

5º PLANO DE AÇÃO NACIONAL EM GOVERNO ABERTO

COMPROMISSO 5: 'Promover a abertura e integração de bases de dados das cadeias agropecuárias com aderência aos interesses da sociedade'

Relatório final do MARCO 4 (Protocolo de metadados para integração e transparência de dados da cadeia agropecuária com base em padrões da Web e conceito FAIR.)

1. Introdução

Neste documento são apresentados os procedimentos e material para evidenciar o alcance do **Marco 4**, intitulado *Protocolo de metadados para integração e transparência de dados da cadeia agropecuária com base em padrões da Web e conceito FAIR*.¹ do **Compromisso 5** (*Promover a abertura e integração de bases de dados das cadeias agropecuárias com aderência aos interesses da sociedade*) do **5º Plano de Ação Nacional em Governo Aberto**². O protocolo de metadados será utilizado como balizador para a proposta de integração de dados abertos no próximo marco.

2. Métodos

A metodologia aplicada no marco 4 foi o desenvolvimento e criação de um protocolo de metadados com ênfase em dados das cadeias agropecuárias relativas à comercialização.

O processo para a criação do protocolo foi dividido em 4 etapas:

- a. Definição das bases de dados utilizadas para consulta 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 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, todas instituições participantes do Compromisso 5. Foram

¹ O Marco 3 foi coordenado pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa), em parceria com o Observatório do Código Florestal (OCF), Centro de Inteligência Artificial da Universidade de São Paulo (C4AI/USP), Universidade Federal de Minas Gerais (UFMG) e Núcleo de Informação e Coordenação do Ponto BR (NIC.br).

² Disponível em: <<https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/5o-plano-de-acao-brasileiro/5-plano-acao-nacional-30-03-2022.pdf>>, páginas 35-39. Acesso: 27.07.2022.



selecionados os dados dessas fontes correspondentes a informações agropecuárias relativas a dados de comercialização.

- b. 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. Não foram considerados os períodos temporais para essa extração, uma vez que o interesse foi mapear os termos (propriedades de metadados) comuns entre as bases de dados. Também foi verificado o nível de aderência ou detalhamento sobre os dados originais que eram apresentados nos sites das 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.

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

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

- d. 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 preenchimentos são apresentados na etapa de resultados.

3. Análise e discussão de resultados

- a. Etapa para a seleção das fontes de dados utilizadas:

Para a etapa de seleção de fontes, conforme descrito na parte de métodos: “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 é apresentado e exemplificado na Tabela 1. Esse exemplo foi extraído de dados da fonte CEPEA, os quais foram agrupados para o modelo descritivo.

Tabela 1: Exemplos de seleção de fonte com metadado descritivo

| FONTE: CEPEA - Exemplo | | | |
|------------------------|-----------|---|---|
| Campo Abreviado | limitação | 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 | https://www.cepea.esalq.usp.br/br/metodologia/metodologia-acucar- | https://www.cepea.esalq.usp.br/br/metodologia/metodologia- |

Fonte: Dados da pesquisa.

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, IPEA e CONAB e estabelecer metadados comuns entre as três bases, conforme a tabela 2.

Tabela 2: Metadados descritivos da relação entre as três fontes.

| METADADOS | CEPEA | IPEA | CONAB |
|-------------------------|-------|------|-------|
| produtoNome | x | x | x |
| indicador | x | | |
| índice | | x | |
| preçosAgrícolas | | | x |
| regiãoReferência | x | | x |
| data | x | x | x |
| periodicidade | x | x | x |
| unidadeMonetária | x | x | x |
| valorMonetário | x | x | x |
| fonte | x | x | x |
| medidaReferênciaValor | x | x | x |
| medidaReferênciaUnidade | x | x | x |
| metodologiaID | x | x | |

Fonte: dados da pesquisa (2021).

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

Para a definição formal dos termos dos metadados propostos, inicialmente é apresentada na tabela 3 a composição resumida dos elementos dos metadados. No início da tabela 3 são apresentadas as propriedades formais do modelo, inclusive com o nome proposto de AlmesCORE, que se trata de um acrônimo para “Agricultural and Livestock Metadata Element Set Core”. Em sequência, é apresentada a lista de termos e uma sucinta definição dos termos.

