

CONSTRUÇÃO COLABORATIVA DE REPRESENTAÇÕES PARA A DISSEMINAÇÃO DE DADOS AGRÍCOLAS: Um estudo no Portal CoDAF

Collaborative construction of representations for agricultural data dissemination: a study in the CoDAF Portal.

Fábio Mosso Moreira¹, Diana Vilas Boas Souto Aleixo², Pedro Henrique Santos Bisi³, Leonardo F. Franchi⁴, Ricardo César Gonçalves Sant'Ana⁵

(1) UNESP, Marília, fabiomoreira@tupa.unesp.br

(2) UNESP, Marília, dianavbsouto@gmail.com

(3) UNESP, Marília, phbisi@gmail.com

(4) UNESP, Tupã, franchileonardo@gmail.com

(5) UNESP, Marília, ricardosantana@marilia.unesp.br

Resumo:

O ingresso das novas tecnologias informacionais na sociedade viabilizou o compartilhamento de dados e informações de forma dialógica, podendo ser considerado uma alternativa para aproximar os produtores às informações que os auxilia no desenvolvimento de suas atividades. Neste sentido, o objetivo deste trabalho é descrever o processo de construção colaborativa de representações de dados agrícolas, por meio de um estudo realizado no Portal Competências Digitais para Agricultura Familiar (Portal CoDAF), de modo a identificar os atributos que compõem a estrutura das representações e apontar os principais agentes e atividades envolvidos no processo. A metodologia baseou-se no Ciclo de Vida dos Dados como modelo para estruturação do fluxo informacional estudado. Como resultado parcial, pôde-se identificar um conjunto de atributos para compor a estrutura de disponibilização dos dados. Conclui-se que a participação dos agentes internos e externos de forma colaborativa é essencial para possibilitar a recuperação dos dados, assim como a definição de uma estrutura padrão para representação das fontes.

Palavras-chave: Fonte de dados. Acesso a dados. Ciclo de Vida dos Dados. Portal CoDAF. Dados agrícolas.

Abstract:

The admission new informational technologies in society, made possible data and information dialogical sharing, and may be considered an alternative for helping to bring producers to informations that helps you development of their activities. In this sense, the objective of the present work is describe the collaborative construction process of agricultural data representations, by study realized in Digital Skills for Family Farming Portal (CoDAF Portal) to identify the attributes which compose the representations structure and to indicate the key actors and activities involved in process. The methodology was based on the Data Life Cycle as a model for structuring of informational flow. As a partial result, it could be identified an set of attributes to provide data availability structure. The conclusion is that the participation of internal and external agents in a collaborative way is essential to allow for retrieving and recovery of the data, as well the definition a standard framework for representation of sources.

Keywords: Data sources. Access to data. Data Life Cycle. CoDAF Portal. Agricultural Data.

1 Introdução

As novas Tecnologias da Informação e Comunicação (TIC), em especial a Internet e a Web, podem ser consideradas importantes alternativas para viabilizar o compartilhamento de dados e informações dialogicamente, reduzindo a unidirecionalidade emissor-receptor, como, por exemplo, quando se observa na comunicação por meio de mídias tradicionais (ex: TV, rádio e jornais).

Um dos setores que pode se beneficiar com o uso destas tecnologias é a Agricultura.

O seguimento é composto por pequenos, médios e grandes produtores rurais, que juntos tem importante participação na geração de capital para o país, representando 23,5% do PIB Nacional (CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA, 2016).

Vieiro e Silveira (2011) ressaltam que as possibilidades proporcionadas pelas TIC para o meio rural são diversas, com destaque ao seu papel na ampliação de horizontes e incorporação de novas expectativas; constituição de grupos de comercialização;

novas políticas públicas; estimativas de safras e desempenhos nas bolsas de valores e *commodities*; serviços bancários; cooperativas de crédito e de produção; educação à distância e assistência técnica. Neste sentido, o meio rural não pode mais ser visto como um local distante e atrasado pelos que vivem no meio urbano e industrial, mas sim como um ícone de diversidade que está em constante desenvolvimento, demandando cada vez mais informações atualizadas e constantes.

