Business Intelligence and Data Warehouse in Agrarian Sector: A Bibliometric Study

Luciano Moraes da Luz Brum[1], Vinícius do Nascimento Lampert[1,2] & Sandro da Silva Camargo[1]

[1] Graduate Program in Applied Computing, Federal University of Pampa, Bagé, Brazil

[2] Brazilian Agricultural Research Corporation, Bagé, Brazil

Correspondence: Luciano Moraes da Luz Brum, Av. Maria Anunciação Gomes Godoy, 1650, Federal University of Pampa, Bagé, Rio Grande do Sul, Brazil. Tel: 55-539-9959-8544. E-mail: lucianobrum18@gmail.com

Abstract

Business Intelligence with Data Warehouse technologies are known in the literature as solutions that allow access to business data dynamically and analytical operations on them. Scientific literature lacks works that investigate the current use of these technologies in the agrarian sector, at the international level in the last 10 years. This work presents a bibliometric analysis, which was done through the ProKnow-C methodology, of the application of Business Intelligence and Data Warehouse technologies in the agrarian sector. The objective is to investigate the dissemination of such technologies in this sector in national and international scale. The main findings were the following: number of papers in last years are increasing. Majority of papers were found in the journal named Computers and Eletronics in Agriculture, with a great number of colaborations between authors of France. Few colaborations between authors from different countries were found. Sandro Bimonte was the most cited author. France and India highlight in researches approaching Data Warehouse and Business Intelligence usage in agrarian sciences. The majority of references from Bibliographic Portfolio were from 2001-2010. 66% of papers use some open source technology. Star schema is the most used modelling technique and the use of Unified Modeling Language by authors of France in agricultural Data Warehouse modelling is encouraged. The main limitations were the impossibility of free access in some databases, absence of research on proprietary solutions of technology market in the rural sector and few number of keyword searches.

Keywords: agribusiness, decision making, decision support systems, ProKnow-C

1. Introduction

The need for technologies to optimize decision making processes in organizations is increasing, as large volumes of data are generated daily and the requirement for useful strategic information is growing (Shahid et al., 2016). In a similar context of other kind of organizations, rural producers, farmers and professionals in the agrarian sector also need to subsidize their decisions, to optimize productivity or reduce the risks and uncertainties inherent in their activities. In Brazil, for example, the decision-making processes of small and medium-sized rural producers are often based on experience, intuition and informality rather than data, statistics and analytical methods with scientific support (Chavez, Magalhães, Benedetti, Blos, & Silva, 2010; Hofer, Pacheco, Souza, & Protil, 2011). In France, the use of Information Technologies (IT) by farmers is limited and generally, they are non-skilled users (Bimonte et al., 2013).

A concept that relates decision-making processes, large volumes of data, extraction of useful information and knowledge discovery is Business Intelligence (BI). BI has capabilities focused on creating and managing information, allowing employees in an organization to make better decisions (Diaz & Caralt, 2011). One of the components of a BI solution is the Data Warehouse (DW). DW is a subject-oriented, integrated, non-volatile and time-variant data source (Inmon, 2002). The DW, together with ETL (Extraction, Transformation and Load) process, help in organizing, integrating and cleaning the data, as well as facilitating the use of data visualization tools and knowledge extraction, such as data mining tools, online analytical processing (OLAP), reporting, ad-hoc queries, among others. DW is widely used in different sectors, such as: government, business, financial, health, industry, education, agribusiness, among others (Rai, Dubey, Chaturvedi, & Malhotra, 2008; Shahid et al., 2016).

Although there are some BI reviews presented in the literature (Bergeron, 2000; Chee et al., 2009; Obeidat, North, Richardson, Rattanak, North, 2015; Llave, 2017), current and published papers in high impact journals that synthesize the use of these technologies as solutions in the agrarian sector and related areas in an international scope were not found. One of the ways of building current knowledge on a topic is through bibliometrics. Bibliometry is a technique for quantitative analysis that makes it possible to produce statistics to measure the production and dissemination of scientific knowledge on a theme (Araújo, 2006). Bibliometrics can be done by the analysis of citations, collaboration between countries, authors and institutions, quantification of the content of the articles, impact factor, place of publication, keywords, among others (Ellegaard & Wallin, 2015).

It is observed a research gap that this study aims to fill, by answering the following guiding question of this article: What is the scientific production related to the use of BI and DW technologies in the agrarian sector, in national and international conferences and journals? From this issue, questions arise such as: the main authors and countries that contribute to these researches, the main events and journals with relevant publications, the level of interest in this topic over the years, citations and recognition of the works found and quantification of specific aspects of the most relevant works. Therefore, the objective of this review is to present a quantitative bibliometric analysis to review how BI and DW technologies in the agrarian sector have been applied, using the ProKnow-C methodology (L. Ensslin, S. R. Ensslin, Lacerda, & Tasca, 2010).

This study is organized as follows: Section 2 addresses the basic principles of BI and DW. Section 3 presents the materials and methods of the work. Section 4 presents the execution of ProKnow-C and the results found. Section 5 presents a discussion and synthesis of the results and characteristics of the works selected for the final Bibliographic Portfolio (BP) that will contribute scientifically to a better knowledge about the topic of interest. Section 6 presents the final considerations and proposals for future work.

2. Business Intelligence and Data Warehouse

There are different definitions for the term Business Intelligence in literature, but most of them agree on the importance of subsidizing and optimizing the decision-making of an organization using such techniques (Turban, Sharda, & Delen, 2011; Chaudhuri, Dayal, & Narasayya, 2011; Diaz & Caralt, 2011). BI is the set of methodologies, applications, practices and capacities focused on the creation and administration of information, allowing the employees of an organization to make better decisions (Diaz & Caralt, 2011). BI is not a tool or technology, although today it helps on aspects that optimize data storage, processing, analysis and visualization. The goal of BI is "to enable interactive access (sometimes real time) to data, to enable manipulation of data, and to give business managers and analysts the ability to conduct appropriate analysis" (Turban et al., 2011, p. 19).

The generic BI architecture has the following components (Kimball & Ross, 2013):

•   Data sources: The organization's internal or external data repositories, which provide data for the DW. The databases are managed by the Database Management System (DBMS), which are present in systems like Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), Supply Chain Management (SCM) and Online Transaction Processing (OLTP). It is also possible to have repositories like files, spreadsheets, sensor data and unstructured data, among others.

•   Extraction-Transformation-Load (ETL): The data must be extracted from the data repositories, transformed, filtered, processed and redefined. This process targets the data quality for the next step, the load for the DW for analytical processes.

•   Presentation Area: Location where data is organized, stored and made available for user queries, analytic applications and BI tools. Usually, the repository is the DW, a data repository that provides a global, common and integrated view of data—independently how they are going to be used later by users. This kind of repository has the following properties: stable, coherent, reliable and with historic information (Diaz & Caralt, 2011, p. 32, Our traduction)

•   BI applications: They are the capabilities provided to the users to access the presentation area to obtain information for an analytical decision making. It can be simply as an *ad-hoc* query or it can be a complex application of data mining.

DW architectures usually have the following elements:

•   Data Mart (DM): These are a subset of the DW data, after being transformed and filtered in the ETL process, to meet the demands of a specific set of users or department of an organization.

• Stage Area: Intermediate data repository that sits between operational data sources and DW, which is intended to aid in the ETL process and should result in improved data quality.

• Metadata: structured and codified data that helps to describe, organize, identify, discover and administer instances or data. In short, they are data about the data.

