

capítulo

3

Redes e parcerias para abertura de dados agrícolas

Fernando Elias Correa
Debora Pignatari Drucker
Filipi Miranda Soares
Kelly Rosa Braghetto
Alexandre Cláudio Botazzo Delbem
Fernando Santos Osório
Antônio Mauro Saraiva

Introdução

O agronegócio brasileiro e atividades correlatas geraram para o ano de 2023 a expectativa de crescimento do Produto Interno Bruto (PIB) do agronegócio na ordem de 13,2% (Ipea, 2020). O setor, mesmo diante da pandemia, continuou a expandir sua importância na economia brasileira.

Contudo, apesar dos números positivos, os desafios para a atividade agropecuária são muitos, seja pelas características de perecibilidade dos seus produtos, pelo perfil de grande parte de seus produtores, pelas formas de distribuição prevalentes, ou pela forma de consumo muitas vezes in natura. Assim, dado o caráter altamente distribuído desse mercado, ele se apresenta como um objeto de estudo e de implementação complexo do ponto de vista da organização dos dados ou, neste contexto, os agrodados.

Parte desse processo tem sido desempenhado pelo uso de tecnologias empregadas em diferentes setores agrícolas – desde a produção, comercialização e distribuição – e tem exigido avanços significativos de empresas e entidades especializadas. Como exemplo, podemos citar o uso de dispositivos e software para obtenção e análise massiva de dados, o que tem sido conhecido como Agricultura Digital. Com isso, investimentos e pesquisas têm sido aplicados por entidades, como mostram os trabalhos junto à Embrapa (Bolfe et al., 2020) e à Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) (Zaparolli, 2020). No entanto, na mesma proporção, crescem os desafios relativos à integração, geração, armazenamento, tratamento e disponibilização de agrodados.

Assim, torna-se necessário pesquisar, desenvolver e utilizar métodos que permitam a aplicação de modelos para gestão de dados com enfoque no domínio de aplicação do agronegócio (Corrêa, 2019). Em alinhamento à gestão de dados está a necessidade de facilitar o acesso a dados em plataformas, a definição clara de seus metadados, a geração de informações levando em conta a usabilidade e acurácia dos dados.

Neste contexto, no Centro para Inteligência Artificial (C4AI) da Universidade de São Paulo (USP)¹ tem-se trabalhado em procedimentos e modelagem dos agrodados, com definição de modelos de metadados e aplicação de princípios FAIR (*Findable, Accessible, Interoperable, Reusable*). Os trabalhos se desenvolvem em três linhas sinérgicas, com objetivo de melhorar e ampliar a acessibilidade, acurácia e qualidade

¹ Para mais informações, consulte: <https://c4ai.inova.usp.br/>.

dos agrodados, permitindo assim sua reprodutibilidade e seu reúso. Assim, o primeiro tema abordado neste capítulo será a publicação de um website e ferramentas de apoio, como ontologias e termos de metadados, a serem divulgados no domínio público como www.almescore.org – que se refere ao acrônimo para Agricultural and Livestock Metadata Element Set Core (C4AI-Agribio, 2023). Na sequência, será descrito o processo de trabalho no desenvolvimento do Compromisso 5 do 5º Plano de Ação Nacional Brasileiro para Governo Aberto, intitulado “Promover a abertura e integração de bases de dados das cadeias agropecuárias com aderência aos interesses da sociedade” (Brasil, 2021). O Centro coordenou o Marco 4, “Protocolo de metadados para integração e transparência de dados da cadeia agropecuária com base em padrões da Web e conceitos FAIR”. E por fim, a terceira linha de ação é a atuação para o desenvolvimento da Rede GO Fair Agro Brasil, a qual possui um grupo de trabalho denominado Dados, Metadados e Repositórios, que iniciou os trabalhos colaborativos no ano de 2022, e está em execução para agregar diferentes instituições públicas e privadas que lidam com agrodados.

