



012

REGISTROS DUPLICADOS EM BASES DE DADOS: O CASO DA EMBRAPA¹

Elizabeth Denise Câmara Trevisan²

As Bases de Dados da Pesquisa Agropecuária (BDPA) têm por objetivo contribuir para o uso e a disseminação da informação técnico-científica produzida e colecionada pelas Bibliotecas das Unidades de Pesquisa da Embrapa. É composta por três bases de dados de informações documentais: Acervo Documental, Coleção de Periódicos e Produção Técnico-Científica da Embrapa. Pode ser acessada pela página da Biblioteca da *Embrapa Florestas* (<http://www.cnpf.embrapa.br/biblio/basedado.htm>), sendo de domínio público. Contava, em novembro de 2006, com 436.373 registros válidos, gerados pela junção dos arquivos de todas as Bibliotecas que fazem parte do Sistema Embrapa de Bibliotecas. Nesse total estão incluídos os registros duplicados, que podem chegar a 25% desse total. O Grupo de Trabalho de Referência – GT Refer, entre outras atribuições, tem por objetivo diminuir o número dessas duplicações. O objetivo deste trabalho é avaliar a metodologia atualmente utilizada pelo GT Refer. Existem três campos, que se não estiverem absolutamente iguais, vão gerar duplicação de registros na BDPA, sendo eles autoria, título e ano de publicação. As duplicações acontecem essencialmente por erros de digitação e/ou divergência na descrição bibliográfica dos documentos. As correções dos registros duplicados por erros de digitação são solicitadas basicamente pelo Grupo de Trabalho, por correio eletrônico, às Bibliotecas responsáveis por essas duplicações. As outras duplicações são geradas por diferenças no campo autoria, onde se descreve o ponto de acesso principal do documento. Esses casos são analisados pelo GT Refer para posterior solicitação de correção. Por diversas razões, nem sempre essas solicitações são atendidas. Por esse fato e pelo grande número de registros já duplicados somado às novas duplicações que serão geradas, essa metodologia não está sendo eficiente na eliminação dos registros duplicados. A melhor solução para o problema seria uma nova versão do Ainfo em Sistema Web, como levantado em discussões anteriores por bibliotecários da Embrapa. Esse sistema *online* possibilitaria correções em tempo real e evitaria re-trabalho. Como melhoria parcial, sugere-se que no caso de mais de um registro com um mesmo ISBN, o sistema gere relatórios com subsequente envio automático de mensagem eletrônica. Esses registros ficariam apartados da BDPA até que fossem devidamente corrigidos.

¹Trabalho realizado na *Embrapa Florestas*

²Analista da *Embrapa Florestas*, beth@cnpf.embrapa.br