Conventional database systems, which follow the Entity-Relationship (ER) model, generally meet operational demands, focusing on availability and performance of transaction processing, like OLTP (Online Transaction Processing) systems, for example. A DW, oriented and optimized for queries and analytical processing, is modeled in dimensional schemas (Kimball & Ross, 2013). Dimensional models have fact and dimension tables. Dimension tables contain descriptive data about the business subjects and fact tables contain quantitative data, metrics, measurements or key-performance indicators about the business. The literature presents the following dimensional models: Star schema—one fact table connected with the dimensions in star format, snowflake squema—one fact table connected with the dimensions in a normalized form and fact constellation—dimensions shared between fact tables (Han, Kamber, & Pei, 2011).

There are other important concepts about BI like DW architecture selection, differences between the modelling schemas, granularity of data, dimensions type, metrics type, among others. It is suggested some recognized books for more details (Kimball & Ross, 2013; Inmon, 2002; Han et al., 2011).

3. Materials and Methods

The research is descriptive and bibliometric, focusing on the investigation of the scientific literature, through articles published in national and international conferences and journals with the theme related to the use of BI and DW in the agrarian sector. The study is characterized as documentary through secondary data of scientific articles available in the bibliographic databases analyzed. The study is qualitative and quantitative for describing and measuring aspects of scientific production.

The ProKnow-C (Knowledge Development Process-Constructivist) intervention instrument is used in this research (Ensslin et al., 2010). Figure 1 shows the number of papers that mention or apply ProKnow-C over the years. This review was conducted in June 2018. The search keyword used was "ProKnow-C" through searches on title, abstract, and keywords (except Google Scholar, where you can search by title or in all text).
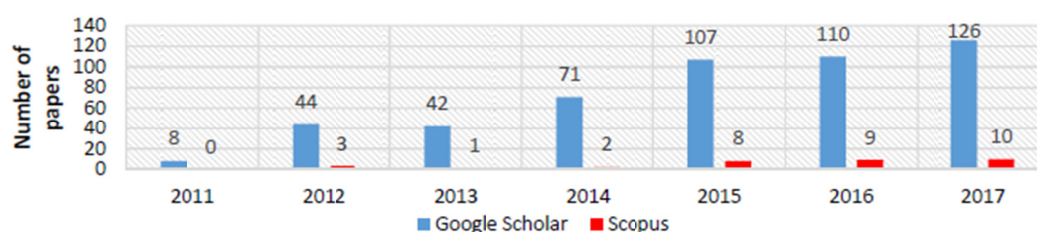

Figure 1. Number of papers by year in Google Scholar and Scopus

Source: Authors, 2018.

It is possible to see an increase in the number of citations to this method over the years in these databases, with few exceptions. This demonstrates a possible growing interest of the scientific community. Some articles that use this methodology have been produced with a significant number of citations (Afonso, Souza, & Ensslin, 2012; L. Ensslin, S. R. Ensslin, & Pacheco, 2012; Waiczyk & Ensslin, 2013).

ProKnow-C has four phases: selection of the BP of articles on the research subject, bibliometric analysis of BP, systemic analysis and definition of the research question and research objective. BP selection is subdivided into four stages: selection of keywords, selection of databases, search of articles in databases with keywords and article filtering. The bibliometric analysis is subdivided in the analysis of four attributes of the articles of the portfolio: relevance of journals, scientific recognition, major authors and most used keywords (Ensslin et al., 2012). The next section will approach ProKnow-C execution.

4. Results

This work is delimited in the following criteria: (i) articles published in journals and conferences, (ii) stages of BP selection and bibliometric analysis of ProKnow-C, (iii) published papers with temporal delimitation (2008-2018), (iv) a few number of keywords used and only one search carried out in each database, (v) databases

and search mechanisms of *CAPES* (Coordination of Improvement of Higher Education Personnel) Periodicals with free access to articles of the agrarian sciences; (vi) alignment of articles with the theme according to researchers' perception.

The execution of ProKnow-C starts from the definition of the research axes of interest of the researchers. Only one search axis was defined: use of BI/DW as solutions in the agrarian sector.

4.1 Bibliographic Portfolio Selection

The selection of BP was performed by selecting the keywords, selecting the appropriate databases for this search, performing the search of the articles in the databases with the keywords and filtering of raw article bank. It is important to emphasize that it is possible that the keywords do not contemplate most of the scientific articles related to the research area. It is necessary to perform a keyword adherence test after selecting the raw article bank to identify the need to adjust the keywords. Text mining for further analysis can also be used.

The keywords defined for the search of articles in the databases and the combinations of logical instructions (AND and OR) used were the following: ("business intelligence" OR "inteligencia de negocio" OR "data warehouse" OR "data warehousing" OR "armazem de dados" OR "almacen de datos") AND (agricultur* OR livestock OR pecuaria OR ganaderia OR agribusiness OR agronegocio OR agroindustria).

The next step was the definition of the bibliographic databases to execute the searches. Only databases and search engines available in *CAPES* and aligned with the area of knowledge relevant to the research topic (Agrarian Sciences) were considered. *CAPES* plays a key role in the expansion and consolidation of the stricto sensu postgraduate and graduate courses in Brazil. *CAPES* has a virtual library that brings together and makes available to teaching and research institutions in Brazil the international scientific production. Of the 197 possibilities, were selected: *AGRIS, BDPA, EBSCO, Google Scholar, ProQuest, SciELO, ScienceDirect, SCOPUS, SpringerLink, Taylor & Francis Online, Web of Science* and *Wiley Online Library*. In the *CAPES* portal, only the categories "complete text", "references with abstracts" and "search tools" were considered.

The search for articles was mostly done in the title, abstract, and keyword fields simultaneously. In search engines that did not allow this combination, the search were performed in the full text. 1435 papers in twelve databases with defined keywords were found. Figure 2 shows the number of publications per database. In *Google Scholar*, 20200 results with the keywords initially set were found, but the tool displays only the first 100 pages. Only the works of first ten pages of the search tool, sorted by relevance, were imported, resulting in 173 papers.
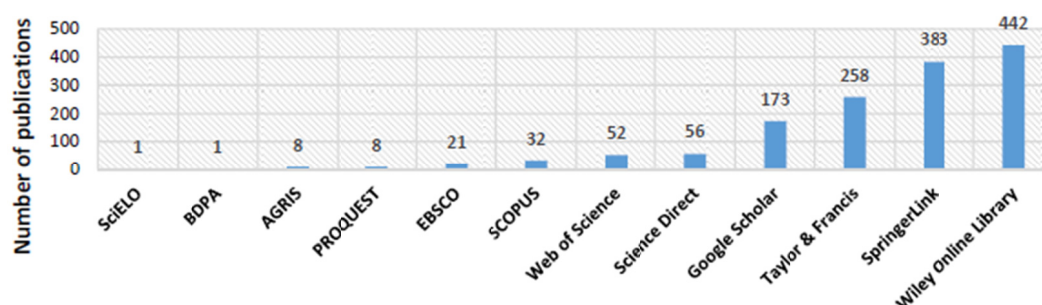

Figure 2. Number of articles found in databases

Source: Authors, 2018.

Finally, the adherence of the keywords to the research topic is tested. In this step, a randomized paper was selected from each database to verify the adherence of the keywords, totaling 12 papers. 77% of the selected papers had at least one of the keywords defined at the beginning of this research, therefore such words were considered adequate for this research.

In order to perform the filtering of the 1435 works, five steps were performed sequentially with the help of the following tools: *Mendeley v. 1.19.1* and *Excel 2013*.

a)   Elimination of redundant works (170 of 1435, 11.85%). Figure 3 shows the year distribution of the works found without the redundancies.