Modelo de metadados AlmesCORE

O princípio trabalhado para a integração dos agrodados foi considerar a gestão dos dados analisando diferentes atores, como produtores, fornecedores e utilizadores de dados. Para tanto, o conhecimento sobre os dados, a forma como foi gerado, características como periodicidade, localização e frequência, dentre outras, devem estar claras e disponíveis a todos os atores envolvidos.

Assim, algumas boas práticas devem ser aplicadas pelos produtores dos agrodados. Um aspecto fundamental é gerar metadados, que são informações que detalham os dados e devem estar disponíveis junto aos conjuntos de dados descritos. Contudo, a definição dos termos para os metadados também deverá ser precedida de uma padronização ou integração entre os especialistas, instituições e fontes provedoras de dados, pois ações isoladas ou exclusivas de geração de metadados por fornecedores poderão gerar termos e conjuntos de metadados incompatíveis entre si. Assim, o AlmesCORE, um acrônimo para “*Agricultural and Livestock Metadata Element Set Core*”, foi criado com o objetivo de prover um modelo de metadados orientado ao domínio de agrodados, acompanhado de ferramentas, formatos e meios de visualização de modo a atender a diferentes produtores, fornecedores e utilizadores de dados.

Foram estabelecidos alguns passos para esse objetivo, iniciado pela avaliação de modelos disponíveis de metadados para agrodados, seguido pela avaliação dos

provedores de dados existentes para o subdomínio de dados de comercialização. Após isso, foi proposto o modelo de metadados, com sua publicação em formato de RDF (*Resource Description Framework*) e a criação de um site de divulgação do AlmesCORE.

Para a avaliação de modelos de metadados existentes não foram encontrados modelos formais seguindo as descrições dos termos, ontologias e formatos como RDF disponíveis para agrodados. Foram encontradas aplicações de metadados diretas em plataformas provedoras de dados como a do Instituto Brasileiro de Geografia e Estatística (IBGE) com os registros já preenchidos. Existe uma proposta de modelo feita pela Organização das Nações Unidas para a Alimentação e a Agricultura (FAO), e que não está mais sendo atualizado, chamado de AgMES, mas que trata de alguns descritores, com poucos termos de metadados e que não são aplicáveis para descrever boa parte das características de dados dos provedores de agrodados no Brasil.

Assim, foram feitas as avaliações de três principais provedores de agrodados de comercialização no Brasil, o Centro de Estudos Avançados em Economia Aplicada da Universidade de São Paulo (Cepea/USP), a Companhia Nacional de Abastecimento (Conab) e o Instituto de Pesquisa Econômica Aplicada (Ipea) para buscar um denominador comum de termos para descrição dos dados desses provedores.

Com a definição dos termos de metadados, foram serializados em RDF/XML e Shapes Constraint Language (SHACL) e implementados para disponibilização do modelo para a comunidade usar e implementar em seus provedores, gerando assim os registros de metadados compatíveis entre distintos provedores de agrodados.

De modo a ampliar a divulgação e usabilidade do modelo de metadado, foi criado o portal do AlmesCORE na internet (C4AI-Agribio, 2023), o qual registrou todo o histórico e atualizações do projeto, novas implementações, suporte e contato. O objetivo é tornar a principal forma de interação com a comunidade e utilizadores do AlmesCORE.

Protocolo de metadados para integração de agrodados

A metodologia aplicada no Marco 4 do Compromisso 5 do 5º Plano de Ação brasileiro, no âmbito da Parceria para Governo Aberto (Open Government Partnership (OGP)), desenvolveu um protocolo de metadados com ênfase em dados das cadeias agropecuárias relativas à comercialização (Brasil, 2021).

O processo para a criação do protocolo foi dividido em quatro etapas:

- Definição das bases de dados utilizadas para consulta e extração de metadados.

Neste processo foram selecionadas três fontes de dados, notoriamente reconhecidas por publicações de informações relativas a dados econômicos e agropecuários. As fontes selecionadas na pesquisa foram o Cepea/USP², a Conab³ e o Ipea⁴, todas instituições participantes do Compromisso 5. Foram selecionados os dados dessas fontes correspondentes a informações agropecuárias relativas a dados de comercialização.

