Center for Artificial Intelligence

A research center to advance artificial intelligence in Brazil

A partnership of:

USP

IBM

FAPESP

Building a Community-Based FAIR Metadata Schema for Brazilian Agriculture and Livestock Trading Data

Paper authors: Filipi Miranda Soares, Fernando Elias Corrêa, Luis Ferreira Pires, Luiz Olavo Bonino da Silva Santos, Debora Pignatari Drucker, Kelly Rosa Braghetto, Dilvan de Abreu Moreira, Alexandre Cláudio Botazzo Delbem, Roberto Fray da Silva, Celso Oviedo da Silva Lopes, Antonio Mauro Saraiva.

Presenter: Filipi Miranda Soares (University of São Paulo, Brazil & University of Twente, Netherlands)

Sept 2022

Introduction

The interoperability of information systems is largely affected by the availability of standards that define common vocabularies and procedures. The large variety of incompatible data practices hinders automated data exchange. Especially in agriculture, interoperability poses a significant challenge due to the multidisciplinary and interdisciplinary nature of the field.

Center for Artificial Intelligence

A research center to advance artificial intelligence in Brazil

A partnership of:

USP

IBM

FAPESP

Introduction

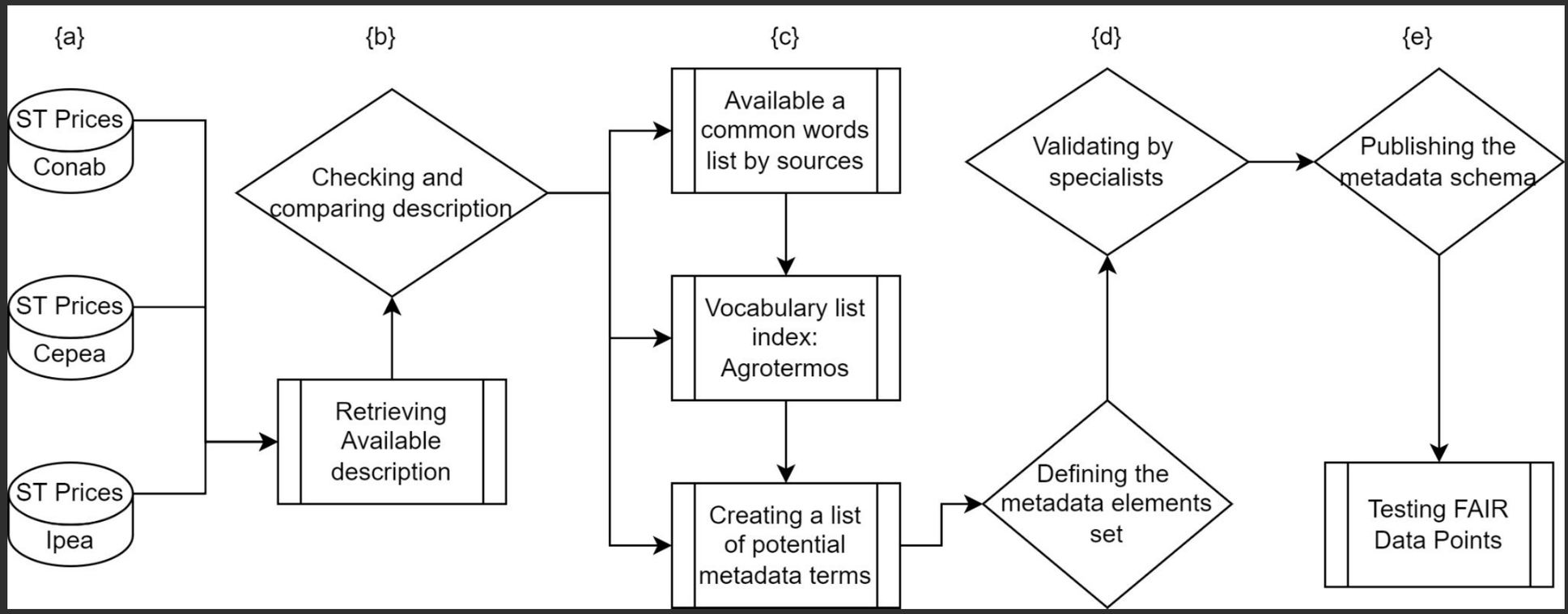
We present an initial effort to build the Agriculture and Livestock Metadata Elements Set (Almes Core), a metadata schema for agricultural trading data to allow data interoperability between three of the main **Brazilian data providers** on agricultural commodities: the **Center for Advanced Studies in Applied Economics (Cepea)**, the **Institute of Applied Economic Research (Ipea)**, and the **National Supply Company (Conab)**.