Figure 3. Publications number by year

Source: Authors, 2018.

b)   Reading titles and abstracts for the discarding of papers not relevant to the topic in research. If abstracts and titles did not address research and implementations of BI and DW solutions in the agrarian sector, they were eliminated. Some works did not have the abstract available, so they were also discarded (1108 of 1265, 87.59%).

c)   Only the articles with open access, written in english, portuguese or spanish languages, with correct year of publication and without duplicates were considered (51 of 157; 32.48%).

d)   Analysis of the number of citations of the articles and analysis of the year for articles excluded for reinsertion. *Google Scholar* was used to analyze the citations of each paper. Papers that had at least one citation were considered, being therefore removed 25 works. Of the works removed, 8 were reinserted into the set of articles included for the next step, since they are very recent works (2017 and 2018) and did not have opportunities to be cited (17 of 106; 16.04%). Some works consider that for the reinsertion of articles in BP, they must be recent and have at least one of the authors of the most cited papers of BP, as criteria for reinsertion. This approach were not considered due to small number of publications and to provide visibility for papers and authors who have not had this opportunity yet.

e)   Full reading of the remaining 89 papers. In this stage, 60 papers were discarded (60 of 89, 67.42%) due to not showing sufficient aspects of a DW or BI solution applied in agrarian sector. Objectively, it was considered if the articles presented the dimensional models, architecture of the solution, technologies and methods used and the problem to be solved. Also, the adequacy with this investigation and presentation of consistent results were considered. The remaining 29 were evaluated in the bibliometric analysis stage.

4.2 Bibliometric Analysis

In this section is presented the analysis of BP and bibliographic references of the articles present in BP. The information quantified and analyzed were: citations, time distribution of papers in BP, papers by author and periodicals, and journals relevance. The same information were analyzed from the BP references, with the exception of citations. The objective of this stage is to analyze the information and generate knowledge about the subject, considering the scientific recognition of the articles, the most relevant periodicals, most important authors and analysis of keywords of the works.

4.2.1 Bibliometric Analysis of BP

Table 1 shows the 29 papers present in BP, sorted by citations.

Table 1. Selected papers to BP

Title	Periodical/Conference	Authors	*Cit.	Year
[1] An integrated approach for agricultural ecosystem management	IEEE Transactions on Systems, Man, and Cybernetics, Part C	Xu, L., Liang, N., & Gao, Q.	114	2008
[2] Dimensional issues in agricultural data warehouse designs	Computers and Eletronics in Agriculture	Nilakanta, S., Scheibe, K., & Rai, A.	60	2008
[3] Analysis of mealybug incidence on the cotton crop using ADSS-OLAP (Online Analytical Processing) tool	Computers and Eletronics in Agriculture	Abdullah, A.	29	2009
[4] The use of UML to design agricultural data warehouses	Agricultural Engineering	Pinet, F., et al.	27	2010
[5] Precise design of environmental data warehouses	Operational Research	Pinet, F., & Schneider, M.	24	2010
[6] EIS Pesticides: An environmental information system to characterize agricultural activities and calculate agro-environmental indicators at embedded watershed scales	Agricultural Systems	Vernier, F., et al.	21	2013

[7] Multidimensional modeling and analysis of large and complex watercourse data: An OLAP-based solution	Ecological Informatics	Boulil, K., Le Ber, F., Bimonte, S., Grac, C., & Cernesson, F.	20	2014
[8] Design and development of data mart for animal resources	Computers and Eletronics in Agriculture	Rai, A., Dubey, V., Chaturvedi, K. K., & Malhotra, P. K.	20	2008
[9] A data warehouse of muscle characteristics and beef quality in France and a demonstration of potential applications	Italian Journal of Animal Science	Chriki, S. et al.	14	2013
[10] Definition and analysis of new agricultural farm energetic indicators using spatial OLAP	Lecture Notes in Computer Science	Bimonte, S., Boulil, K., Chanet, J., & Pradel, M.	13	2012
[11] A quality-aware spatial data warehouse for querying hydroecological data	Computers & Geosciences	Berrahou, L., et al.	11	2015
[12] On-line Analytical Processing in Agriculture using Multidimensional Cubes	J. of the Indian Society of Agricultural Statistics	Chaturvedi, K. K., Rai, A., Dubey, V., & Malhotra, P. K.	11	2008
[13] Guaranteeing the quality of multidimensional analysis in data warehouses of simulation results: Application to pesticide transfer data produced by the MACRO model	Ecological Informatics	Boulil, K., et al	9	2013
[14] Spatial OLAP integrity constraints: From UML-based specification to automatic implementation: Application to energetic data in agriculture	Journal of Decision Systems	Boulil, K., Bimonte, S., & Pinet, F.	9	2014
[15] Data integration as the key to building a decision support system for groundwater Management: Case of Saiss aquifers, Morocco	Groundwater for Sustainable Development	Laraichi, S., Hammani, A., & Bouignane, A.	8	2016
[16] Multidimensional Schema for Agricultural Data Warehouse	Int. J. of Research in Engineering and Technology	Gupta, A. K., & Mazumdar, B. D.	7	2013
[17] Spatial Online Analytical Processing for Hotspots Distribution Based on Socio-economic Factors in Riau Province Indonesia	Procedia Environmental Sciences	Thariqa, P., & Sitanggang, I. S.	5	2015
[18] The data storage and analysis system of the Swiss National Forest Inventory	Computers and Eletronics in Agriculture	Traub, B., Meile, R., Speich, S., & Rösler, E.	5	2017
[19] Multidimensional analysis model for highly pathogenic avian influenza using data cube and data mining techniques	Biosystems Engineering	Xu, Z., Lee, J., Park, D., & Chung, Y	4	2017
[20] Integrated modeling of agricultural scenarios (IMAS) to support pesticide action plans: the case of the Coulonge drinking water catchment area (SW France)	Environmental Science and Pollution Research	Vernier, F., et al.	3	2016
[21] BovReveals: uma plataforma OLAP e data mining para tomada de decisão na pecuária de corte	Brazilian Congress of Agroinformatics	Mota, F. M., Souza, K., Ishii, R., & Gomes, R. D. C.	2	2017
[22] A data warehouse to explore multidimensional simulated data from a spatially distributed agro-hydrological model to improve catchment nitrogen management	Environmental Modelling & Software	Bouadi, T., et al.	1	2017
[23] A system for the rapid design and implementation of Personalized Agricultural Key Performance Indicators issued from sensor data	Computers and Eletronics in Agriculture	Bimonte, S., Naoufal, E., & Gineste, L.	1	2016
[24] An Online Analytical Processing (OLAP) Database for Agricultural Policy Data: a Greek Case Study	CEUR Workshop Proc.	Maliappis, M., & Kremmydas, D.	1	2015
[25] Aplicação de Business Intelligence nos dados de diagnóstico de unidades familiares de produção na Paraíba: um estudo de caso na EMATER-PB	InterScientia	Junior, N. M. C., & Farias, T. M. T.	1	2017
[26] Data Cubes Integration in Spatial OLAP for Agricultural Commodities	Earth and Environmental Science	Putri, A. I., & Sitanggang, I. S.	1	2017
[27] A new decision-support system for the historical analysis of integrated pest management activities on olive crops based on climatic data	Computers and Eletronics in Agriculture	Zazaa, C., et al	0	2018
[28] An Efficient Data Warehouse for Crop Yield Prediction	Int. Conference on Precision Agriculture	Ngo, V. M., Le-Khac, N. A., & Kechadi, M. T.	0	2018
[29] Investigating Factors that Influence Rice Yields of Bangladesh using Data Warehousing, Machine Learning, and Visualization	Int. J. on Modern Education and Computer Science	Ahmed, F., Nandi, D., Rahman, M., & Hasan, K. T.	0	2017

Note: The * means the number of citations.

Source: Authors, 2018.

