

O problema da padronização das afiliações de autores na base de dados *Web of Science*: o caso Embrapa e sua solução

Roberto de Camargo Penteado Filho

Doutor; Empresa Brasileira de Pesquisa Agropecuária, Brasília, DF, Brasil;
roberto.penteado@embrapa.br

Wilson Corrêa da Fonseca Júnior

Doutor; Empresa Brasileira de Pesquisa Agropecuária, Brasília, DF, Brasil;
wilson.fonseca@embrapa.br

Resumo: A produção científica e tecnológica é fundamental no processo de inovação de um país ou instituição científica. No entanto, a avaliação bibliométrica desse desempenho enfrenta há muitos anos um grande desafio: o problema da falta de exatidão das informações registradas em documentos científicos e bases de dados. Este artigo possui como principal objetivo verificar em que medida esse problema ainda persiste na produção científica brasileira, por meio de investigações bibliográfica e bibliométrica. Nesse segundo caso, foi realizado, como exemplo, um estudo básico sobre a visibilidade da Empresa Brasileira de Pesquisa Agropecuária na base de dados *Web of Science*. O resultado demonstrou uma grande incidência de problemas de padronização, refletidos no percentual de 11,93% de variações relacionadas ao nome e sigla originais da empresa. Diante desse resultado, a instituição publicou uma resolução normativa interna regulamentando a afiliação de seus empregados em publicações nacionais e internacionais.

Palavras-chave: Afiliação institucional. Padronização. Embrapa. *Web of Science*. Política de comunicação.

1 Introdução

A falta de exatidão de informações em documentos científicos e bases de dados se encontra entre os principais problemas do processo de avaliação da produção científica e tecnológica de um país, instituição ou pesquisador. Esse problema se deve a diversos fatores, tais como a presença de erros ortográficos ou de impressão, equívocos na classificação de dados e ausência de padronização de informações institucionais, entre outros.

Diversos desses fatores vêm sendo registrados na literatura pelo menos desde a década de 1970, quando Bourne (1977), por exemplo, chamou a atenção para a frequência e o impacto dos erros de ortografia em bases de dados bibliográficas. Naquela época, ao analisar cerca de 3.600 termos indexados em 11 diferentes bases de dados ele encontrou a presença de erros ortográficos com variação de 23% a menos de 0,5% entre uma base e outra. De acordo com Hood e Wilson (2003), essas ocorrências ainda se fazem presentes porque as bases de dados foram criadas originalmente com a finalidade principal de recuperação de informações, não de realização de estudos bibliométricos.

Atualmente, entre os principais problemas que dificultam a correta avaliação da produção científica de um país, instituição ou pesquisador se encontra a ausência de padronização dos nomes de autores e de suas afiliações em bases de dados em ciência e tecnologia (C&T). Em trabalho recente de análise sobre a produção científica da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), Penteadó Filho, Fonseca Júnior e Avila (2016) também se empenharam no tratamento dos principais fatores associados a esse problema. Os resultados desse levantamento realizado na base de dados *Web of Science* (WoS), tendo como referência teórico-metodológica a literatura especializada, se encontram sintetizados neste artigo.

2 Problemas e dificuldades de padronização nos estudos bibliométricos

Os principais problemas e dificuldades encontrados na extração de informações úteis em bases de dados eletrônicas para a realização de estudos bibliométricos são fartamente abordados pela literatura especializada (BOURNE, 1977; HOOD; WILSON, 2003; TAŞKIN; AL, 2013). Hood e Wilson (2003), por exemplo, subdividiram esses problemas e dificuldades em três categorias. A primeira delas se refere aos problemas de erro ou de falta de consistência do dado (nível micro), tais como a presença de erros ortográficos, de indexações equivocadas de palavras-chave, de variações na abreviatura de autores ou de datas, duplicidade de títulos de periódicos e diferentes traduções do nome original de determinada instituição. O segundo grupo se refere a problemas relacionados ao uso de bases

de dados para propósitos bibliométricos (nível macro). Nesse caso, as principais ocorrências são a sobreposição de informações com a duplicação de registros, a limitação da cobertura geográfica, temporal ou temática da base de dados, o atraso na indexação da literatura recente, a ausência de campos relevantes de busca e a falta de padronização de dados, entre outras. O terceiro grupo de dificuldades se refere a problemas derivados das próprias bases, tais como a disponibilidade de ferramentas amigáveis, porém inadequadas à análise bibliométrica ou a impossibilidade de realização de estudos *offline*.

Diversos desses problemas e dificuldades foram encontrados por Penteado Filho (2006) ao realizar a auditoria do acervo da produção científica da Embrapa com a utilização de diversos *softwares* (Infotrans, Dataview, Matrisme e MS Excel). Entre os problemas detectados estavam erros de catalogação, ortográficos e de digitação, duplicações de registros e presença de autores homônimos. Mais recentemente, ao abordar o problema da padronização das afiliações de autores em índices de citação, Taşkin e Al (2014) analisaram na base de dados *Web of Science* todos os tipos de documentos (artigos, anais de congressos ou cartas), publicados entre 1928 e 2009, que mencionassem nos campos “endereço” e “país” a Turquia como país de origem. Entre os principais resultados desse trabalho encontra-se uma tabela contendo as 20 universidades turcas mais produtivas, com seus respectivos números de publicações, quantidade de erros de indexação e percentual desses erros em relação ao total de publicações. Nesse caso, os percentuais variaram de 12,1% a 0,3% entre uma instituição e outra e contemplaram diversos tipos de problemas, tais como: erros de caracteres ou de ortografia, erros de indexação, erros de tradução realizados pelos autores, além de problemas de padronização dos endereços das universidades.

