveis.

Os dados antigos encontram-se arquivados em papel ou arquivos digitais, em microcomputadores ou em pastas compartilhadas em rede local, ou ainda em disquetes ou CDs. Todavia, essa forma de armazenamento de dados apresenta riscos de perda, por danos à mídia ou exclusão acidental, além de dificultar o acesso a dados antigos.

Sistemas gerenciadores de documentos podem ser utilizados para garantir o armazenamento, localização, recuperação e controle de acesso a documentos, de forma simples e confiável.

Este resumo descreve a organização de um repositório de dados do melhoramento de arroz da Embrapa, utilizando-se o sistema gerenciador de documentos KnowledgeTree (www.knowledgetree.com).

## Material e métodos

O repositório de dados do melhoramento de arroz da Embrapa contém os resultados dos ensaios do programa de melhoramento de arroz irrigado e de terras altas, coordenado pela Embrapa Arroz e Feijão no período de 1983 a 2007. Tratam-se de arquivos texto, formatados para análise no pacote estatístico SAS (<arquivo.sas>), incluindo na maioria dos casos os comandos utilizados para sua análise. Estes arquivos ocupam pouco espaço de memória, e podem ser importados para planilhas eletrônicas se necessário.

Para criação deste repositório está sendo utilizado o software KnowledgeTree, desenvolvido pela Jam Warehouse e licenciado segundo termos da licença *KnowledgeTree Public License (open source e freeware)*. Este programa foi traduzido para a língua portuguesa do Brasil e o pacote de linguagem está disponibilizado para a comunidade. Uma tela deste sistema é apresentada na Figura 1.

Os documentos a serem inseridos no repositório são associados a metadados, baseados no cabeçalho dos arquivos (Tabela 1). Os metadados são usados para classificar e/ou informar sobre o conteúdo do documento, facilitando também o processo de busca. A

busca pode ser simples ou avançada, utilizando-se os metadados ou o próprio texto do arquivo. Na busca avançada é possível informar e associar diversos critérios para recuperar os documentos, tornando a busca mais precisa.

**Tabela 1.** Descrição dos metadados utilizados no banco de dados do melhoramento de arroz.

Metadado	Descrição
Sistema	Irigado ou terras altas
Categoria	Ensaio de rendimento, entomologia, fitopatologia, molecular, qualidade de grãos, recursos genéticos, etc.
Tipo	Ensaio de Valor de Cultivo e Uso (VCU), Preliminar (EP), Regional (ER), etc.
População	Elite, Seleção Recorrente, Germoplasma, etc.
Produção Média	Produção média de grãos, em kg/ha
CV	Coefficiente de variação da produção
CV (Classe)	Baixo, médio, alto e muito alto
Ano	Ano que o ensaio foi colhido
Estado	Estado da Federação
Município	Local de condução do ensaio

Até o presente, o repositório encontra-se na Intranet da Embrapa Arroz e Feijão e pode ser acessado, via navegador web, por qualquer usuário devidamente autenticado. Nele existem dois níveis de acesso: administrador, com atribuição de entrada e exclusão de arquivos e alteração de metadados; e usuários, os quais terão acesso apenas para leitura dos metadados, com opção de baixar os arquivos de dados.

Além do controle de acesso, metadados para descrição dos documentos e um serviço de busca para localizá-los, o repositório ainda fornece ferramentas para controle de versão e associação de documentos relacionados entre si. O uso destas ferramentas é previsto para uma fase posterior do trabalho.