Tabela 3: Modelo de metadados - AlmesCORE

| | |
|-------------------------|--|
| Title: | AlmesCORE Metadata Terms |
| Creator: | AlmesCORE Maintenance Group |
| Date issued: | 2022-03-25 |
| Last update: | 2022-05-09 |
| Document Status: | Version 1.0 |
| Description: | Este documento é a versão mais atualizada da especificação dos termos de metadados mantidos pelo C4AI AgriBio Data Management Team, incluindo classes, propriedades, esquemas de codificação de vocabulário, e esquemas de codificação de sintaxe. |

Lista de termos - simplificado (sem a descrição formal)

| Termo | Definição |
|--|--|
| Grupo do produto | Tipos de composição dos produtos. |
| Tema | Indica o tema principal investigado na operação estatística. |
| Nome do produto | Nome do produto agrícola ou pecuário. |
| Nome verbatim | Nome original em linguagem natural pelo qual a série é popularmente conhecida. |
| Editor | A entidade responsável por publicar uma série de dados. |
| Criador | A entidade responsável pela criação dos dados. |
| Referência | um recurso relacionado que é referenciado, citado ou apontado pelo recurso descrito. |
| Tipo de dado | Natureza dos dados que compõem o registro. |
| Periodicidade | Frequência temporal da publicação dos dados. |
| Unidade | Indica a medida ou quantidade do produto. |
| Localização | O nome do local ou região a qual a série de dados se refere. |
| Início do período de referência | Data de início do período de registro da série de dados. |
| Fim do período de referência | Data de encerramento do período de registro da série de dados. |
| Licença | Um documento legal que dá permissão oficial para se fazer algo com o recurso. |
| Metodologia | Resumo dos métodos utilizados para o processo de obtenção de dados para a produção da série. |

A lista completa das propriedades, definições, obrigatoriedades e recomendações estão no anexo 1.

- c. 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 3 etapas:

- i. Foi criado um formulário para preenchimento de modo instrutivo de um registro para o protocolo de metadados, que está disponível em: https://docs.google.com/forms/d/e/1FAIpQLSeMYrOG37BTLsAd8uzGs6qOXVuhNyrew1zhjjIVcBFekIOoUw/viewform?usp=sf_link
- ii. Foi criado um formulário para avaliação da usabilidade do modelo de metadados, que está disponível em:



https://docs.google.com/forms/d/e/1FAIpQLSficEMtfnsx2hSpheKNMW43FId_CaUgKkxF9yDAaQmdW7yisQ/viewform?usp=sf_link

4) Considerações finais

A publicação do esquema de metadados nos padrões da Web Semântica será realizada em etapa posterior. Será aplicada a metodologia de reúso de ontologias para se realizar a especificação formal do esquema de metadados. 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.

Testes de anotação de dados com as propriedades do AlmesCORE serão realizados. Serão criados FAIR Data Points (FDPs) para verificar a aderência dos conjuntos de dados descritos com o conjunto de elementos AlmesCORE aos FAIR Data Principles.

São Paulo, 8 de setembro de 2022

- ANEXO I -

Definição formal dos termos do modelo de elementos de metadados:

| | |
|------------------------|--|
| Title: | AlmesCORE Metadata Terms |
| Creator: | AlmesCORE Maintenance Group |
| Date Issued: | 2022-03-25 |
| Last update: | 2022-05-09 |
| Document Status | Version 1.0 |
| Description: | Este documento é a versão mais atualizada da especificação dos termos de metadados mantidos pelo C4AI AgriBio Data Management Team, incluindo classes, propriedades, esquemas de codificação de vocabulário, e esquemas de codificação de sintaxe. |
| Index of Terms | |
| Label: | Grupo do produto |
| Definition: | Tipos de composição dos produtos. |
| Obligation: | Sim |
| Comment: | A melhor prática recomendada é utilizar um vocabulário controlado. Exemplo: grão, hortifrutí, carne. |
| Label: | Tema |
| URI: | https://metadados.ibge.gov.br/consulta/estatisticos/temas |
| Definition: | Indica o tema principal investigado na operação estatística. |
| Obligation: | Não |
| Comment: | A melhor prática recomendada é utilizar um vocabulário controlado. Exemplo: série de preços, custo, coeficiente técnico ou zootécnico. |
| Reference: | https://metadados.ibge.gov.br/consulta/estatisticos/temas |
| Label: | Nome do produto |
| Definition: | Nome do produto agrícola ou pecuário. |
| Obligation: | Sim |
| Comment: | A melhor prática recomendada é utilizar um vocabulário controlado. Exemplo: soja, milho, boi gordo. |
| Label: | Nome verbatim |
| Type of Term: | propriedade |
| Definition: | Nome original em linguagem natural pelo qual a série é popularmente conhecida. |
| Obligation: | Não |
| Comment: | |
| Label: | Editor |
| URI: | http://purl.org/dc/terms/publisher |
| Definition: | A entidade responsável por publicar uma série de dados. |
| Obligation: | Sim |
| Comment: | A melhor prática recomendada é informar o nome por extenso do editor, juntamente com a URL do respectivo site. |
| Reference: | http://purl.org/dc/terms |