Segundo Moreira et al. (2015), os dados e informações provenientes da agricultura estão distribuídos entre diversos agentes e instituições, como exemplo, produtores, universidades, institutos de pesquisa, serviços de extensão rural, iniciativa pública e privada, e organizações não governamentais (ONGs). Muitos destes conteúdos são disponibilizados em formato digital e podem ser acessadas via Internet e da Web, contudo, podem existir problemas na recuperação, tais como a heterogeneidade na definição dos domínios e a ambiguidade léxica.

Além dos problemas intrínsecos ao contexto da agricultura, Akerlof (1970) ressalta que a obtenção de conteúdos informacionais pode ser prejudicada quando a disponibilização das bases de dados não ocorre por meio de uma linguagem de fácil compreensão pelo usuário. Este fato se agrava quando a disponibilização é de responsabilidade de instituições públicas, as quais podem vir a fornecer serviços com acesso parcial, superficial e de difícil compreensão dos dados (LOPES; SANT'ANA, 2013), além de recursos com características heterogêneas e conjuntos de dados pulverizados em diferentes ambientes informacionais (SANT'ANA; RODRIGUES, 2013).

Quando se trata de um processo de compartilhamento de dados, deve-se considerar que o objeto em questão possui algumas particularidades, como o fato de um dado, por si só, não transmitir uma mensagem ou representar algum conhecimento (MOREIRA; SANT'ANA, JORENTE, 2016). O dado, como elemento básico, é formado por signos e não contém intrinsecamente, um componente semântico,

mas somente elementos sintáticos (SANTOS; SANT'ANA, 2002).

As características relacionadas à baixa carga semântica e à alta estruturação inerentes aos dados tornam o processo de recuperação diferente dos processos de recuperação da informação via mecanismos de buscas (JANOWICZ et al., 2012). Segundo Van Rijsbergen (1979), no processo de obtenção de dados busca-se por uma correspondência exata à necessidade, enquanto na recuperação da informação espera-se uma correspondência aproximada (a melhor correspondência possível); na recuperação de dados a inferência utilizada é a dedução lógica, ao passo que na recuperação da informação é possível realizar uma inferência indutiva (possui grau de certeza ou incerteza); na recuperação de dados a linguagem utilizada é uma linguagem artificial restrita por sintaxes específicas, enquanto na recuperação da informação utiliza-se de uma linguagem natural que pode ser expressa apenas com as especificações necessárias; ainda sobre a linguagem de consulta utilizada, na recuperação de dados a linguagem, por ser amarrada a uma sintaxe, possui maior sensibilidade ao erro, enquanto na recuperação da informação, pequenos erros podem ser contornados.

Uma possível solução ao problema da recuperação de dados agrícolas seria a construção de representações, a fim de facilitar a localização das bases e uso dos dados. Entretanto, devido à sua baixa carga semântica (SANTOS; SANT'ANA, 2013), os dados requerem esforços adicionais ainda maiores para sua representação, já que não bastam elementos que os descrevam como um todo e que propiciem sua recuperação.

Para Sant'Ana (2017), são necessários, ainda, elementos que permitam a sua interpretação por quem os acesse, com informações que detalhem sua estrutura e possibilitem a interpretação de cada atributo que os compõe. É importante considerar neste processo, aspectos como o formato das representações, o momento em que os dados devem receber tratamento e o responsável pela representação.

Estas questões indicam a necessidade da elaboração de estudos orientados a

buscar alternativas que venham a amenizar os problemas encontrados na recuperação de dados agrícolas, tais como a criação de camadas de representação de forma colaborativa. Ressalta-se que estas iniciativas podem ser realizadas tendo como base teorias e princípios da Ciência da Informação, que por sua vez deve ultrapassar a fronteira da informação interpretada para atingir o nível rígido dos dados (MOREIRA; VALENTIM; SANT'ANA, 2016).