- Definição e mapeamento de termos correlatos e comuns nas bases de dados previamente mapeadas.

Após a seleção das bases de dados, foram extraídos os dados relacionados à temática de comercialização dessas fontes. Para isso não foram considerados os períodos temporais, uma vez que o interesse foi mapear os termos (propriedades de metadados) comuns entre as bases de dados. Na etapa de resultados, são apresentados os recortes de consultas e extrações de dados feitas dessas fontes para mapeamento dos termos correlatos.

- Definição das propriedades, terminologia e termos do modelo de metadados.

Foram definidos os termos, as propriedades, as nomenclaturas e as relações necessárias para a proposta e criação dos termos de metadados para cadeias agropecuárias. Também foram definidos o nome e as ferramentas para a publicação do modelo. Os materiais produzidos são apresentados na etapa de resultados.

- Criação dos formulários de avaliação e explicação dos termos de metadados para os participantes do Marco 4 para validação.

Como forma de apresentar e validar o modelo de metadados proposto, foram criados e divulgados para os membros do Marco 4, antes de sua publicação, material para validação e simulação de preenchimento dos registros de metadados. Os documentos de validação e preenchimento são apresentados na etapa de resultados.

Para a etapa de seleção de fontes, conforme os métodos descritos em definição e mapeamento de termos correlatos e comuns nas bases de dados previamente mapeadas, foram empregados esforços para se definir um conjunto mínimo de termos de metadados para dados de comercialização. Um conjunto de metadados descritivos, ainda sem especificação formal, foi gerado e apresentado na Tabela 3.1.

²Para mais informações, consulte: <https://www.cepea.esalq.usp.br/br>.

³Para mais informações, consulte: <https://www.conab.gov.br/>.

⁴Para mais informações, consulte: <https://www.ipea.gov.br/>.

Esse exemplo foi extraído de dados do Cepea/USP, os quais foram agrupados para o modelo descritivo

Tabela 3.1. Exemplos de seleção de fonte com metadados descritivos.

Cepea - Exemplo			
Campo abreviado	Lim ⁽¹⁾	Ex1: Açúcar cristal	Ex2: Indicador açúcar santos
aggr_agri	1	Primário	primário
grup_prod	1	Sucroalcooleiro	Sucroalcooleiro
temas	1	Indicador	Indicador
produto	1	Açúcar	açúcar
nome_fantasia	1	Indicador de Açúcar Cristal - à vista	Indicador de Açúcar Cristal - Santos - à vista
tipo_dado	1	série temporal	série temporal
fonte	1	Cepea	Cepea
disseminacao	1	site, api, excel	site, api, excel
variaveis	0	data; valor (R\$); valor (US\$)	data; valor (R\$); valor (US\$)
freq	0	Diária	diária
unidade	0	saca 50 kg	saca 50 kg
localidade	0	São Paulo	Santos
periodicidade	0	20/05/2003 - contínua	23/01/2013 - contínua
metodologia	0	Escola Superior de Agricultura Luiz de Queiroz (2024a)	Escola Superior de Agricultura Luiz de Queiroz (2024b)

⁽¹⁾ Limitação (1 - Obrigatório / 0 - facultativo)

Fonte: Escola Superior de Agricultura Luiz de Queiroz (2024a, 2024b).

Os metadados administrativos ainda serão definidos conforme recomendações dos princípios FAIR e da norma ISO/IEC 11179. O processo inicial foi extrair os metadados em potencial das bases do Cepea/USP, Ipea e Conab e estabelecer metadados comuns entre as três bases, conforme apresentado na Tabela 3.2.

Tabela 3.2. Metadados descritivos da relação entre as três fontes.