cepea

ipea Instituto de Pesquisa Econômica Aplicada

Conab Companhia Nacional de Abastecimento

Methodology



Results

- A careful analysis of the Cepea, Ipea and Conab datasets allowed us to understand what kind of data these data sets publish and how they publish it (descriptive information they use to describe data).

Metadada	Cepea	Ipea	Conab
Product name	Produto	nome	produto
Theme	Indicador	índice	preço médio
Location	Região	[not specified]	Unidade da Federação
Start of reference period	Data	data	mês/ano
End of reference period	Data	data	mês/ano
Reference	Fonte	fonte	[not specified]
Methodology	Metodologia	metodologia	[not specified]
Unity	Nota (medidade de referência)	unidade	[part of] produto
License	[not specified]	[not specified]	As informações podem ser reproduzidas, desde que citada a fonte CONAB.
Frequency	PERIODICIDADE	frequência	preços médios mensais ou semanais

Results

- *Product name* and *Product group* properties give the names of products that are the object in the trade operation, so we extracted these names from the datasets to list them.

Cepea Terms	Conab Terms	Legend:	noMatch	closeMatch	totalMatch
	abacaxi				
	abobora				
	abobrinha				
	açaí				
	acerola				
Açúcar	açúcar				
Açúcar cristal	açúcar				
Açúcar refinado	açúcar				
	alcool				
Alface	alface				
	algodão em caroço				
Algodão em pluma	algodão em pluma				

Term Name	Source	URL
Açúcar	Cepea	https://www.cepea.esalq.usp.br/br
Açúcar cristal	Cepea	https://www.cepea.esalq.usp.br/br
Açúcar refinado	Cepea	https://www.cepea.esalq.usp.br/br
Alface	Cepea	https://www.cepea.esalq.usp.br/br
Algodão em pluma	Cepea	https://www.cepea.esalq.usp.br/br
Algodoeira	Cepea	https://www.cepea.esalq.usp.br/br
Arroz em Casca	Cepea	https://www.cepea.esalq.usp.br/br
Avicultor	Cepea	https://www.cepea.esalq.usp.br/br
Banana	Cepea	https://www.cepea.esalq.usp.br/br
Batata	Cepea	https://www.cepea.esalq.usp.br/br
Bezerro	Cepea	https://www.cepea.esalq.usp.br/br

Results

Based on the most representative descriptors common within three datasets, a list of 15 potential metadata properties for agricultural commodities was set:

- ***Product group***: types of products. The best practice is to use a controlled vocabulary such as **Agrotermos**. Ex.: grain, vegetable, meat.
- ***Theme***: main topic investigated in the statistical operation. The best practice is to use a controlled vocabulary. Ex.: regional average price, indicator, index, cost, technical or zootechnical coefficient.
- ***Product name***: name of the agricultural or livestock product. The best practice is to use a controlled vocabulary such as Agrotermos. Examples: soy, corn, cattle.

Results

- ***Verbatim name***: natural language name given to the data series in the original dataset.
- *Publisher*: entity responsible for making the resource available .
- ***Creator***: entity responsible for creating the resource .
- References: related resources that are referenced, cited, or otherwise pointed to by the described resource .
- ***Data type***: kind of the data that make up the record. The best practice is to use a controlled vocabulary. Examples: multivariate, univariate, synthetic, sequential, time series, text.
- ***Frequency***: temporal frequency of data publication. Examples: daily, monthly, bimonthly, yearly.