A construção de representações para dados agrícolas pode ser realizada sob a responsabilidade do detentor, ou mesmo com a participação de outrem que venha a realizá-la de forma colaborativa. Para tanto, adota-se como objeto desta pesquisa um ambiente informacional digital especializado na disponibilização de conteúdos agrícolas e que disponibiliza representações para viabilizar a disseminação dos dados.

O Portal CoDAF¹ (website desenvolvido pelo Projeto Competências Digitais para Agricultura Familiar, da Faculdade de Ciências e Engenharia UNESP/Tupã) é um ambiente que proporciona áreas para que pequenos produtores divulguem conteúdos sobre sua produção de maneira on-line, estes podem qualificar seus produtos, fornecer características sobre seu empreendimento rural e indicar locais onde ocorre a comercialização de seus produtos. No Portal também são disponibilizados conteúdos como notícias sobre agricultura e tecnologia, e obtidos conjunto de dados agrícolas, sendo o último foco desta pesquisa.

2 Objetivos

Esta pesquisa tem como objetivo descrever o processo de construção colaborativa de representações de dados agrícolas por meio de um estudo realizado no Portal CoDAF, identificando a estrutura das representações e apontando os principais agentes e atividades envolvidas neste processo.

3 Procedimentos Metodológicos

Para atender ao objetivo proposto na pesquisa, o procedimento metodológico utilizado baseou-se em um modelo para estruturar fluxos informacionais que envolvem processos de compartilhamento de dados. Segundo Sant'Ana (2016), o Ciclo de Vida dos Dados (Figura 1) pode ser utilizado para estudar fatores e características que propiciem ampliação do equilíbrio entre os atores envolvidos no processo e a máxima otimização do uso dos dados. O modelo se baseia em uma estrutura básica para contextualizar momentos, características e requisitos em um aspecto cíclico de fluxo de dados.

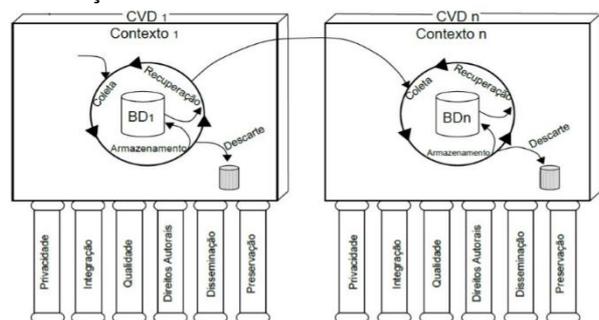
A estrutura do Ciclo de Vida dos Dados é composta pelas seguintes fases (SANT'ANA, 2016):

- **COLETA:** identificação das necessidades informacionais que irão nortear as escolhas dos dados necessários e a articulação de estratégias para localizar e avaliar estes dados, definindo as ferramentas necessárias para coleta.
- **ARMAZENAMENTO:** definição de quais dados serão disponibilizados e quais serão apenas armazenados, e qual estrutura física e lógica será utilizada para o armazenamento.
- **RECUPERAÇÃO:** formulação de estratégias para viabilizar que estes dados sejam encontrados, acessados e passíveis de interpretação (preferencialmente, e em muitos casos obrigatoriamente, por máquinas).
- **DESCARTE:** consiste na limpeza ou desativação da base, ou apenas a atividade de apagar o registro inteiro ou algum atributo específico.

Para Sant'Ana (2016) cada uma das fases são permeadas por seis objetivos específicos envolvidos no fluxo informacional: Privacidade, Integração, Qualidade, Direitos Autorais, Disseminação e Preservação. Esta pesquisa se insere principalmente no objetivo da Disseminação, uma vez que a construção de representações de dados pode aperfeiçoar seu processo de recuperação.

¹ Disponível em: <codaf.tupa.unesp.br>. Acesso em: 11/07/2017.

Figura 1: Ciclo de Vida dos Dados para Ciência da Informação.



Fonte: Sant'Ana (2016).