The first element analyzed is the number of papers' citations and the cumulative frequency of citations, as shown in Figure 4. It is possible to observe three papers without citations. These papers were reinserted in the review process due to the recent year of publication. It is verified that 89% of citations are concentrated in 13 of the 29 papers (44.83%).
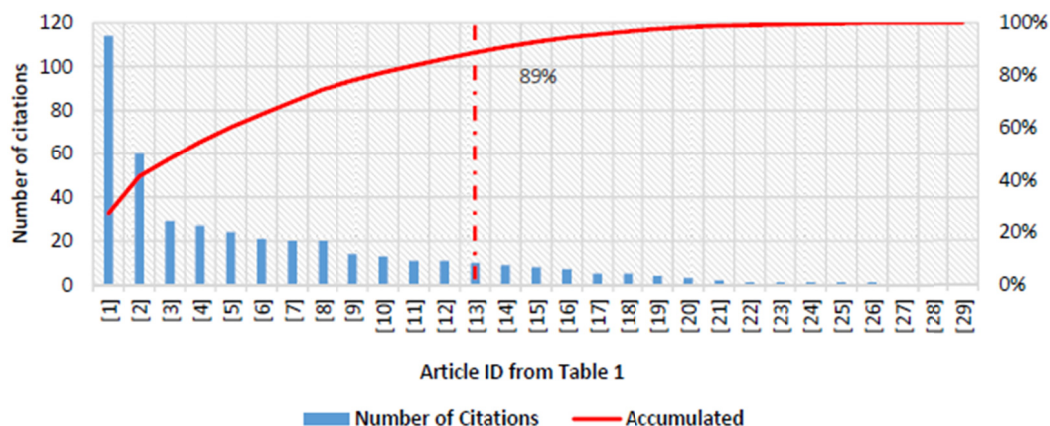

Figure 4. BP's papers citations and cumulative frequency

Source: Authors, 2018.

In Figure 5, the citations and the number of BP's papers per year are shown. As expected, most citations focus on the earliest period (83% of citations in the period 2008-2013).
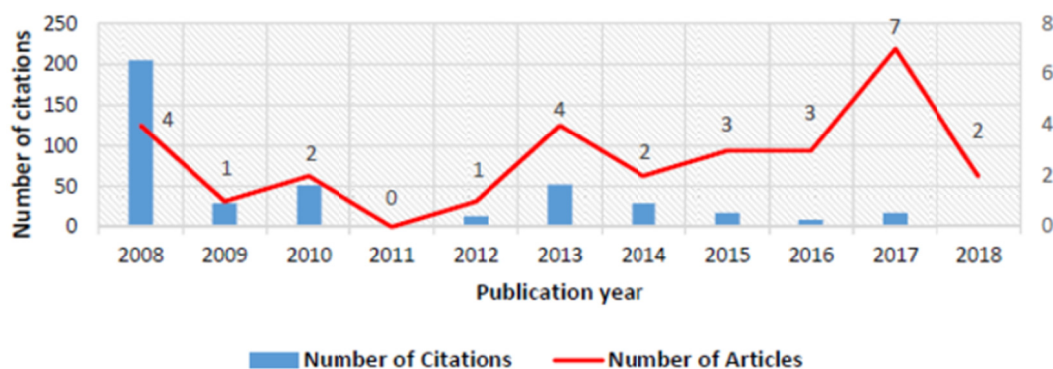
Figure 5. Number of BP's works and it's citations per year

Source: Authors, 2018.

Figure 6 shows, among the total of 97, the authors with the highest number of BP's publications. The authors with one publication in BP were not considered. The six authors with the greatest number of works (except Rai, A.) develop their activities in a collaborative way, all of them being from institutions of France.
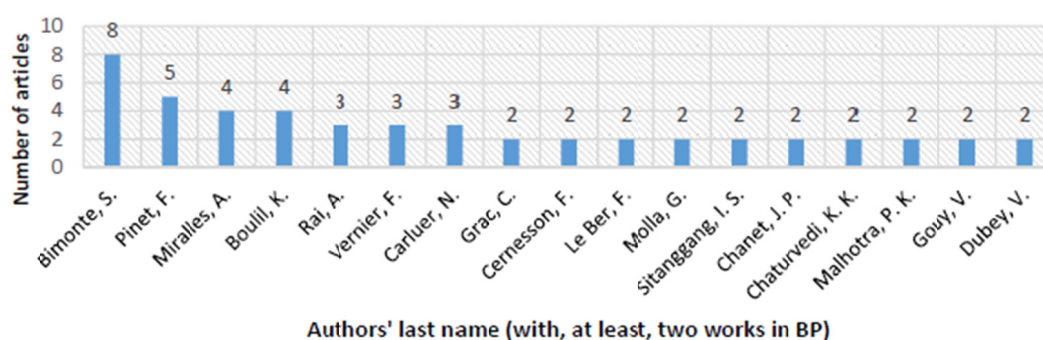
Figure 6. Papers in BP by author

Source: Authors, 2018.

Figure 7 presents an analysis of the collaboration between authors in the BP's papers. The vertex colors indicates the authors' affiliation country and the vertex size represents the number of collaborations with different authors. Bimonte, S. and Miralles, A. carry out a greater number of works in collaboration with other authors and France has excelled in the scientific production in DW in the agrarian sciences with 13 papers of BP followed by India with 4 papers. In this count, the presence of at least one of the authors of a given country by paper is considered. Three authors are affiliated with two institutions from different countries.
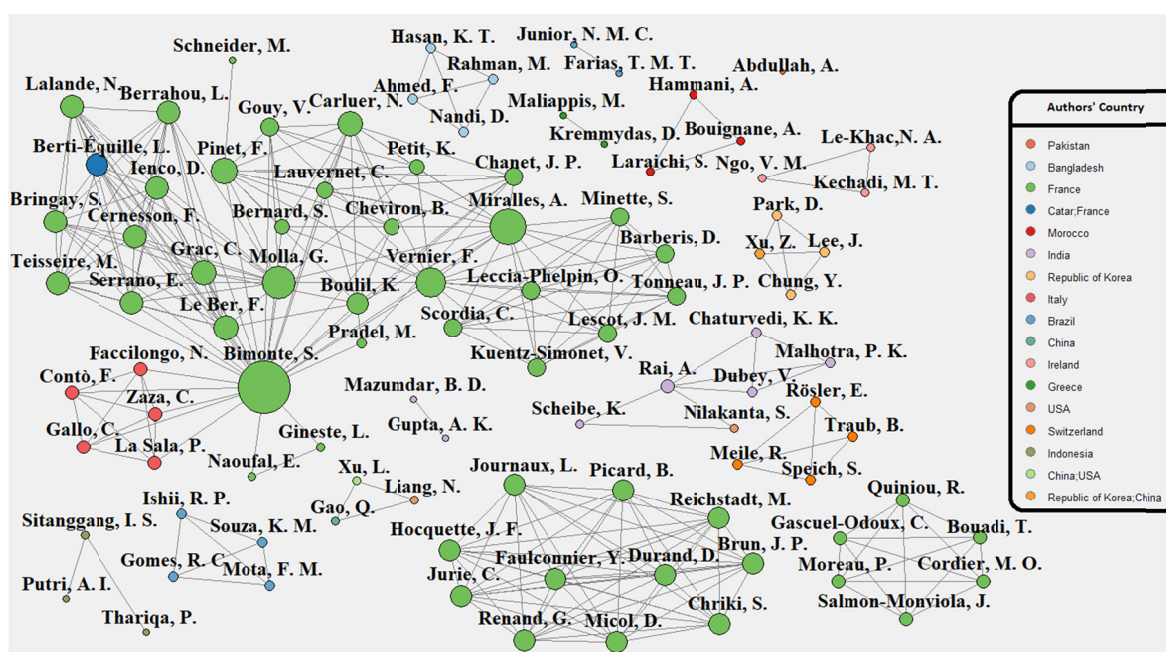

Figure 7. Analysis of the collaboration network between authors and countries in BP. Source: Authors, 2018

From BP's papers, 3% (n = 1) were done with one author, 21% (n = 6) with two authors, 21% (n = 6) with three authors, 24% (n = 7) with four authors and 31% (n = 9) with more than four authors. This number shows that the development of works on BI and DW technologies in the agrarian sector has required a considerable number of people. This can be explained by the complexity of the subjects and by the interaction of at least two areas of knowledge: computer science to the developers and Information Technology (IT) experts and agrarian sciences to the problem domain experts.