Metadado	Cepea	Ipea	Conab
produtoNome	x	x	x
indicador	x	–	–
indice	–	x	–
preçosAgrícolas	–	–	x
regiãoReferência	x	–	x
data	x	x	x
periodicidade	x	x	x
unidadeMonetária	x	x	x
valorMonetário	x	x	x
medidaReferênciaValor	x	x	x
medidaReferênciaUnidade	x	x	x
metodologiaID	x	x	–

Para a definição formal dos termos dos metadados propostos, inicialmente foi apresentada a composição resumida dos elementos aplicados a um conjunto de metadados de dados de comercialização agrícola, mas espera-se que nos próximos anos outros domínios da agricultura também sejam cobertos pelo esquema de metadados (Soares et al., 2022). A descrição resumida dos elementos está disponível em C4AI-Agribio (2023).

A última etapa foi a criação dos formulários de avaliação e explicação dos termos de metadados para os participantes do Marco 4 para validação

Os termos de metadados foram submetidos à avaliação dos integrantes do Marco 4 do Compromisso 5. Este processo foi dividido em duas etapas. Primeiro, foi criado um formulário para preenchimento de modo instrutivo de um registro para o protocolo de metadados e disponibilizado online para preenchimento dos participantes.

Em sequência, foi criado um formulário para avaliação da usabilidade do modelo de metadados e disponibilizado para os participantes.

A publicação do esquema de metadados está em processo de finalização. Foi aplicada a metodologia de reuso de ontologias para se realizar a especificação formal do esquema de metadados em RDF/XML e SHACL. Essa metodologia prevê que ontologias ou outros vocabulários controlados existentes para um determinado domínio possam ser reutilizados na construção de novos vocabulários controlados, evitando-se assim o trabalho de criar termos que já existam.

Na evolução dos trabalhos, serão realizados testes de anotações de dados com as propriedades do AlmesCORE e criados na ferramenta *FAIR Data Points* (Santos et al., 2023) para verificar a aderência dos conjuntos de dados descritos com o conjunto de elementos AlmesCORE aos *FAIR Data Principles*. Serão aplicados em conjunto de dados reais de provedores de dados como o Cepea, Conab e Ipea.

GT de dados, metadados e repositórios – rede GO FAIR Agro Brasil

A rede GO FAIR é uma iniciativa internacional iniciada em 2018 na Europa e atualmente presente em diversos países. Ela é dirigida pelas partes interessadas e autogovernada e visa implementar os princípios FAIR (Wilkinson et al., 2016), tornando os dados localizáveis, acessíveis, interoperáveis e reutilizáveis. O Brasil foi um dos primeiros países a integrar a iniciativa GO FAIR (Sales et al., 2020) e conta com Redes de Implementação (RIs) temáticas, as quais contribuem para a disseminação e a adoção dos princípios FAIR no Brasil em diferentes domínios do conhecimento. As atividades de implementação da RI GO FAIR Agro Brasil foram iniciadas em 2021 no XIII Congresso Brasileiro de Agroinformática, promovido pela Associação Brasileira de Agroinformática (SBIAgro), com a elaboração do Manifesto⁵ da rede e adesão de dezenas de intuições e de centenas de pesquisadores de todo o Brasil (Drucker et al., 2021).

Ao longo de 2022, foram realizadas oficinas para a construção coletiva da RI, resultando na criação de três grupos de trabalhos: GT Dados, Metadados e Repositórios; GT Políticas e Planos de Gestão de Dados e GT Ontologias. O primeiro conta com a coordenação do grupo que lidera o AlmesCORE e representa a continuidade das ações anteriormente descritas para a criação e implementação de um modelo de metadados

⁵ Disponível em: <https://go-fair-agro.github.io/>.

para agrodados. O grupo de trabalho foi estabelecido inicialmente com as seguintes atividades:

- Mapeamento e análise de padrões de dados e metadados para domínios de conhecimento de dados agropecuários.
- Criação e definição de termos e modelos de metadados e extensão para modelos existentes, aplicados a especificidades dos dados agropecuários, com ampla participação da comunidade de prática.
- Desenvolvimento e aplicação de tipologia de dados, metadados e padrões para garantia de qualidade e (re)usabilidade dos dados na fase de inferência e de aplicação em Inteligência Artificial (IA), compatível com os dados de treinamento.