Após delimitado o modelo no qual a descrição do fluxo informacional será embasada, foram coletadas informações no Portal CoDAF por meio da observação da área contendo as representações de dados divulgados pelo portal. Ao total foram analisadas oito fontes² de dados agrícolas que estavam acessíveis até a data em que ocorreram as coletas (Julho/2017): “Dados sobre produção, importação e exportação de produtos agrícolas”; “Centro produtores de leite no estado de São Paulo”; “IBGE como fonte de Dados para a Agricultura”; “Dados sobre o repasse de recursos financeiros do Programa Nacional de Fortalecimento da Agricultura Familiar (PRONAF)”; “Dados sobre preço recebido pelos agricultores (Laranja)”; “Base Cartográfica Contínua do Brasil - Escala 1:1.000.000”; “Fonte de Dados sobre preço médio de Hortifrutícolas”; e “Dados sobre recursos do MDS/MDA aplicados no Programa de Aquisição de Alimentos”.

A partir do conteúdo observado, foram definidos os principais atributos que compõem a estrutura das representações. Para apresentar as informações analisadas, elaborou-se um fluxograma que ilustra o processo de construção colaborativa de representações de dados agrícolas no Portal CoDAF.

4 Resultados

O conteúdo disponibilizado no Portal CoDAF é gerado tanto pelos integrantes do projeto quanto pela participação de agentes externos, como no cadastro de propriedades

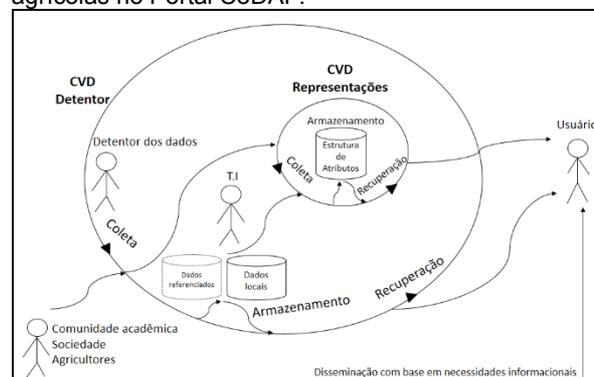
² Disponível em: <<http://codaf.tupa.unesp.br/agricultura-familiar/fontes-de-dados>>. Acesso em: 11/07/2017.

rurais, no envio das notícias para publicação, e na construção das representações de dados agrícolas, que buscam viabilizar a disseminação de fontes e utilização dos dados. As atividades envolvidas na fase da coleta deste fluxo informacional baseiam-se na premissa da colaboração, e são realizadas por meio de atividades junto à comunidade acadêmica, à sociedade e aos produtores, por meio das funcionalidades do Portal ou através da realização de cursos e oficinas.

O conteúdo gerado na coleta é armazenado localmente por um banco de dados mantido pelo CoDAF. Também existem referências a conteúdos de fontes externas por meio de *hiperlinks*, apontados nas notícias publicadas e nas representações de dados disponíveis, configurando uma base referenciada. Na fase de armazenamento é importante destacar o papel do suporte tecnológico utilizado, garantido pelos recursos informáticos e pela mediação dos integrantes responsáveis pela gestão das ferramentas tecnológicas (gerenciador de conteúdo, manutenção do servidor).

Após a coleta e armazenamento, o conteúdo fica disponível no Portal CoDAF para ser recuperado pelo usuário. Como citado anteriormente, encontram-se problemas no processo de recuperação de dados que podem ser amenizados com a construção de representações. Este processo pode ser considerado um ciclo de dados interno ao ciclo de dados principal do detentor (Figura 2).

Figura 2: Descrição do processo de disponibilização de conteúdo e construção das representações de dados agrícolas no Portal CoDAF.



Fonte: Autores.