3 Metodologia

Este levantamento reuniu a produção científica da Embrapa nas bases de dados *Science Citations Index Expanded* (SCI-EXPANDED), *Social Science Citation Index* (SSCI) e *Arts & Humanities Citation Index* (A&HCI) da *Thomson Scientific* (ISI), conhecidas sob a denominação de *Web of Science* (WoS). Para isso foram

considerados todos os registros, em diversas línguas e tipos de documentos (artigos, editoriais, resenhas, entre outros) com as possíveis denominações que fizessem referência à Embrapa no campo “Afiliação de autor” no período compreendido entre 1973 e 17 de julho de 2015.

A busca teve início pela expressão mais simples. Depois, foram sendo incorporadas outras variações de acordo com as inconsistências encontradas nos artigos referentes ao nome da Embrapa ou de seus centros de pesquisa. O acréscimo dessas variações elevou o número de registros de 15.956 para 17.794 entre a primeira e terceira etapas. A quarta etapa contemplou uma investigação sobre registros únicos dos centros de pesquisa da Empresa (324 registros), o que aumentou o resultado final para 18.118 registros. A diferença entre esse total e o número de registros da primeira busca (15.956) é de 2.162 registros. Isso significa que, se a busca sobre a produção científica da Embrapa se limitasse aos termos originais adotados na primeira etapa, a sigla/nome da instituição, 11,93% dos documentos não estariam contemplados na sua produção científica. Veja abaixo a evolução das estratégias de busca no item “pesquisa avançada” da WoS do nome da Embrapa e seus resultados:

- a) 1. AD=(Embrapa) OR AD=(EMBRAPA) — Registros: 15.956 (17/07/2015);

- b) 2. AD=(EMBRAPA) OR AD=(EMPRESA BRASILEIRA PESQUISA AGROPECUARIA) OR AD=(Brazilian Org Agr Res) OR AD=(BRAZILIAN ENTERPRISE AGR RES) OR AD=(Brazilian Agr Res Corp) OR AD=(BRAZILIAN AGR RES ENTERPRISE) OR AD=(Brazilian Enterprise Agropecuary) OR AD=(Brazilian Agropecuary Res Corp) OR AD=(BRAZILIAN ORG AGR RES) OR AD=(BRAZILIAN AGR RES CORP) OR AD=(BRAZILIAN ENTERPRISE AGROPECUARY) OR AD=(BRAZILIAN AGROPECUARY RES CORP) — Registros: 17.438 (17/07/2015);

c) 3. AD=(Ambrapa) OR AD=(AMBRAPA) OR AD=(Brazilian Agr Res Corp) OR AD=(BRAZILIAN AGR RES CORP) OR AD=(BRAZILIAN AGR RES ENTERPRISE) OR AD=(Brazilian Agr Res Enterprise) OR AD=(Brazilian Agropecuary Res Corp) OR AD=(BRAZILIAN AGROPECUARY RES CORP) OR AD=(Brazilian Corp Agr) OR AD=(BRAZILIAN CORP AGR) OR AD=(Brazilian Corp Agr Res) OR AD=(BRAZILIAN CORP AGR RES) OR AD=(BRAZILIAN ENTERPRISE AGR RES) OR AD=(Brazilian Enterprise Agr Res) OR AD=(Brazilian Enterprise Agropecuary) OR AD=(BRAZILIAN ENTERPRISE AGROPECUARY) OR AD=(Brazilian Org Agr Res) OR AD=(BRAZILIAN ORG AGR RES) OR AD=(Brazilian Res Agr Corp) OR AD=(BRAZILIAN RES AGR CORP) OR AD=(Einbrapa) OR AD=(EINBRAPA) OR AD=(Embapra) OR AD=(EMBAPRA) OR AD=(Embarapa) OR AD=(EMBRA) OR AD=(Embra) OR AD=(Embraba) OR AD=(EMBRABA) OR AD=(Embraoa) OR AD=(EMBRAOA) OR AD=(Embrapa) OR AD=(EMBRAPA) OR AD=(EMBRAPA* NOT EMBRAPA) OR AD=(EMBRAPA* NOT EMBRAPII) OR AD=(Embrape) OR AD=(EMBRAPE) OR AD=(Embrapo) OR AD=(EMBRAPO) OR AD=(Embrara) OR AD=(EMBRARA) OR AD=(Embrpa) OR AD=(EMBRPA) OR AD=(EMBTAPA) OR AD=(Embtapa) OR AD=(Empbrapa) OR AD=(EMPBRAPA) OR AD=(Empera Brasileira Pesquisas Agropecuaria) OR AD=(EMPERA BRASILEIRA PESQUISAS AGROPECUARIA) OR AD=(Empersa) OR AD=(Empraba) OR AD=(EMPRABA) OR AD=(EMPRAPA) OR AD=(Emprapa) OR AD=(Emprea Brasileira Pesquisa Agropecuaria) OR AD=(EMPREA BRASILEIRA PESQUISA AGROPECUARIA) OR AD=(EMPRESA BRASILEIRA DE PESQUISA AGROPECUARIA) OR AD=(Empresa Brasileira de Pesquisa Agropecuaria) OR AD=(Empresa Brasileira Pedquisa Agropecuaria) OR AD=(EMPRESA BRASILEIRA PESQUISA AGR) OR AD=(Empresa Brasileira Pesquisa Agr) OR AD=(Empresa Brasileira Pesquisa Agr* NOT Empresa Brasileira Pesquisa Agr) OR AD=(Empresa Brasileira Pesquisa