O GT Dados, Metadados e Repositórios está dando continuidade à implementação do AlmesCORE e promovendo discussões para alcançar os objetivos estabelecidos junto à Rede GO FAIR Agro Brasil. Com a validação do modelo para dados de comercialização, espera-se expandir a metodologia para outros subdomínios relevantes aos agrodados e promover a adoção dos padrões pela comunidade.

Conclusões

A crescente quantidade de agrodados gerada nos últimos anos e a tendência de aumento futuro demanda que produtores, fornecedores e usuários trabalhem de forma harmônica para que os maiores benefícios possíveis à sociedade sejam alcançados. Além dos aspectos de governança tratados em outros capítulos do presente livro, é necessário desenvolver e promover a adoção de padrões que viabilizem a implementação dos princípios FAIR. A articulação de redes e parcerias é fundamental para viabilizar que padrões de dados e metadados sejam representativos das práticas da comunidade, uma vez que iniciativas como a OGP e a Rede GO FAIR Agro reúnem pessoas e instituições que enfrentam desafios semelhantes e procuram soluções que podem ser compartilhadas.

A estratégia apresentada neste capítulo de: 1) propor o AlmesCORE no âmbito de um centro de inteligência artificial, buscando tecnologias atuais para implementá-lo; 2) articular sua validação no âmbito de uma iniciativa reconhecidamente bem-sucedida como o compromisso pelo Governo Aberto, a qual reuniu organizações com desafios comuns e permitiu a validação da proposta para padronização de dados de comercialização agrícola; e 3) ampliar o escopo disciplinar e agregar outros atores

relevantes no contexto da Rede GO FAIR Agro Brasil mostrou-se bem-sucedida e aponta para futuros resultados que serão valiosos para promover a adoção de padrões de dados na agricultura.

Referências

BOLFE, E. L.; BARBEDO, J. G. A.; MASSRUHÁ, S. M. F. S.; SOUZA, K. X. S. de; ASSAD, E. D. Desafios, tendências e oportunidades em Agricultura Digital no Brasil. In: MASSRUHÁ, S. M. F. S.; LEITE, M. A. de A.; OLIVEIRA, S. R. de M.; MEIRA, C. A. A.; LUCHIARI JUNIOR, A.; BOLFE, E. L. (ed.). **Agricultura Digital: pesquisa, desenvolvimento e inovação nas cadeias produtivas**. Brasília, DF: Embrapa, 2020. cap. 16, p. 380-406. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/handle/doc/1126283>. Acesso em: 22 set. 2023.

BRASIL. Controladoria-Geral da União. **5º Plano de ação nacional em governo aberto**. Brasília, DF, 2021. Disponível em: https://repositorio.cgu.gov.br/bitstream/1/67649/3/5_plano_acao_nacional_4_2_2022.pdf. Acesso em: 20 jan. 2022.

C4AI-AGRIBIO. **AlmesCore**: Agricultural and Livestock Metadata Elemento Set. 2023. Disponível em: <https://www.almescore.org/>. Acesso em: 29 jan. 2024.

CORRÊA, P. L. P. Modelo organizacional para gestão integrada de dados da biodiversidade brasileira. In: DIAS, G. A.; OLIVEIRA, B. M. J. F. de (org.). **Dados científicos: perspectivas e desafios**. João Pessoa: Editora UFPB, 2019. cap. 7, p. 153-175.

DRUCKER, D. P.; CRUZ, S. M. S. da; SARAIVA, A. M.; FORTALEZA, J. M.; BERTIN, P. R. B.; SIMAO, V. P. M.; TELLES, M. A.; SILVA, A. R. da; SANTOS, P. S. S.; MACARIO, C. G. do N. Implantação da Rede Temática GO-FAIR Agro Brasil: primeiros passos. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 13., 2021, Bagé. **Anais [...]**. Bagé: Unipampa, 2021. p. 164-171. Organizado por Ana Paula Lüdtke Ferreira. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/handle/doc/1136670>. Acesso em: 19 jun. 2023.