Após analisar o conteúdo das oito representações de dados disponíveis no Portal CoDAF, identificaram-se os seguintes atributos que podem definir sua estrutura:

1. Título: denomina o conteúdo dos dados presentes na representação.
2. Introdução: resumo que sintetiza o contexto do conteúdo dos dados representados (ex: o que, como, onde, por quê).
3. Informações sobre a origem dos dados:
 - a. Nome e descrição da instituição mantenedora: instituição detentora dos dados presentes na representação, tais como: área de atuação, empresa privada ou pública, quais seus objetivos, que informações estão disponibilizadas em seu portal;
 - b. Título da fonte originária: Rótulo de cabeçalho da fonte originária;
 - c. Endereço eletrônico: *hiperlink* de acesso para a fonte originária;
 - d. Descrição do processo de navegação: trilha hierárquica para chegar até a fonte originária a partir da página inicial da instituição detentora dos dados.
4. Informação sobre os dados presentes no(s) recurso(s) informacional(is) disponíveis na fonte originária:
 - a. Formato(s) do(s) arquivo(s) acessíveis na fonte originária: (ex: “.xls”, “.odt”, “.html”, “.csv”);
 - b. Atributos dos dados: descrição dos atributos (colunas), o que representam (registros), qual o tipo de variável que recebe (ex: data, texto, número, valor).
5. Demonstração de uma possível visualização para os dados presentes nos recursos utilizados.
 - a. Exibição de gráficos ou elemento de visualização

construídos a partir do uso dos dados acessíveis no recurso;

b. Descrição das informações que podem ser obtidas através da interpretação das visualizações geradas e uso dos dados presentes em algum recurso.

5 Considerações parciais

A pesquisa baseou-se no Ciclo de Vida dos Dados para analisar a construção colaborativa de representações de dados agrícolas por meio de um estudo realizado no Portal CoDAF. Com a descrição do processo foi possível observar os principais agentes envolvidos neste ciclo, em quais fases atuam e as atividades relacionadas a cada um.

Com a definição de uma estrutura padrão, obtida a partir da análise das representações disponíveis no Portal CoDAF, as atividades de construção colaborativa podem ser ampliadas por meio do fornecimento de uma aplicação para coleta online, a fim de aumentar a participação de agentes externos e consequentemente viabilizar a disseminação do conteúdo.

A participação de agentes externos é fundamental uma vez que este público pode conhecer melhor as necessidades informacionais e competências de quem busca os dados, como ilustrado na Figura 2.

Considera-se que as ações descritas e a estrutura definida para construção das representações podem ser implementadas por outras iniciativas que busquem amenizar os problemas relacionados com a recuperação e disseminação de dados no contexto da agricultura.

Referências

- AKERLOF, G A. The Market for “Lemons”: qualitative uncertainty and the market mechanism. **The quarterly journals of economics**, v.84, n.3, p. 488-500, 1970. Disponível em:<<https://www.iei.liu.se/nek/730q83/artiklar/1.328833/AkerlofMarketforLemons.pdf>>. Acesso em: 29/08/2017.
- CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA [CEPEA-USP]. **PIB do Agronegócio BRASIL**. ESALQ/USP, 2016. Disponível em:<<http://www.cepea.esalq.usp.br/upload/kc>>.

[editor/files/Relatorio PIBAGRO](#)

[Brasil_DEZEMBRO.pdf](#)>. Acesso em: 14/04/2017.

JANOWICZ, K.; SCHEIDER, S.; PEHLE, T.; HART, G. Geospatial semantics and linked spatiotemporal data: past, present, and future. **Semantic Web**, v. 3, n. 4, p. 321-332, 2012. Disponível

em:<<http://content.iospress.com/download/semantic-web/sw077?id=semantic-web%2Fsw077>>. Acesso em: 28/08/2017.

LOPES, R. de C. C.; SANT'ANA, R. C. G. Percepção dos usuários sobre o processo de acesso a dados sobre saúde em sítios do governo federal. In: XIV - ENANCIB, 2013, Florianópolis. **Anais...**, 2013. Disponível em:<<http://enancib.ibict.br/index.php/enancib/xivenancib/paper/view/4370/3493>>. Acesso em: 29/08/2017.