Results

- ***Unity***: measure or quantity of the product. Examples: 'arroba' (Brazilian unit of weight, equivalent to 15kg, represented by the symbol @); a bag of 30kg.
- ***Location***: place or region to which the data series refers. The best practice is to use geographic coordinates (decimal latitude and longitude) or insert names according to Geonames .
- ***Start of reference period***: date of publication of the first dataset of a data series. The best-recommended practice is to adopt date encoding schemes such as ISO 8601 or ABNT NBR 5892.
- ***End of reference period***: date of publication of the last dataset of a data series. The best-recommended practice is to adopt date encoding schemes such as ISO 8601 or ABNT NBR 5892. In the case of a current series, this field must be left empty.

Results

- ***License***: A legal document that defines the policy to do something with the resource . The best-recommended practice is to adopt a URI to indicate the license.
- ***Methodology***: summary of the methods used to generate the dataset. The best practice is to indicate the resource URI published in an open Web access format .

Final considerations

Next steps:

- *Validating the metadata properties*: specialists from nine Brazilian organizations participating in the 5th Brazilian Open Government Action Plan.
- *Publishing the metadata schema in Semantic Web standards (reusing vocabularies from Agroportal).*
- *Datasets described with the Almes Core properties will be implemented as FAIR Data Points*.

References

Shamin, A., Frolova, O., Makarychev, V., Yashkova, N., Kornilova, L., Akimov, A. Digital transformation of agricultural industry. IOP Conference Series: Earth and Environmental Science, 346(1), (2019). https://doi.org/10.1088/1755-1315/346/1/012029

Saraiva, A. M., Costa, W. F., Xavier, F., Albertini, B. de C., Pfeifer, R. A. C., Júnior, M. A., Simplício, A. K. V. Dados na Agricultura Digital: ciclo, padronização, qualidade, compartilhamento e segurança. In Daniel Marçal Queiroz, D. S. M. Valente, F. de A. C. Pinto, & A. Borém (eds.).: Agricultura Digital. UFV, Viçosa, pp. 308–325 (2020).

Duncan, E., Rotz, S., Magnan, A., Bronson, K. Disciplining land through data: The role of agricultural technologies in farmland assetisation. Sociologia Ruralis 62(2), 231–249 (2022). https://doi.org/10.1111/SORU.12369

Greenberg, J. Understanding Metadata and Metadata Schemes. Cataloging & Classification Quarterly, 40(3-4), 17–36 (2005). https://doi.org/10.1300/J104V40N03_02

Pomerantz, J. Metadata. The MIT Press, Cambridge (2015).

Coyle, K. Understanding Metadata and Its Purpose. The Journal of Academic Librarianship, 31(2), 160–163 (2005). https://doi.org/10.1016/J.ACALIB.2004.12.010

Mayernik, M. S. Metadata. Knowledge Organization, 47(8), 696–713 (2020). https://doi.org/10.5771/0943-7444-2020-8-696

Smiraglia, R. P. Metadata: a cataloger's primer. Routledge, Abingdon-on-Thames (2005).

Chan, L. M., Zeng, M. L. Metadata Interoperability and Standardization - A Study of Methodology Part I. D-Lib Magazine, 12(6), (2006). https://doi.org/10.1045/june2006-chan

Food and Agriculture Organization of the United Nations (FAO). AgMES 1.1 Namespace Specification (2010). http://aims.fao.org/standards/agmes/namespace-specification, last accessed 2022/07/04.

Tullis, T., Albert, B. Measuring the User Experience. 2nd ed. Elsevier, Amsterdam, (2013).

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 1–9 (2016). https://doi.org/10.1038/sdata.2016.18

Bonino, L., Kaliyaperumal, R., Kees Burger, A., Kuzniar, A. Gavai. FAIR Data Point Specification (2021), https://github.com/FAIRDataTeam/FAIRDataPoint/wiki, last accessed 2022/07/04.

Center for Artificial Intelligence

A research center to advance artificial intelligence in Brazil

A partnership of:

USP

IBM

FAPESP

Thank you!

Filipi M. Soares

filipisoares@usp.br

f.mirandasoares@utwente.nl