The journals with the largest number of BP's publications are Computers and Electronics in Agriculture, which owns 6 of the 29 BP's publications and aggregates 27.25% of the citations, and Ecological Informatics, which owns 2 of the 29 BP's publications and aggregates 7.11% of the citations. The publication vehicle of the other 21 papers appeared only once.

4.2.2 Bibliometric Analysis of BP's References

The bibliometric analysis of the references was carried out considering only the authors and periodicals of the works present in the BP. Figure 8 shows the time distribution of the works present in the BP references, including any type of reference and possible repetitions between them. It is possible to observe a great concentration of citations to works between 2000 and 2010 and works done before 1995 have a smaller number of citations.
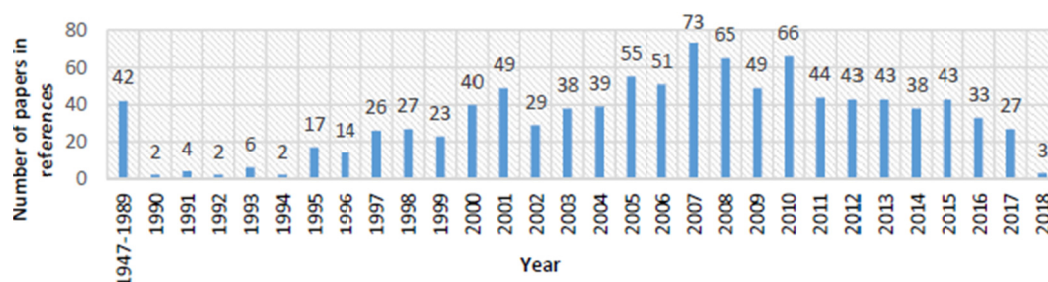

Figure 8. Time distribution of the references present in the BP's articles

Source: Authors, 2018.

Figure 9 presents the main articles and authors in a combined analysis of the number of citations from the BP's studies and the number of citations to the most cited author of the papers in the BP's references. The numbers represent each of the works from Table 1. The cut-off point defined in the Figure 9 for the delimitation of relevance of the works was 20% of the articles highlighted in the attributes under analysis. The Pareto principle was used to define the cut-off points.
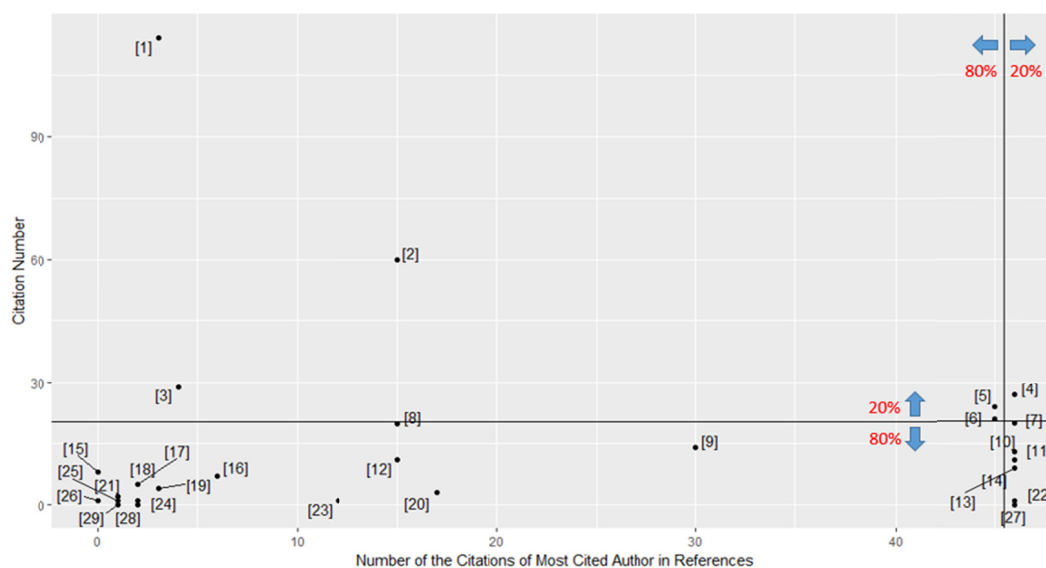

Figure 9. Citations to BP's papers and most cited authors by work present in references

Source: Authors, 2018.

In the group of highlighted articles and authors in the quadrant I, stand out the Pinet et al. (2010). In the quadrant II, in the group of highlighted articles related to citations, with more than 20, are presented the works of Xu, Liang, and Gao (2008), Nilakanta, Scheibe, and Rai (2008), Abdullah (2009), Pinet and Schneider (2010), and Vernier et al. (2013). In the quadrant IV, in the group of prominent authors with more than 45 citations in the references of PB, the works of Boulil et al. (2014), Bimonte et al. (2012), Berrahou et al. (2015), Boulil et al. (2013), Boulil et al. (2014), Bouadi et al. (2017) and Zaza et al. (2018) are hightlighted. In quadrant III are the

remaining 15 works, with 20 citations or less to the articles and 47 citations or less to the most cited author of the works in the BP references.

Figure 10 presents a combined analysis of periodicals and conferences of BP's articles and their references. The shape's color identifies the journals and conferences. The shape's size is directly proportional to the journal's JCR (*Journal Citation Report*) value for the year of 2017. The JCR is one of the journals' impact assessment metrics in the scientific community and is calculated as follows:

$$JCR = \frac{Sum\ of\ Citations\ in\ the\ Last\ Two\ Years}{Sum\ of\ all\ Publications\ in\ Last\ Two\ Years} \qquad (1)$$

Figure 10 shows four groups of journals and conferences. *Computers and Eletronics in Agriculture* is highlighted in quadrant I, with six articles in BP and it appears more than 40 times in BP's articles references, showing that this vehicle is the most appropriate and relevant for publications related to BI and DW solutions applied in the agrarian sciences. *Ecological Informatics* appears in quadrant II with two articles in BP, but appears eight times in BP references. *Environmental Modelling & Software* is highlighted in quadrant IV, with only one article in the BP and appears 21 times in BP references and also is the journal with the highest JCR from the BP's articles. The remaining 20 journals and conferences appears in quadrant III, each one with only one article in BP and all of them appears between 0 and 8 times in BP articles.
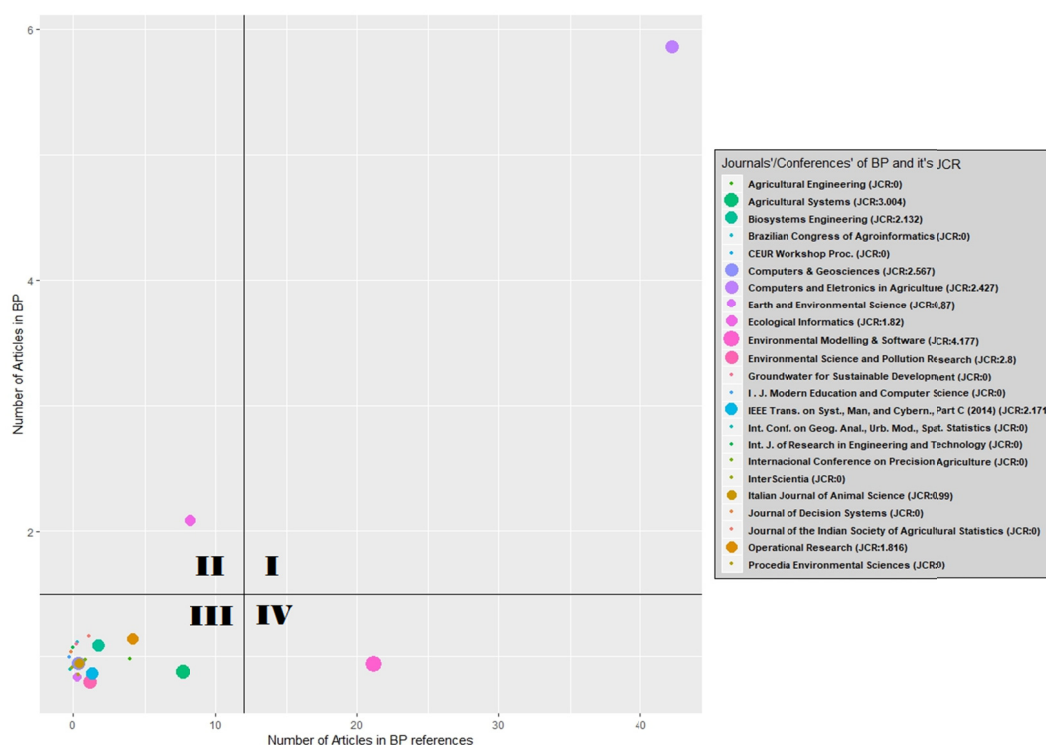