Agropecuaria) OR AD=(EMPRESA BRASILEIRA PESQUISA AGROPECUARIA) OR AD=(Empresa Brasileira Pesquisa Agropecuarias) OR AD=(EMPRESA BRASILEIRA PESQUISA AGROPECUARIAS) OR AD=(Empresa Brasileira Pesquisa Agropecuciria) OR AD=(EMPRESA BRASILEIRA PESQUISA AGROPECUCIRIA) OR AD=(Empresa Brasileira Pesquisa Agropecudria) OR AD=(EMPRESA BRASILEIRA PESQUISA AGROPECUDRIA) OR AD=(Empresa Brasileira Pesquisa Agropecuraria) OR AD=(EMPRESA BRASILEIRA PESQUISA AGROPECURARIA) OR AD=(EMPRESA BRASILEIRA PESQUISA AGROPECURIA) OR AD=(Empresa Brasileira Pesquisa Agropecuria) OR AD=(Empresa Brasileira Pesquisas Agr) OR AD=(EMPRESA BRASILEIRA PESQUISAS AGR) OR AD=(Empresa Brasileira Pesquisa Agropecuaria) OR AD=(EMPRESA BRASILERA PESQUISA AGROPECUARIA) OR AD=(Empresa Brasileria Pesquisa Agr) OR AD=(Empresa brasileria Pesquisa Agropecuaria) OR AD=(EMPRESA BRASILERIA PESQUISA AGROPECUARIA) OR AD=(Empresa Brasileira Pesquisa Agr) OR AD=(EMPRESA BRASILIERA PESQUISA AGR) OR AD=(Empresa Brasilleira Pesquisa Agropecuaria) OR AD=(EMPRESA BRASILLEIRA PESQUISA AGROPECUARIA) OR AD=(Empresa Brasileira Pesquisa Agropecuria) OR AD=(EMPRESA BRASILEIRA PESQUISA AGROPECURIA) OR AD=(Empress Brasileira Pesquisa Agropecuaria) OR AD=(EMPRESS BRASILEIRA PESQUISA AGROPECUARIA) OR AD=(Ernbrapa) OR AD=(ERNBRAPA) — Registros: 17.794 (17/07/2015);

- d) 4. (“Ctr Nacl Pesquisa & Desenvolvimento Instrumentaca” OR “Ctr Nacl Pesquisa & Gado Corte” OR “Ctr Nacl Pesquisa Agrobiol” OR “Ctr Nacl Pesquisa Algodao” OR “Ctr Nacl Pesquisa Arroz & Feijao” OR “Ctr Nacl Pesquisa Caprinos” OR “Ctr Nacl Pesquisa Florestas” OR “Ctr Nacl Pesquisa Gado Leite” OR “Ctr Nacl Pesquisa Hort” OR “Ctr Nacl Pesquisa Milho & Sorgo” OR “Ctr Nacl Pesquisa Monitoramento &

Avaiacao Impact” OR “Ctr Nacl Pesquisa Recursos Genet & Biotecnol”
OR “Ctr Nacl Pesquisa Soja” OR “Ctr Nacl Pesquisa Solos” OR “Ctr Nacl
Pesquisa Suinos & Aves” OR “Ctr Nacl Pesquisa Tecnol Agroind
Alimentos CTAA” OR “Ctr Nacl Pesquisa Trigo” OR “Ctr Nacl Recursos
Genet & Biotecnol Cenargen” OR “Ctr Pesquisa Agr Trop Semiarido” OR
“Ctr Pesquisa Agrflorestal Amazonia Oriental CPATU” OR “Ctr Pesquisa
Agroflorestal Acre” OR “Ctr Pesquisa Agroflorestal Amazonia Ocidental”
OR “Ctr Pesquisa Agroflorestal Roraima CPAF” OR “Ctr Pesquisa
Agroflorestal Roraima” OR “Ctr Pesquisa Agropecuaria Clima
Temperado” OR “Ctr Pesquisa Agropecuaria Meio Norte” OR “Ctr
Pesquisa Agropecuaria Pantanaal” OR “Ctr Pesquisa Agropecuaria
Pantanal” OR “Ctr Pesquisa Agropecuaria Tabuleiros Costeiros” OR “Ctr
Pesquisa Agropecuaria Trop Semiarido” OR “Ctr Pesquisa Agropecuaria
Clima Temperado” OR “Cenargen” OR “CENARGEN” OR “CNPAB”
OR “CNPAF” OR “CNPGL” OR “CNPMF” OR “CNPMS” OR “CNPSO”
OR “CPA Pantanal” OR “CPAC” OR “CPAMN” OR “CPPSE”)

NOT

(“Ambrapa” OR “AMBRAPA” OR “Brazilian Agr Res Corp” OR
“BRAZILIAN AGR RES CORP” OR “BRAZILIAN AGR RES
ENTERPRISE” OR “Brazilian Agr Res Enterprise” OR “Brazilian
Agropecuaria Res Corp” OR “BRAZILIAN AGROPECUARY RES
CORP” OR “Brazilian Corp Agr” OR “BRAZILIAN CORP AGR” OR
“Brazilian Corp Agr Res” OR “BRAZILIAN CORP AGR RES” OR
“BRAZILIAN ENTERPRISE AGR RES” OR “Brazilian Enterprise Agr
Res” OR “Brazilian Enterprise Agropecuaria” OR “BRAZILIAN
ENTERPRISE AGROPECUARY” OR “Brazilian Org Agr Res” OR
“BRAZILIAN ORG AGR RES” OR “Brazilian Res Agr Corp” OR
“BRAZILIAN RES AGR CORP” OR “Einbrapa” OR “EINBRAPA” OR
“Embapra” OR “EMBAPRA” OR “Embarapa” OR “EMBRA” OR
“Embra” OR “Embraba” OR “EMBRABA” OR “Embraoa” OR
“EMBRAOA” OR “Embrapa” OR “EMBRAPA” OR “EMBRAPA* NOT
EMBRAPA” OR “Embrape” OR “EMBRAPE” OR “Embrapo” OR