| | |
|--------------------|---|
| Label: | Criador |
| URI: | http://purl.org/dc/terms/creator |
| Definition: | A entidade responsável pela criação dos dados. |
| Obligation: | Não |
| Comment: | A melhor prática recomendada é indicar nome completo do responsável. |
| Reference: | http://purl.org/dc/terms |
| Label: | Referência |
| URI: | http://purl.org/dc/terms/references |
| Definition: | um recurso relacionado que é referenciado, citado ou apontado pelo recurso descrito. |
| Obligation: | Sim |
| Comment: | Este campo deve ser utilizado para indicar a URI da série de dados publicada na fonte original dos dados. |
| Reference: | http://purl.org/dc/terms |
| Label: | Tipo de dado |
| Definition: | Natureza dos dados que compõem o registro. |
| Obligation: | Não |
| Datatype: | String |
| Comment: | A melhor prática recomendada é utilizar um vocabulário controlado. Exemplo: multivariado, univariado, sintético, sequencial, série temporal, texto. |
| Reference: | https://archive.ics.uci.edu/ml/index.php |
| Label: | Periodicidade |
| Definition: | Frequência temporal da publicação dos dados. |
| Obligation: | Não |
| Comment: | Exemplo: diário, mensal, bimestral, anual. |
| Label: | Unidade |
| Definition: | Indica a medida ou quantidade do produto. |
| Obligation: | Sim |
| Comment: | Exemplos: @; sc/15Kg. |
| Label: | Localização |
| Definition: | O nome do local ou região a qual a série de dados se refere. |
| Obligation: | Sim |
| Comment: | A localização pode ser definida utilizando-se de coordenadas geográficas, ou nomes de locais conforme o IBGE ou Geonames < http://www.geonames.org/ >. |
| Label: | Início do período de referência |
| Definition: | Data de início do período de registro da série de dados. |
| Obligation: | Não |
| Comment: | A melhor prática recomendada é adotar esquemas de codificação de datas como a ISO 8601 ou ABNT NBR 5892. |
| Label: | Fim do período de referência |
| Definition: | Data de encerramento do período de registro da série de dados. |
| Obligation: | Não |
| Comment: | A melhor prática recomendada é adotar esquemas de codificação de datas como a ISO 8601 ou ABNT NBR 5892. Caso se trate de uma série contínua, este campo deve ser deixado vazio. |

| | |
|--------------------|--|
| Label: | Licença |
| URI: | http://purl.org/dc/terms/license |
| Definition: | Um documento legal que dá permissão oficial para se fazer algo com o recurso. |
| Obligation: | Sim |
| Comment: | A melhor prática recomendada é atribuir uma licença do tipo Creative Commons, disponível em: < https://creativecommons.org/licenses/by/4.0/deed.pt_BR >. |
| Source: | http://purl.org/dc/terms |
| Label: | Metodologia |
| Definition: | Resumo dos métodos utilizados para o processo de obtenção de dados para a produção da série. |
| Obligation: | Não |
| Comment: | A melhor prática recomendada é indicar a URI do recurso publicado em um formato de acesso aberto na Web. |
| Reference: | https://metadados.ibge.gov.br/consulta/glossario/estatistico |