ESCOLA SUPERIOR DE AGRICULTURA LUIZ DE QUEIROZ. Centro de Estudos Avançados em Economia Agrícola. **Indicador açúcar cristal – Santos (FOB)**. Disponível em: <https://cepea.esalq.usp.br/br/metodologia/indicador-acucar-cristal-santos-fob.aspx>. Acesso em: 25 out. 2024a.

ESCOLA SUPERIOR DE AGRICULTURA LUIZ DE QUEIROZ. Centro de Estudos Avançados em Economia Agrícola. **Metodologia do açúcar cristal branco CEPEA/ESALQ – São Paulo**. Disponível em: <https://www.cepea.esalq.usp.br/br/metodologia/metodologia-do-acucar-cristal-branco-cepea-esalq-sao-paulo.aspx>. Acesso em: 25 out. 2024b.

IPEA. Conjuntura agrícola brasileira. In: _____. **Carta de Conjuntura**, n. 49, p. 1-4, 4 trim. 2020. Nota de conjuntura 22.

SALES, L.; HENNING, P.; VEIGA, V.; COSTA, M. M.; SAYÃO, L. F.; SANTOS, L. O. B. da S.; PIRES, L. F. GO FAIR Brazil: a challenge for Brazilian data science. **Data Intelligence**, v. 2, n. 1-2, p. 238–245, Nov. 2019. DOI: https://doi.org/10.1162/dint_a_00046.

SANTOS, L. O. B. da S.; BURGER, K.; KALIYAPERUMAL, R.; WILKINSON, M. D. FAIR data point: a FAIR-oriented approach for metadata publication. **Data Intelligence**, v. 5, n. 1, p. 163-183, Winter 2023.

DOI: https://doi.org/10.1162/dint_a_00160.

SOARES, F. M.; CORRÊA, F. E.; PIRES, L. F.; SANTOS, L. O. B. da S.; DRUCKER, D. P.; BRAGHETTO, K. L.; MOREIRA, D. de A.; DELBEM, A. C. B.; SILVA, R. F. da; LOPES, C. O. da S.; SARAIVA, A. M. Building a community-based FAIR metadata schema for Brazilian agriculture and livestock trading data. **CEUR Workshop Proceedings**, v. 3235, 2022. 6 p. Edition of the 18th International Conference on Semantic Systems, Vienna, Austria. SEMANTiCS 2022. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/handle/doc/1147744>. Acesso em: 29 jan. 2024.

WILKINSON, M. D.; DUMONTIER, M.; AALBERSBERG, I. J.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J. W.; SANTOS, L. B. da S.; BOURNE, P. E.; BOUWMAN, J.; BROOKES, A. J.; CLARK, T.; CROSAS, M.; DILLO, I.; DUMON, O.; EDMUNDS, S.; EVELO, C. T.; FINKERS, R.; GONZALEZ-BELTRAN, A.; GRAY, A. J. G.; GROTH, P.; GOBLE, C.; GRETHE, J. S.; HERINGA, J.; 't HOEN, P. A. C.; HOOFT, R.; KUHN, T.; KOK, R.; KOK, J.; LUSHER, S. J.; MARTONE, M. E.; MONS, A.; PACKER, A. L.; PERSSON, B.; ROCCA-SERRA, P.; ROOS, M.; SCHAIK, R. van; SANSONE, S. A.; SCHULTES, E.; SENGSTAG, T.; SLATER, T.; STRAWN, G.; SWERTZ, M. A.; THOMPSON, M.; VAN DER LEI, J.; VAN MULLIGEN, E.; VELTEROP, J.; WAAGMEESTER, A.; WITTENBURG, P.; WOLSTENCROFT, K.; ZHAO, J.; MONS, B. The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, v. 13, p. 1-8, Mar. 2016. DOI: <https://doi.org/10.1038/sdata.2016.18>.

ZAPAROLLI, D. Agricultura 4.0. **Pesquisa Fapesp**, v. 287, jan. 2020.