“EMBRAPO” OR “Embrara” OR “EMBRARA” OR “Embrpa” OR
“EMBRPA” OR “EMBTAPA” OR “Embtapa” OR “Empbrapa” OR
“EMPBRAPA” OR “Empera Brasileira Pesquisas Agropecuaria” OR
“EMPERA BRASILEIRA PESQUISAS AGROPECUARIA” OR
“Empersa” OR “Empraba” OR “EMPRABA” OR “EMPRAPA” OR
“Emprapa” OR “Emprea Brasileira Pesquisa Agropecuaria” OR
“EMPREA BRASILEIRA PESQUISA AGROPECUARIA” OR
“EMPRESA BRASILEIRA DE PESQUISA AGROPECUARIA” OR
“Empresa Brasileira de Pesquisa Agropecuaria” OR “Empresa Brasileira
Pedquisa Agropecuaria” OR “EMPRESA BRASILEIRA PESQUISA
AGR” OR “Empresa Brasileira Pesquisa Agr” OR “Empresa Brasileira
Pesquisa Agr* NOT Empresa Brasileira Pesquisa Agr” OR “Empresa
Brasileira Pesquisa Agropecuaria” OR “EMPRESA BRASILEIRA
PESQUISA AGROPECUARIA” OR “Empresa Brasileira Pesquisa
Agropecuarias” OR “EMPRESA BRASILEIRA PESQUISA
AGROPECUARIAS” OR “Empresa Brasileira Pesquisa Agropecuciria”
OR “EMPRESA BRASILEIRA PESQUISA AGROPECUCIRIA” OR
“Empresa Brasileira Pesquisa Agropecudria” OR “EMPRESA
BRASILEIRA PESQUISA AGROPECUDRIA” OR “Empresa Brasileira
Pesquisa Agropecuraria” OR “EMPRESA BRASILEIRA PESQUISA
AGROPECURARIA” OR “EMPRESA BRASILEIRA PESQUISA
AGROPECURIA” OR “Empresa Brasileira Pesquisa Agropecuria” OR
“Empresa Brasileira Pesquisas Agr” OR “EMPRESA BRASILEIRA
PESQUISAS AGR” OR “Empresa Brasileira Pesquisa Agropecuaria” OR
“EMPRESA BRASILERA PESQUISA AGROPECUARIA” OR
“Empresa Brasileria Pesquisa Agr” OR “Empresa brasileria Pesquisa
Agropecuaria” OR “EMPRESA BRASILERIA PESQUISA
AGROPECUARIA” OR “Empresa Brasiliera Pesquisa Agr” OR
“EMPRESA BRASILIERA PESQUISA AGR” OR “Empresa Brasilleira
Pesquisa Agropecuaria” OR “EMPRESA BRASILLEIRA PESQUISA
AGROPECUARIA” OR “Empresa Brasileira Pesquisa Agropecuria” OR
“EMPRESA BRASILEIRA PESQUISA AGROPECURIA” OR “Empress

Brasileira Pesquisa Agropecuária” OR “EMPRESS BRASILEIRA PESQUISA AGROPECUARIA” OR “Embrapa” OR “ERNBRAPA”) — Registros: 324 (22/07/2015).

Resultado geral da busca: $17.794 + 324 = 18.118$

Percentual de variações não contempladas na busca inicial: 11,93%

4 Análise e discussão dos resultados

A primeira parte da busca permitiu estabelecer uma lista visando codificar as expressões adotadas pelos autores em seus artigos científicos e incorporadas pela base, na forma de assinaturas-síntese criadas em 1996 pela Embrapa. Nessa mesma ocasião também foi adotada a denominação “*Brazilian Agricultural Research Corporation*” para o seu nome em inglês. A criação de assinaturas-síntese foi a forma encontrada pela instituição para associar o nome e a sigla específica de cada um de seus centros de pesquisa à sigla Embrapa e se encontra entre as principais medidas de sua Política de Comunicação (EMBRAPA, 2002). Por exemplo, o Centro de Pesquisa Agropecuária do Pantanal (CPAP) e o Centro Nacional de Pesquisa em Gado de Leite (CNPGL) se tornaram, respectivamente, com a criação da assinatura-síntese, Embrapa Pantanal e Embrapa Gado de Leite. Esse trabalho de decodificação permitiu posteriormente a realização de uma proposta de solução para os problemas de padronização do nome da Embrapa em documentos científicos. Diante do exposto, apresentamos no Quadro 1 essa lista de decodificação.

Quadro 1 – Decodificação das expressões de busca encontradas na base *Web of Science* nas assinaturas-síntese dos centros da Embrapa

EXPRESSÕES	ASSINATURAS-SÍNTESE
Ctr Nacl Pesquisa Arroz & Feijao	Embrapa Arroz e Feijão
Ctr Nacl Pesquisa Trigo	Embrapa Trigo
Ctr Nacl Pesquisa Soja	Embrapa Soja
Ctr Nacl Pesquisa Suínos & Aves	Embrapa Suínos e Aves
Ctr Nacl Pesquisa Milho & Sorgo	Embrapa Milho e Sorgo
Ctr Nacl Pesquisa Algodao	Embrapa Algodão
Ctr Nacl Pesquisa Gado Leite	Embrapa Gado de Leite
Ctr Nacl Pesquisa & Gado Corte	Embrapa Gado de Corte