MOREIRA, F. M.; SANT'ANA, R. C. G.; SANTAREM SEGUNDO, J. E.; VIDOTTI, S. A. B. G. Tecnologias da Web Semântica para a recuperação de dados agrícolas: um estudo sobre o International Information System of the Agricultural Science and Technology (AGRIS). **Em Questão**, v. 21, p. 173-192, 2015. Disponível

em:<<http://seer.ufrgs.br/index.php/EmQuestao/article/view/50317>>. Acesso em: 29/08/2017.

MOREIRA, F. M.; SANTANA, R. C. G.; JORENTE, M. J. V. A Complexidade na disponibilização e acesso a dados governamentais na Web. **Perspectivas em Ciência da Informação**, v. 21, p. 70-88, 2016. Disponível

em:<<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2540>>. Acesso em: 29/08/2017.

MOREIRA, F. M.; VALENTIM, M. L. P.; SANT'ANA, R. C. G. Interdisciplinaridades em Ciência da Informação: um estudo do compartilhamento de dados governamentais na web. In: III Encontro Internacional Dados, Informação e Tecnologia, 3, Marília, 2016. **Anais...** do III Encontro Internacional Dados, Informação e Tecnologia, 2016. Disponível

em:<<http://gpnti.marilia.unesp.br:8085/index.php/3DTI/3dti/paper/view/340/150>>. Acesso em: 29/08/2017.

SANT'ANA, R. C. G. Ciclo de Vida dos Dados: Uma perspectiva a partir da Ciência da Informação. **Informação e Informação**,

Londrina, v.21, n.2, p.116-142, 2016.

Disponível

em:<<http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940>>. Acesso em: 29/08/2017.

SANT'ANA, R. C. G. Reflexões sobre a Representação no Ciclo de Vida dos Dados. In: I Encontro de Representação Documental (EnReDo), 1, São Carlos, 2017. **Anais...**, 2017. Disponível

em:<<http://www.telescopium.ufscar.br/index.php/enredo/enredo/paper/viewFile/109/96>>. Acesso em: 29/08/2017.

SANTANA, R. C. G.; RODRIGUES, F. A. Visualização de afinidades entre parlamentares mediante dados de votações no Senado Brasileiro. **Informação & Sociedade: Estudos**, João Pessoa, v.23, n.1, p.49-59, 2013. Disponível

em:<<https://repositorio.unesp.br/handle/11449/75885>>. Acesso em: 29/08/2017.

SANTOS, P. L. A. C.; SANT'ANA, R. C. G. Transferência da informação: análise para valoração de unidades de conhecimento. **DataGramaZero**, Rio de Janeiro, v.3, n.2, 2002. Disponível

em:<<http://basessibi.c3sl.ufpr.br/brapci/index.php/article/view/0000001259/f8d0251a874410d87c7a0bcb589fc725>>. Acesso em: 29/08/2017.

SANTOS, P. L. V. A. C.; SANT'ANA, R. C. G. Dado e Granularidade na perspectiva da Informação e Tecnologia: uma interpretação pela Ciência da Informação. **Ciência da Informação**, Brasília, v. 42, p. 199-209, 2013. Disponível

em:<<http://revista.ibict.br/ciinf/article/view/1382>>. Acesso em: 29/08/2017.

VAN RIJSBERGEN, C. J. **Information Retrieval**. Ed. 2. Butterworth-Heinemann Newton, MA, USA, 1979. Disponível

em:<http://openlib.org/home/krichel/courses/lis618/readings/rijsbergen79_infor_retriev.pdf>. Acesso em: 29/08/2017.

VIERO, V. SILVEIRA, A. Apropriação de Tecnologias de Informação e Comunicação no meio rural brasileiro. **Cadernos de Ciência e Tecnologia – Embrapa**, Brasília, v. 28, n. 1, p. 257-277, 2011. Disponível

em:<<https://seer.sct.embrapa.br/index.php/cc/article/view/12042>>. Acesso em: 29/08/2017.