## Resultados

O sistema de gerenciamento de dados do melhoramento de arroz da Embrapa está em fase de montagem. O trabalho teve início em agosto de 2006 e tem previsão de término em julho de 2007. Até janeiro de 2007, o sistema recebeu 1870

1. Bolsista da Embrapa Arroz e Feijão, CP 179, Santo Antônio de Goiás, GO, CEP 75375-000, [carol@cnpaf.embrapa.br](mailto:carol@cnpaf.embrapa.br)

2. Pesquisador da Embrapa Arroz e Feijão, [flavio@cnpaf.embrapa.br](mailto:flavio@cnpaf.embrapa.br)

3. Pesquisador da Embrapa Arroz e Feijão, [peixoto@cnpaf.embrapa.br](mailto:peixoto@cnpaf.embrapa.br)

4. Desenvolvedor de Software da Embrapa Arroz e Feijão, [sergio@cnpaf.embrapa.br](mailto:sergio@cnpaf.embrapa.br)

arquivos, o que corresponde a 62,3% da meta de 3000 arquivos retroativos a serem inseridos. Dos arquivos inseridos, aproximadamente 61,8% são dados de arroz de terras altas, indicando que a Embrapa possivelmente tenha dado mais ênfase ao melhoramento para este sistema do que para o sistema irrigado (Tabela 2). É importante ressaltar, no entanto, que os dados inseridos no sistema até o momento representam aqueles mais facilmente disponíveis.

**Tabela 2.** Número de ensaios inseridos no sistema de gerenciamento até janeiro de 2007, classificados por ano e sistema de cultivo.

Ano	Sistema		Total
	Irrigado	Terras Altas	
1982 a 1989	78	283	361
1990	15	0	15
1991	36	18	54
1992	37	37	74
1993	50	45	95
1994	38	29	67
1995	23	76	99
1996	63	60	123
1997	35	63	98
1998	61	56	117
1999	39	54	93
2000	9	74	83
2001	36	70	106
2002	35	22	57
2003	62	105	167
2004	54	82	136
2005	44	81	125
Total	715	1155	1870
	38,2%	61,8%	100%

## Discussão

Programas de melhoramento de plantas geram um grande volume de dados ao longo do tempo. Além da utilidade imediata para seleção de famílias e linhas, os dados podem ser usados em análises plurianuais. Para isso é necessário que os dados sejam mantidos de forma segura e acessível, e para isto os dados do melhoramento do arroz estão sendo organizados no repositório.

A coleção de dados de ensaios de arroz da Embrapa será útil, entre outros, para: 1) comprovação de valor de

cultivo e uso (VCU) de materiais testados em vários anos; 2) estudos de ganho genético do programa, em que se avalia a variação de um dado caráter ao longo do tempo, descontada a variação ambiental [1]; 3) estudos de estabilidade e adaptabilidade, em que genótipos são comparados quanto a aspectos da interação genótipo x ambiente [2]; 4) análise de associação entre genótipos e fenótipos para identificação de marcadores relacionados a caracteres agronômicos [3]; 5) modelagem ambiental para seleção de locais para condução de ensaios futuros e recomendação específica de cultivares para ambientes restritos [4]. Estes dados poderão servir como um recurso para pesquisas ao nível de pós-graduação em melhoramento de plantas.

Como trabalhos futuros, poderão ser investigadas e implantadas técnicas para agrupamento e extração de conjuntos parciais dos dados de ensaios, a partir dos documentos disponíveis no repositório, permitindo mais facilmente o uso dos mesmos. O presente sistema representa um avanço quanto à organização dos dados do melhoramento, e um primeiro passo em direção à montagem de um banco de dados com todas as funcionalidades oferecidas pelos modernos sistemas de busca e filtragem atualmente disponíveis.

## Agradecimentos

Agradecemos a toda a equipe do melhoramento de arroz que contribuiu para a geração dos dados dos ensaios.

## Referências

- [1] BRESEGHELLO, F., MORAIS, O. P., RANGEL, P. H. N. (1998). A new method to estimate genetic gain in annual crops. *Genet. Mol. Biol.* 21, 551-555.
- [2] DUARTE, J. B., ZIMMERMANN, M. J. O. (1992). Comparison of three methods used for the study of adaptation and phenotypic stability in the common bean (*Phaseolus vulgaris* L.). *Revista Brasileira de Genética* 15, 125-136.
- [3] BRESEGHELLO, F., SORRELLS, M. E. (2006). Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Sci.* 46, 1323-1330.
- [4] ANNICCHIARICO, P., BELLAH, F., CHIARI, T. 2005. Defining subregions and estimating benefits for a specific-adaptation strategy by breeding programs: a case study. *Crop Sci.* 45, 1741-1749.