EXPRESSÕES	ASSINATURAS-SÍNTESE
Ctr Nacl Pesquisa Caprinos	Embrapa Caprinos e Ovinos
Ctr Nacl Pesquisa Recursos Genet & Biotecnol	Embrapa Recursos Genéticos e Biotecnologia
Ctr Nacl Pesquisa Agrobiol	Embrapa Agrobiologia
Ctr Pesquisa Agropecuaria Pantanal	Embrapa Pantanal
Ctr Nacl Pesquisa Hort	Embrapa Horticultura
Ctr Pesquisa Agropecuaria Trop Semiarido	Embrapa Semiárido
Ctr Nacl Pesquisa Solos	Embrapa Solos
Ctr Nacl Pesquisa Florestas	Embrapa Florestas
Ctr Pesquisa Agroflorestal Acre	Embrapa Acre
Ctr Pesquisa Agroflorestal Amazonia Ocidental	Embrapa Amazônia Ocidental
Ctr Pesquisa Agropecuaria Tabuleiros Costeiros	Embrapa Tabuleiros Costeiros
Ctr Pesquisa Agropecuaria Clima Temperado	Embrapa Clima Temperado
Ctr Pesquisa Agroflorestal Roraima	Embrapa Roraima
Ctr Nacl Pesquisa & Desenvolvimento Instrumentaca	Embrapa Instrumentação
Ctr Pesquisa Agr Trop Semiarido	Embrapa Semiárido
Ctr Nacl Pesquisa Tecnol Agroind Alimentos CTAA	Embrapa Agroindústria de Alimentos
Ctr Pesquisa Agropecuaria Meio Norte	Embrapa Meio Norte
Ctr Nacl Recursos Genet & Biotecnol Cenargen	Embrapa Recursos Genéticos e Biotecnologia
Ctr Pesquisa Agropecuaria Clima Temperado	Embrapa Clima Temperado
Ctr Nacl Pesquisa Monitoramento & Avaiacao Impact	Embrapa Monitoramento por Satélite
Ctr Pesquisa Agrflorestal Amazonia Oriental CPATU	Embrapa Amazônia Oriental
Ctr Pesquisa Agroflorestal Roraima CPAF	Embrapa Roraima
Ctr Pesquisa Agropecuaria Pantanaal	Embrapa Pantanal
Cenargen	Embrapa Recursos Genéticos e Biotecnologia
CENARGEN	Embrapa Recursos Genéticos e Biotecnologia
CNPAB	Embrapa Agrobiologia
CNPAF	Embrapa Arroz e Feijão
CNPGL	Embrapa Gado de Leite
CNPMF	Embrapa Mandioca e Fruticultura
CNPMS	Embrapa Milho e Sorgo
CNPSo	Embrapa Soja
CPA Pantanal	Embrapa Pantanal
CPAC	Embrapa Cerrados
CPAMN	Embrapa Meio-Norte
CPPSE	Embrapa Pecuária Sudeste

Fonte: Dados da pesquisa.

Outra iniciativa adotada neste estudo visando a posterior padronização do nome da Embrapa em bases de dados foi selecionar uma amostra das expressões recolhidas a partir dos mecanismos de busca para levantar a origem e os principais tipos de problemas de padronização encontrados. Essa investigação exploratória indicou que grande parte dos problemas estava na **informação sobre a**

instituição fornecida pelo autor, ou ainda na redução do nome completo da instituição pela base de dados ou mesmo no cadastramento equivocado do nome da instituição pela base. Essas diversas formas equivocadas de menção ao nome da Embrapa, bem como sua origem e descrição dos principais tipos de problema são apresentados no Quadro 2.

Quadro 2 – Levantamento preliminar dos problemas das expressões relacionadas à Embrapa na base *Web of Science*.

EXPRESSÃO	OCORRÊNCIAS	ORIGEM E TIPO DE PROBLEMA
Ambrapa	1	Erro de cadastramento da base
Brazilian Agr Res Corp	440	Erro de informação pelo autor e nome resumido pela base
BRAZILIAN AGR RES ENTERPRISE	6	Erro de informação pelo autor
Brazilian Agropecuary Res Corp	1	Erro de informação pelo autor
Brazilian Corp Agr Res	13	Erro de informação pelo autor
BRAZILIAN ENTERPRISE AGR RES	30	Erro de informação pelo autor
Brazilian Enterprise Agropecuary	1	Erro de informação pelo autor
Brazilian Org Agr Res	4	Nome resumido pela base
Ctr Nacl Pesquisa	7	Erro de informação pelo autor e nome reduzido pela base
Ctr Pesquisa Agr*	671	Erro de informação pelo autor e nome reduzido pela base
Einbrapa	1	Erro de cadastramento pela base
Embapra	1	Erro de digitação pelo autor
Embarapa	2	Erro de cadastramento pela base
Embra	5	Erro de digitação pelo autor
Embraba	3	Erro de digitação pelo autor e de cadastramento pela base
Embraoa	1	Erro de cadastramento pela base
Embrap	9	Nome reduzido pela base
Embrape	1	Erro de informação pelo autor e de cadastramento pela base
Embrapo	2	Erro de cadastramento pela base

EXPRESSÃO	OCORRÊNCIAS	ORIGEM E TIPO DE PROBLEMA
Empera Brasileira Pesquisas Agropecuaria	1	Erro de informação pelo autor
Empersa Brasileira Pesquisa Agropecuaria	1	Erro de informação pelo autor e de cadastramento pela base
EMPRAPA	26	Erro de informação pelo autor
Empresa Brasileira Pedquisa Agropecuaria	1	Erro de cadastramento pela base
Empresa Brasileira Pesquisa Agr	610	Redução do nome pela base
Empresa Brasileira Pesquisa Agropecuaria	1.684	Redução do nome pela base
Empresa Brasileira Pesquisa Agropecuciria	2	Erro de cadastramento pela base
Empresa Brasileira Pesquisa Agropecudria	3	Erro de cadastramento pela base
Empresa Brasileira Pesquisa Agropecuraria	1	Erro de cadastramento pela base
Empresa Brasileira Pesquisa Agropecuria	30	Erro de cadastramento pela base
Empresa Brasileira Pesquisas Agr	8	Erro de informação pelo autor e de redução do nome pela base
Empresa Brasileira Pesquisa Agropecuaria	3	Erro de cadastramento pela base
Empresa brasileria Pesquisa Agropecuaria	12	Erro de cadastramento pela base
Empresa Brasileira Pesquisa Agr	1	Erro de informação pelo autor e de redução do nome pela base
Empresa Brasileira Pesquisa Agropecuaria	1	Erro de cadastramento pela base
Ernbrapa	3	Erro de cadastramento pela base

Fonte: Dados da pesquisa.