Figure 10. Relevance of periodicals and conferences of BP articles and its references

Note. The x-axis shows how many articles, by journal or conference, appeared in references from the BP. The y-axis represents how many articles were found for the BP, by journal or conference. The color just identifies each periodical and conference. The size of the circle is proportional to the JCR. To prevent overplotting, specifically in quadrant III, the jitter function was used.

Source: Authors, 2018.

Figure 11 shows a word cloud with the most used keywords in BP's articles. The word clouds are a conventional and popular solution to represent, in a graphical way, the words frequency of a document (Viégas, Wattenberg, & Feinberg, 2009). The color has no meaning, but the size reflects the frequency of the keyword. Three papers of BP do not present any keyword in its portable document format (pdf) (Bimonte, Naoufal, & Gineste, 2016; Putri & Sitanggang, 2017; Pinet et al., 2010). The words "Data Warehouse", "Decision Support Systems" and "Business Intelligence" were united, to analyse the frequency of these words of interest. The eight most frequent

keywords from the group of 126 found, accumulating 32.90% of total number, were: Data Warehouse (n = 16), Model (n = 14), OLAP (n = 12), Data (n = 9), Agriculture (n = 9), System (n = 6), Multidimensional (n = 5) and Spatial (n = 5). Also, 35.71% (n = 5) of the initial keywords defined were present in articles of final BP. This analysis suggests that the words "OLAP", "Spatial" and "multidimensional" are suitable for this kind of research and could be used in future works.

Figure 11. Keywords extracted from articles of BP

Source: Authors, 2018.

5. Discussion and Synthesis

Some synthesis are presented below, from papers in quadrants I and II from Figure 9. Pinet et al. (2010) presents the use of UML to design an agricultural DW to analyse the pesticides use in agriculture, in France, from a spatial perspective. Xu, Liang, and Gao (2008), presents an Agricultural Ecosystem Enterprise Information System, that extracts data on terrain, land use, planting, and others, and integrates them for agricultural and ecosystem management. Nilakanta, Sheibe, and Rai (2008) show some problems in dimensional modeling of a DW for the Indian government, to agricultural sector. The main issues approached were the granularities of location and time dimensions, overlapping time domains and aggregation and disaggregation of information at different dimensional hierarchies. Abdullah (2009) presents an Agriculture Decision Support System with detailed pest-scouting data from more than 2300 farmers of Pakistan, from 2005 and 2006. This system was used for comparison of which group of pesticides contributed to the control of the mealybug through analysis of pesticide and time dimension in two periods. Pinet and Schneider (2010) approach the use of UML for building multidimensional models and choosing different criteria according to analysis requirements. The method is used for analyze the use of agricultural fertilizers, by amout, per commune and fertilizer. Vernier et al. (2013) present an environmental information system (EIS Pesticide) using Spatial DW technology to analyse the impact of agriculture, through pesticide use, at different embedded watershed scales. UML was used and made the communication easier between the participants involved.

Main Technologies: It is worth mention that 66% (n = 19) of papers mentioned the use of at least one open source or free software to develop its solutions. Figure 12 shows the main technologies and methods used as solutions in all 29 papers from BP. Some papers did not mention any detail about the implementations or technologies used. Open source technologies in scientific environment are more used and encouraged in this subject.

About the techniques used to model the DW, 13 papers applied Star schema, 6 applied Snowflake schema and 4 applied the Constellation schema. Also, 8 papers use Unified Modeling Language (UML) to help in interactions between computer scientists and the domain experts to design the solutions, mainly used by authors from France (Pinet & Schneider, 2010; Pinet et al., 2010; Bimonte et al., 2012; Boulil et al., 2013; Vernier et al., 2013; Boulil et al., 2014a; Boulil et al., 2014b).
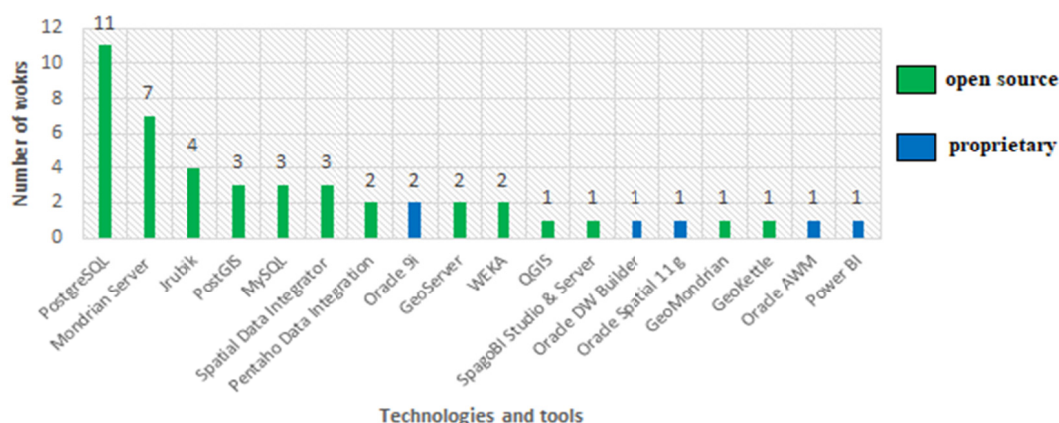
Figure 12. Main Technologies used in articles of BP

Source: Authors, 2018.

The Figure 13 shows the papers number distributed by study area:
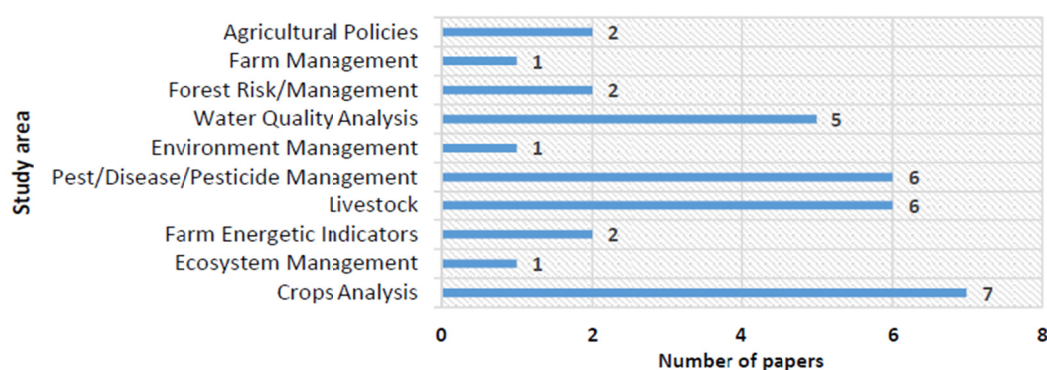

Figure 13. Main areas of study of papers selected

Source: Authors, 2018.