The screenshot displays the KnowledgeTree Document Management System interface. At the top left is the Embrapa logo, and at the top right is the KnowledgeTree logo with the text 'Document Management System'. Below the logos, a navigation bar shows 'panel' and 'Navegar Documentos', along with the user name 'Flávia Braseguella' and options for 'Preferências' and 'Sair'. The breadcrumb path indicates the current location: 'você está aqui: navegar > pasta > melhoramento de arroz > resultados de ensaios > 2004'.

On the left side, there are two panels. The top panel, 'Ações da Pasta', contains links for 'Adicionar Documento', 'Adicionar uma Pasta', 'Renomear', 'Upload volume', 'Transições de pasta', 'Exportar Volume', and 'View Tree'. The bottom panel, 'Pesquisar', features a search input field, a 'pesquisar' button, and a link for 'Busca avançada'.

The main content area, titled 'Navegar Documentos', displays a table of documents. The table has columns for 'Título', 'Criado', 'Modificado', 'Criador', and 'Estado Atual'. Each row represents a document with a checkbox, a file icon, a title (e.g., 'EMC04320 (8Kb)'), a size, creation and modification timestamps, the creator's name ('Ana Carolina Simees'), and a status indicator ('—').

<input type="checkbox"/>	Título		Criado	Modificado	Criador	Estado Atual
<input type="checkbox"/>	<a href="#">EMC04320</a> (8Kb)		2006-07-28 09:59	2006-07-28 10:22	Ana Carolina Simees	—
<input type="checkbox"/>	<a href="#">EMC04324</a> (16Kb)		2006-07-28 10:39	2006-07-28 10:39	Ana Carolina Simees	—
<input type="checkbox"/>	<a href="#">EMC04325</a> (7Kb)		2006-07-28 10:40	2006-07-28 10:40	Ana Carolina Simees	—
<input type="checkbox"/>	<a href="#">EMC04326</a> (8Kb)		2006-07-28 10:42	2006-07-28 10:42	Ana Carolina Simees	—
<input type="checkbox"/>	<a href="#">EMC04327</a> (15Kb)		2006-07-28 10:44	2006-07-28 10:44	Ana Carolina Simees	—
<input type="checkbox"/>	<a href="#">EMC04328</a> (13Kb)		2006-07-28 10:45	2006-07-28 10:45	Ana Carolina Simees	—
<input type="checkbox"/>	<a href="#">EMC04329</a> (20Kb)		2006-07-28 10:47	2006-07-28 10:47	Ana Carolina Simees	—
<input type="checkbox"/>	<a href="#">EMC04330</a> (8Kb)		2006-07-28 10:48	2006-07-28 10:48	Ana Carolina Simees	—
<input type="checkbox"/>	<a href="#">EMC04331</a> (49Kb)		2006-07-28 10:11	2006-07-28 10:23	Ana Carolina Simees	—
<input type="checkbox"/>	<a href="#">EMC04332</a> (22Kb)		2006-07-28 10:12	2006-07-28 10:24	Ana Carolina Simees	—
<input type="checkbox"/>	<a href="#">EMC04333</a> (7Kb)		2006-07-28 10:15	2006-07-28 10:24	Ana Carolina Simees	—
<input type="checkbox"/>	<a href="#">EMC04334</a> (4Kb)		2006-08-02 14:05	2006-08-02 14:05	Ana Carolina Simees	—
<input type="checkbox"/>	<a href="#">EMC04335</a> (3Kb)		2006-07-28 10:18	2006-07-28 10:25	Ana Carolina Simees	—

Figura 1. Imagem da tela do computador mostrando o ambiente do sistema de informação KnowledgeTree, com ensaios de arroz relacionados e opções para visualização de metadados ou para baixar os arquivos texto contendo os dados originais.