Conforme constatado, os problemas da falta de exatidão das informações encontrados nessa busca sobre a produção científica da Embrapa se devem a vários fatores, que vão desde a digitação errônea da sigla da Empresa pelo autor ou pela base de dados, passando pela ausência de uniformização do nome da instituição em português e outras línguas, até a ausência da sigla e/ou do nome da instituição nesses documentos. Apresentamos a seguir um resumo das inconsistências encontradas:

- a) erros de digitação de dados ou de redação no próprio documento científico.

Exemplos: Embapra / Embarapa / Embraba / EMPRAPA;

- b) erros de digitação ou de digitalização de dados do documento científico pela base de dados. Exemplos: Ambrapa / Einbrapa / Embraoa / Empresa brasileira Pesquisa Agropecuciria / Empresa Brasileira Pesquisa Agropecuraria;
- c) erros de digitação e/ou de digitalização de dados no documento científico e na base de dados. Exemplos: Empresa Brasileira de Pesquisas Agropecuária (artigo) / Empresa Brasileira (de) Pesquisas Agropecuária (base de dados);
- d) redução do nome original por extenso da Embrapa pela base de dados. Exemplos: Empresa Brasileira Pesquisas Agr / *Brazilian Agr Res Corp*;
- e) omissão da sigla e/ou nome da instituição no documento científico. Exemplos: Ctr Nacl Pesquisa & Desenvolvimento / Ctr Nacl Pesquisa;
- f) prevalência, no documento científico, da sigla e do nome da unidade e/ou departamento em detrimento da sigla e do nome da Embrapa. Exemplo: Laboratório de Análise do Solo, Centro Nacional de... (sigla da unidade);
- g) existência nos documentos científicos de diversidade de nomes em inglês (ou outras línguas) para a Empresa. Exemplos: *Brazilian Agricultural Research Enterprise* / *Brazilian Corporation of Agricultural Research* / *Brazilian Enterprise for Agricultural Research*;
- h) coexistência no documento científico da identificação da Empresa em português e inglês (ou outras línguas). Exemplo: *Laboratory of bacteriology*, Empresa Brasileira de Pesquisa Agropecuária;
- i) adoção no documento científico de nomes e abreviações genéricas de unidades ou departamentos da Empresa, em português, inglês e outras línguas, que dificultam a busca em base de dados. Exemplos: Natl Ctr / Natl Res Ctr / Ctr Nacl / Ctr Nacl Pesq;
- j) possibilidade de confusão entre as siglas Embrapa (Empresa Brasileira de Pesquisa Agropecuária) e Embrapii (Empresa Brasileira de Pesquisa e Inovação Industrial). Exemplo: Embrap – Empresa Brasileira de Pesquisa e Inovação Industrial – Embrapii / Embrap – Empresa Brasileira de Pesquisa Agropecuária.

4.1. Um problema exclusivo da base de dados

Além dos problemas de incorreção relacionados ao nome e à sigla das instituições, este estudo também constatou a ocorrência de outro grave problema nos resultados do campo de afiliação de autores da base de dados *Web of Science*: erros introduzidos pelo algoritmo de padronização da própria base para siglas e endereços.

Alguns autores informam seus endereços brasileiros sem citar o nome do país. Seus endereços de afiliação, portanto, terminam na sigla do estado. Aparentemente, a WoS criou um algoritmo para introdução de nomes de países e estados quando os autores são omissos. Dessa forma, a sigla ES, adotada no Brasil para abreviar o nome do estado do Espírito Santo, está sendo interpretada pela base como sigla da Espanha; a sigla BA, relativa ao estado da Bahia, é interpretada pela base como Buenos Aires, e o algoritmo da base chega a acrescentar “, Argentina”; ou então Bahia, BA é alterado pela WoS para “Bahia Blanca, Buenos Aires, Argentina”. Endereços que terminam na sigla SC do estado de Santa Catarina estão sendo transpostos para “South Carolina, USA”. A cidade de Santiago de Cuba é associada à Espanha e a de Cachoeiro de Itapemirim está sendo transferida para Belize (nesse caso a sigla encontrada foi BZ) e Vitória para a Espanha. Problemas como esse originados na própria editora da base acabam gerando trabalho adicional na pesquisa bibliométrica e a perda de eficiência das buscas geográficas.

4.2. Um problema não exclusivo da Embrapa

Tais problemas na afiliação dos autores não são exclusivos da Embrapa. Eles afetam todas as instituições com as quais a Embrapa produziu documentos científicos. Veja, no Quadro 3, uma estatística da diversidade de grafias referentes aos nomes dos cinco principais parceiros da Embrapa na elaboração dos seus artigos. Para efeito didático, de demonstrar a extensão e o alcance desse problema de padronização, acrescentamos cada uma dessas listas no Anexo de 165 páginas

disponibilizado em link da Internet da Embrapa. Veja o link para o arquivo no final do trabalho.

Quadro 3 – Estatísticas de artigos e grafias dos cinco principais parceiros da Embrapa em artigos na base *Web of Science*.