Main challenges and issues pointed out: choice of adequate tools for DW/BI development; Quality and means to access available data and information (historical data can be stored in handwritten books or spreadsheets); Uniformity of data collection; Care with data concepts, measurement units and granularity level in different systems; Establishing data patterns to simplify the integration process; Tools and methods to get and guaranteeing data quality in the DW (bad data can lead to bad decisions); DW architecture choice; Information and data ownership; Stakeholders and institutions must be convinced of the benefits provided by such technologies; Highly efficient interactive user interfaces and search tools (Traub et al., 2017); and heterogeneous agricultural production systems (crops grown, climatic conditions, soils types, cultural conditions, among others) make the development of agricultural systems a challenge (Chaturvedi et al., 2008).

6. Conclusions and Future Work

This study presented a bibliometric analysis of scientific literature about the uses of BI and DW in agrarian sector. Different graphics were used to summarize and present the information retrieved. The majority of papers aligned with this research were found in the journal Computers and Eletronics in Agriculture, with a great number of colaborations between authors of France. Few colaborations between authors from different countries were found. Bimonte was the most cited author, with 8 works in BP and 46 citations in BP references. France and India have been highlighted in researches approaching DW/BI usage in agrarian sciences. Some papers were found from China, but we decided to not use them in this research because they are not open access, not allowing other researchers to reproduce our results. The works from BP show that the majority of citations concentrate on old papers, but many recent works approach the DW and BI usage in different subareas of agrarian sector in

many countries. The majority of references from BP were from 2000-2010. Highlighted articles from BP had more than 20 citations and more than 45 citations to the most cited author from BP's papers in bibliographic references. Most of the works are using open-source and free technologies to develop their solutions in agrarian sciences. It could be seen the high use of PostgreSQL and MySQL DBMS and also the use of BI suites like Pentaho and SpagoBI.

Also, the majority of the challenges to develop DW/BI solutions to agrarian sector are regarding to: data format, technologies and modelling techniques used, common and easy communication between developers (computer scientists) and the experts from the specific domain and farmers, and the heterogeneity of source systems. Usually, those are the same problems faced in other DW/BI application areas.

Acknowledgements

References

Abdullah, A. (2009). Analysis of mealybug incidence on the cotton crop using ADSS-OLAP (Online Analytical Processing) tool. *Computers and Electronics in Agriculture, 69*, 59-72. https://doi.org/10.1016/j.compag.2009.07.003

Afonso, M. H. F, Souza, J. V., & Ensslin, S. R. (2012). Como construir conhecimento sobre o tema de pesquisa? Aplicação do processo Proknow-C na busca de literatura sobre avaliação do desenvolvimento sustentável. *Revista de Gestão Social e Ambiental, 5*, 47-62. https://doi.org/10.24857/rgsa.v5i2.424

Ahmed, F., Nandi, D., Rahman, M., & Hasan, K. T. (2017). Investigating Factors that Influence Rice Yields of Bangladesh using Data Warehousing, Machine Learning, and Visualization. *International Journal on Modern Education and Computer Science, 3*, 36-47. https://doi.org/10.5815/ijmecs.2017.03.05

Araújo, C. A. (2006). Bibliometria: evolução histórica e questões atuais. *Em Questão, 12*(1), 11-32.

Bergeron, P. (2000). Regional business intelligence: The view from Canada. *Journal of Information Science, 26*(3), 153-160. https://doi.org/10.1177/016555150002600305

Berrahou, L., Lalande, N., Serrano, E., Molla, G., Berti-Équille, L., Bimonte, S., ... Teisseire, M. (2015). A quality-aware spatial data warehouse for querying hydroecological data. *Computers & Geosciences, 85*, 126-135. https://doi.org/10.1016/j.cageo.2015.09.012

Bimonte, S., Boulil, K., Chanet, J. P., & Pradel, M. (2012). *Definition and analysis of new agricultural farm energetic indicators using spatial OLAP*. International Conference on Geographical Analysis, Urban Modeling, Spatial Statistics, May 18-20, Salvador, Bahia, Brazil. https://doi.org/10.1007/978-3-642-31075-1_28

Bimonte, S., Edoh-Alove, E., Nazih, H., Kang, M., & Rizzi, S. (2013). ProtOLAP: Rapid OLAP prototyping with on-demand data supply. *Proceedings of the sixteenth international workshop on Data warehousing and OLAP (DOLAP)* (pp. 61-66). San Francisco, California, USA. https://doi.org/10.1145/2513190.2513199

Bimonte, S., Naoufal, E., & Gineste, L. (2016). A system for the rapid design and implementation of personalized agricultural key performance indicators issued from sensor data. *Computers and Electronics in Agriculture, 130*, 1-12. https://doi.org/10.1016/j.compag.2016.09.010

Bouadi, T., Cordier, M. O., Moreau, P., Quiniou, R., Salmon-Monviola, J., & Gascuel-Odoux, C. (2017). A data warehouse to explore multidimensional simulated data from a spatially distributed agro-hydrological model to improve catchment nitrogen management. *Environmental Modelling & Software, 97*, 229-242. https://doi.org/10.1016/j.envsoft.2017.07.019

Boulil, K., Bimonte, S., & Pinet, F. (2014). Spatial OLAP integrity constraints: From UML-based specification to automatic implementation: Application to energetic data in agriculture. *Journal of Decision Systems, 23*, 460-480. https://doi.org/10.1080/12460125.2014.934120

Boulil, K., Le Ber, F., Bimonte, S., Grac, C., & Cernesson, F. (2014). Multidimensional modeling and analysis of large and complex watercourse data: An OLAP-based solution. *Ecological Informatics, 24*, 90-106. https://doi.org/10.1016/j.ecoinf.2014.07.001

Boulil, K., Pinet, F., Bimonte, S., Carluer, N., Lauvernet, C., Cheviron, B., ... Chanet, J. P. (2013). Guaranteeing the quality of multidimensional analysis in data warehouses of simulation results: application to pesticide transfer data produced by the macro model. *Ecological Informatics, 16*, 41-52. https://doi.org/10.1016/j.ecoinf.2013.04.004

Chaturvedi, K. K., Rai, A., Dubey, V., & Malhotra, P. K. (2008). On-line Analytical processing in Agriculture using Multidimensional Cubes. *Journal of the Indian Society of Agricultural Statistics, 62*(2), 56-64.

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM, 54*(8), 88-98. https://doi.org/10.1145/1978542.1978562

Chavez, R. Q., Magalhães, A. M., Benedetti, O. I. S., Blos, A. L. F., & Silva, T. N. (2010). Tomada de decisão e empreendedorismo rural: um caso da exploração comercial de ovinos de leite. *Revista Brasileira de Gestão e Desenvolvimento Regional, 6*(3), 3-21.

Chee, T., Chan, L. K., Chuah, M. H., Tan, C. S., Wong, S. F., & Yeoh, W. (2009). Business intelligence systems: State-of-the-art review and contemporary applications. *Symposium on Progress in Information and Communication Technology, 2*(4), 96-101.

Chriki, S., Picard, B., Faulconnier, Y., Micol, D., Brun, J. P., Reichstadt, M., ... Hocquette, J. F. (2013). A data warehouse of muscle characteristics and beef quality in France and a demonstration of potential applications. *Italian Journal of Animal Sciences, 12*, 247-256. https://doi.org/10.4081/ijas.2013.e41

Díaz, J. C., & Caralt, J. C. (2012). *Introducción al Business Intelligence*. Editorial UOC, Barcelona.

Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics, 105*, 1809-1831. https://doi.org/10.1007/s11192-015-1645-z

Ensslin, L., Ensslin, S. R., & Pacheco, G. C. (2012). A study on safety at football matches based on the analysis of the international literature. *Perspectivas em Ciências da Informação, 17*(2), 71-91. https://doi.org/10.15 90/S1413-99362012000200006

Ensslin, L., Ensslin, S. R., Lacerda, R. T. O., & Tasca, J. E. (2010). *ProKnow-C, Knowledge Development Process-Constructivist*. Processo técnico com patente de registro pendente junto ao INPI, Brasil.

Ensslin, S. R., Ensslin, L., Imlau, J. M., & Chaves, L. C. (2014). Processo de mapeamento das publicações científicas de um tema: portfólio bibliográfico e análise bibliométrica sobre avaliação de desempenho de cooperativas de produção agropecuária. *Revista de Economia e Sociologia Rural (Impresso), 52*(1), 587-608. https://doi.org/10.1590/S0103-20032014000300010

Google. (2018). *Google Scholar*. Retrieved from http://scholar.google.com.br

Gupta, A., & Mazumdar, B. (2013). Multidimensional schema for agricultural data warehouse. *International Journal of Research in Engineering and Technology, 2*(3), 245-253. https://doi.org/10.15623/ijret.2013.02 03006

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Elsevier/Morgan Kaufmann Publishers.

Hofer, E., Pacheco, V., Souza, A., & Protil, R. M. (2011). A relevância do controle contábil para o desenvolvimento do agronegócio em pequenas e médias propriedades rurais. *Revista Contabilidade e Controladoria, 3*(1), 27-42. https://doi.org/10.5380/rcc.v3i1.21490

Inmon, W. H. (2002). *Building the Data Warehouse* (3rd ed.). JohnWiley & Sons.

Junior, N. M. C., & Farias, T. M. T. (2017). Aplicação de Business Intelligence nos dados de diagnóstico de unidades familiares de produção na Paraíba: Um estudo de caso na EMATER-PB. *InterScientia, 5*(1), 143-152.

Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd ed.). John Wiley & Sons, Inc.

Laraichi, S., Hammani, A., & Bouignane, A. (2016). Data integration as the key to building a decision support system for groundwater management: case of Saiss aquifers, Morocco. *Groundwater for Sustainable Development, 2*, 7-15. https://doi.org/10.1016/j.gsd.2016.04.003

Llave, M. R. (2017). Business Intelligence and Analytics in Small and Medium-Sized Enterprises: A Systematic Literature Review. *Procedia Computer Science, 121*, 194-205. https://doi.org/10.1016/j.procs.2017.11.027

Maliappis, M., & Kremmydas, D. (2015). An online analytical processing (OLAP) database for agricultural policy data: A Greek case study. *Proceedings of the 7th International Conference on Information and Communication Technologies in Agriculture, Food and Environment (HAICTA 2015), Kavala, Greece, September 17-20, 2015* (pp. 214-225).

Mota, F. D., Souza, K., Ishii, R., & Gomes, R. D. C. (2017). Bovreveals: uma plataforma olap e data mining para tomada de decisão na pecuária de corte. *Brazilian Congress of Agroinformatics, 11*, 29-39.

Ngo, V. M., Le-Khac, N.-A., & Kechadi, M.-T. (2018). An Efficient Data Warehouse for Crop Yield Prediction. *Proceedings of the 14th International Conference on Precision Agriculture, June 24-27, 2018, Montreal, Quebec, Canada*.

Nilakanta, S., Scheibe, K., & Rai, A. (2008). Dimensional issues in agricultural data warehouse designs. *Computers and Electronics in Agriculture, 60*, 263-278. https://doi.org/10.1016/j.compag.2007.09.009

Obeidat, M., North, M., Richardson, R., Rattanak, V., & North, S. (2015). Business intelligence technology, applications, and trends. *International Management Review, 11*(2), 47-56.

Pinet, F., & Schneider, M. (2010). Precise design of environmental data warehouses. *Operational Research, 10*(3), 349-369. https://doi.org/10.1007/s12351-009-0069-z

Pinet, F., Miralles, A., Bimonte, S., Vernier, F., Carluer, N., Gouy, V., & Bernard, S. (2010). *The use of UML to design agricultural data warehouses*. AgEng 2010, International Conference on Agricultural Engineering, Clermont-Ferrand, France.

Putri, A. I., & Sitanggang, I. S. (2017). Data cubes integration in spatial OLAP for agricultural commodities. *IOP Conference Series: Earth and Environmental Science, 58*. https://doi.org/10.1088/1755-1315/58/1/012034

Rai, A., Dubey, V., Chaturvedi, K. K., & Malhotra, P. K. (2008). Design and development of data mart for animal resources. *Computers and Electronics in Agriculture, 64*, 111-119. https://doi.org/10.1016/j.compag.2008.04.009

Shahid, M., Sheikh, U., Raza, B., Shah, M., Kamran, A., Anjum, A., & Javaid, Q. (2016). Application of Data Warehouse in Real Life: State-of-the-art Survey from User Preferences' Perspective. *International Journal of Advanced Computer Science and Applications, 7*(4), 415-426. https://doi.org/10.14569/IJACSA.2016.070455

Thariqa, P., & Sitanggang, I. S. (2015). Spatial online analytical processing for hotspots distribution based on socio-economic factors in Riau Province Indonesia. *Procedia Environmental Sciences, 24*, 277-284. https://doi.org/10.1016/j.proenv.2015.03.036

Traub, B., Meile, R., Speich, S., & Rösler, E. (2017). The data storage and analysis system of the Swiss National Forest Inventory. *Computers and Electronics in Agriculture, 132*, 97-107. https://doi.org/10.1016/j.compag.2016.11.016

Turban, E., Sharda, R. E., & Delen, D. (2011). *Decision Support and Business Intelligence Systems* (9th ed.). [S.l.]: Bookman.

Vernier, F., Leccia-Phelpin, O., Lescot, J. M., Minette, S., Miralles, A., Barberis, D., … Tonneau, J. P. (2016). Integrated modeling of agricultural scenarios (IMAS) to support pesticide action plans: The case of the Coulonge drinking water catchment area (SW France). *Environmental Science and Pollution Research, 24*, 6923-6950. https://doi.org/10.1007/s11356-016-7657-2

Vernier, F., Miralles, A., Pinet, F., Carluer, N., Gouy, V., Molla, G., & Petit, K. (2013). EIS Pesticides: an environmental information system to characterize agricultural activities and calculate agro-environmental indicators at embedded watershed scales. *Agricultural Systems, 122*, 11-21. https://doi.org/10.1016/j.agsy.2013.07.005

Viégas, F. B., Wattenberg, M., Feinberg, J. (2009). Participatory Visualization with Wordle. *IEEE Transactions on Visualization and Computer Graphics, 15*(6), 1137-1144. https://doi.org/10.1109/TVCG.2009.171

Waiczyk, C., & Ensslin, E. R. (2013). Avaliação de produção científica de pesquisadores: mapeamento das publicações científicas. *Revista Contemporânea de Contabilidade, 10*(20), 97-112. https://doi.org/10.5007/2175-8069.2013v10n20p97

Xu, L., Liang, N., & Gao, Q. (2008). An integrated approach for agricultural ecosystem management. *IEEE Transactions on Systems, Man, and Cybernetics, Part C, 38*, 590-599. https://doi.org/10.1109/TSMCC.2007.913894

Xu, Z., Lee, J., Park, D., & Chung, Y. (2017). Multidimensional analysis model for highly pathogenic avian influenza using data cube and data mining techniques. *Biosystems Engineering, 157*, 109-121. https://doi.org/10.1016/j.biosystemseng.2017.03.004

Zaza, C., Bimonte, S., Faccilongo, N., La Sala, P., Conta, F., & Gallo, C. (2018). A new decision-support system for the historical analysis of integrated pest management activities on olive crops based on climatic data. *Computers and Electronics in Agricultur*e, *148*, 237-249. https://doi.org/10.1016/j.compag.2018.03.015

Copyrights