INSTITUIÇÃO	TOTAL ARTIGOS	NÚMERO DE AFILIAÇÕES	PORCENTAGEM
UFV	1.125	656	58
UNESP	1.008	1.223*	121
UnB-BR	916	680	74
USP-ESALQ	870	896*	103
UFLA	678	473	69

Fonte: Dados da pesquisa.

*Nota: Os escores de afiliações da UNESP e USP-ESALQ superiores ao próprio número de artigos podem ser explicados pela existência de dois ou mais autores que, num mesmo artigo, citaram sua afiliação de maneira diferente.

Quando examinados no nível micro, isto é, do centro de pesquisa, esses erros de padronização têm o poder de se multiplicar de forma exponencial. Por exemplo, a Embrapa Recursos Genéticos e Biotecnologia possui 1.545 artigos científicos publicados no período na WoS, nos quais foram encontradas 1.188 diferentes afiliações.

Para efeito de contagem da produção científica na base WoS, cada uma das diferentes grafias (1.188) corresponde a uma instituição. Essa é a principal razão porque, apesar de produzir artigos suficientes para figurar sozinha entre as 50 primeiras instituições do país, a Embrapa Recursos Genéticos e Biotecnologia aparece, na base não tratada, na sua primeira menção, “Embrapa Recursos Genet & Biotecnol, BR-70770900 Brasilia, DF, Brazil”, com 174 artigos. Esse escore corresponderia aproximadamente ao 700º lugar das instituições brasileiras. Essas 1.188 diferentes entradas da Embrapa Recursos Genéticos e Biotecnologia também estão acessíveis no Anexo deste artigo.

O problema se reproduz, por exemplo, numa universidade, quando se desce ao nível do Departamento e vai além, nos dois casos, da Embrapa e de Instituições de Ensino Superior (IES), quando se contempla mais um nível, ou seja, o de Laboratório ou Grupo de Pesquisa. Nesse nível ocorrem os piores erros, que, muitas vezes, impedem inclusive a atribuição do artigo a qualquer instituição.

É comum a inversão da afiliação, citando primeiro o laboratório, em seguida o departamento e, por último, a instituição. O bom senso indica a ordem inversa, sempre; instituição, departamento, laboratório.

A inversão de afiliação é um erro dos autores e carrega consigo a pior falta de exatidão, que é a elisão da instituição. Nesse caso aparecem entradas como “Bioagri Labs, BR-13412000 Piracicaba, SP, Brazil”, “Ctr Terapia Celular, Ribeirao Preto, SP, Brazil”, “Dept Biol Celular, Sao Paulo, Brazil”, “Dept Bioquim, Rio De Janeiro, Brazil”, “Dept Endodont, Rio De Janeiro, Brazil”, “Dept Biol Celular, Brasilia, DF, Brazil”, “Depto Desenvolvimento Ensino, Sao Paulo, Brazil”, “Lab Anim Physiol & Behav, Sao Paulo, Brazil”, “Lab Apoio Anim, Pernambuco, Brazil”, “Lab Biol Celular, Rio De Janeiro, Brazil”, “Lab Citogent, Sao Paulo, Brazil”, “Lab Mass Spectrometry, Brasilia, DF, Brazil”, “Mol Biol Lab, Brasilia, DF, Brazil”. Existem cidades que são sedes de mais de uma universidade. Nesses casos, como identificar a instituição?

Esse erro dos autores é comum também em instituições estrangeiras: “Dept Ecol & Biol, Tucson, AZ 85721 USA”, “DEPT AGR, DIV ENTOMOL, BANGKOK, THAILAND”, “Lab Cytogenet & Gebine Res, B-3000 Louvain, Belgium”, “Lab Invest Aplicada, Cordoba 14080, Spain”, “Opt Labs, Islamabad, Pakistan”, “QTL & Modifier Loci Grp, MRC Mammalian Genet Unit, Harwell, Oxon, England”, “Res Inst, Trichy 9, Tamil Nadu, India”, “Vet Informat & Epidemiol Res Grp, Glasgow G1 1XH, Lanark, Scotland”. Assim, torna-se difícil, quase impossível, descobrir a instituição de origem do(s) autor(es) do artigo e, por consequência, contar corretamente a produção científica da organização.

A menção do CEP ajuda, pois em alguns casos, para classificar a produção científica da Embrapa por centro de pesquisa esse foi o único identificador possível para as cidades onde a Empresa tem mais de um centro de pesquisa. No entanto, via de regra, quando os autores fazem a inversão de afiliação, há uma grande probabilidade de que a instituição responsável perderá esse artigo na contagem de sua produção científica, sobretudo quando se trata de *rankings* nacionais ou internacionais.

Tais efeitos são claramente sentidos no âmbito da Embrapa. A recuperação de 18.188 artigos da Empresa entre 1973 e 2015 a credencia como uma das dez primeiras instituições produtoras de artigos científicos indexados na WoS. No entanto, a maioria dos *rankings* de instituições brasileiras realizados a partir dessa base consegue enxergar melhor as universidades, ao passo que a Embrapa é vista de forma parcial (BRASIL, 2008; GOIS, 2008; GREGOLIN et al., 2005; LETA; CRUZ, 2003). Trata-se de uma perda considerável de visibilidade pública na base WoS, tanto da Empresa como de seus centros de pesquisa, parcialmente corrigida internamente pelo trabalho de acompanhamento da produção científica da Empresa na WoS realizado pela Secretaria de Gestão e Desenvolvimento Institucional (SGI), vinculada à própria Embrapa.

No caso geral das instituições brasileiras esses erros repetidos e constantes acarretam a perda de lugares preciosos nos *rankings* de IES, que estão se tornando cada vez mais populares em todos os continentes.

5 Conclusões e considerações finais

A partir da constatação dos problemas relatados os resultados preliminares deste estudo foram apresentados internamente a um grupo de trabalho coordenado pela Embrapa Informação Tecnológica, que propôs à presidência da Empresa a regulamentação da afiliação institucional dos seus empregados em publicações nacionais e internacionais. Essa proposta se materializou com a edição de uma resolução normativa interna em março de 2016 (EMBRAPA, 2016) com as seguintes determinações:

- a) em todas as publicações nacionais e internacionais, a afiliação institucional do autor deve ser indicada pela assinatura-síntese da unidade. Exemplos: Embrapa Gado de Corte; Embrapa, Departamento de Pesquisa e Desenvolvimento – DPD;
- b) os nomes das unidades centrais e descentralizadas devem ser mantidos sem tradução, em todas as publicações nacionais e internacionais;

- c) o endereço institucional não deve ser traduzido, devendo permanecer em português;
- d) no endereço institucional, apenas o nome da unidade central ou descentralizada deve ser utilizado, sem a indicação de qualquer laboratório ou área, mesmo em artigos em coautoria com universidades;
- e) a Embrapa Informação Tecnológica irá inserir essas orientações no Manual de Editoração da Embrapa (2016) e orientará as unidades no cumprimento dessa determinação;
- f) os Comitês Locais de Publicação (CLPs) devem garantir o cumprimento dessa norma.

Este trabalho e seu Anexo visam chamar a atenção de todos os responsáveis e dirigentes de instituições brasileiras para a extensão do problema.

Agradecimento

Os autores agradecem o apoio recebido de Rosângela Galon Arruda e Alessandra Rodrigues da Silva, analistas da Embrapa Informação Tecnológica, na busca desses resultados.

Referências

BOURNE, Charles Percy. Frequency and impact of spelling errors in bibliographic databases. **Information Processing & Management**, Elmsford, NY, v. 13, n. 1, p. 1-12, 1977.

BRASIL. Ministério da Ciência e Tecnologia. **Indicadores Nacionais de Ciência, Tecnologia e Inovação**. Disponível em: <<http://www.mcti.gov.br/indicadores>>. Acesso em: 18 mai. 2016.

EMBRAPA. Resolução Normativa nº4. **Boletim de Comunicações Administrativas**. Brasília: Embrapa, 2016.

EMBRAPA. **Manual de editoração**. Disponível em: <<http://manual.sct.embrapa.br/editorial/default.jsp>>. Acesso em: 18 mai. 2016.

EMBRAPA. **Política de Comunicação**. Brasília: Embrapa Informação Tecnológica, 2002. Disponível em: <<http://www.youblisher.com/p/123990-Politica-de-Comunicacao/>>. Acesso em: 18 mai. 2016.

GOIS, Antônio. ITA lidera em produtividade científica. **Folha de S. Paulo**, São Paulo, 14 jan. 2008. Caderno Ciência, A10.

GREGOLIN, José Ângelo Rodrigues et al. Análise da produção científica a partir de indicadores bibliográficos. In: LANDI, Francisco Romeu; GUSMÃO, Regina (Coord.). **Indicadores de ciência, tecnologia e inovação em São Paulo 2004**. São Paulo: FAPESP, 2005. 2 v. Disponível em: <<http://www.fapesp.br/2060>>. Acesso em: 18 mai.2016.

HOOD, William; WILSON, Concepción. Informetric studies using databases: opportunities and challenges. **Scientometrics**, Dordrecht, v. 58, n. 3, p. 587-608, 2003.

LETA, Jacqueline; CRUZ, Carlos Henrique de Brito. A produção científica brasileira. In: VIOTTI, Eduardo. B.; MACEDO, Mariano de M. (Orgs.). **Indicadores de ciência, tecnologia e inovação no Brasil**. Campinas: Editora da Unicamp, 2003. p. 123-168.

PENTEADO FILHO, Roberto de Camargo. **Création de systèmes d'intelligence dans une organisation de recherche et développement avec la scientométrie et la médiamétrie**. 2006. 328 p. Tese (Doutorado) - Université du Sud, Toulon Var, Toulon, 2006.

PENTEADO FILHO, Roberto de Camargo; FONSECA JÚNIOR, Wilson Corrêa da; AVILA, Antonio Flavio Dias. **Perfil da produção científica da Embrapa entre 2007 e 2015: oportunidades e desafios**. Documentos (Embrapa SGI) (1679-4680), v. 17, 2016. No prelo.

TAŞKIN, Zehra; AL, Umut. Standardization problem of author affiliations in citation indexes. **Scientometrics**, Dordrecht, v. 98, n. 1, p. 347-368, jan. 2014.

Anexo

Disponível em:

<<http://www22.sede.embrapa.br/web/sge01/estatisticaagricola/anexopadronizacaoEmQuestao.pdf>>.

The problem of standardization of authors' affiliations in the Web of Science database: the Embrapa case and its solution

Abstract: Scientific and technological production is essential in the innovation process of a country or scientific institution. However, bibliometric evaluation of this performance has been facing for many years a major challenge: the lack of accuracy of the information recorded in scientific documents and databases. This article has as main objective to verify to what extent this problem still persists in Brazilian scientific production, through literature and bibliometric investigations. In this second case, it was done, for example, a basic study on the visibility of the Brazilian Agricultural Research Corporation in the database Web of Science. The results showed a high incidence of standardization issues, reflected in the percentage of 11.93% of variations related to the original name and acronym Embrapa. Given this result, the institution issued an internal normative resolution regulating the membership of its employees in national and international publications.

Keywords: Affiliation. Standardization. Embrapa. Web of Science. Communication policy.

Recebido em: 02/09/2016

Aceito em: 08/11/2016