CC&T

ISSNe 0104-1096

Cadernos de Ciência & Tecnologia

www.embrapa.br/cct

Database Models in Agriculture: a systematic review

ABSTRACT – Access to organized agricultural information represents a trend in the development of modern agriculture. Synthesis and analysis of the literature on data models used in agricultural databases can guide future decision making. In this study, we present a systematic review covering the period from 2010 to 2023, using keywords in scientific databases to identify recent approaches in agriculture using database models. We identified several approaches involving modeling, integration, and management of agricultural data. This review provides a comprehensive overview of the latest methodologies, presenting studies that examine data models specific to agriculture, with an emphasis on the area of plant breeding. Besides, techniques and methods available in the literature that can be used in building databases — considering specific needs for making decisions on agricultural data — are pointed up.

Index terms: database models, relational model, database integration.

Modelos de Bancos de Dados na Agricultura: uma revisão sistemática

RESUMO – O acesso a informações agrícolas organizadas representa uma tendência no desenvolvimento da agricultura moderna. A síntese e a análise de literatura sobre os modelos de dados utilizados em bancos de dados agrícolas podem orientar futuras tomadas de decisão. Neste estudo, apresentamos uma revisão sistemática referente ao período de 2010 a 2023, utilizando palavras-chave em bases de dados científicas para identificar abordagens recentes na agricultura com o uso de modelos de bancos de dados. Identificamos diversas abordagens que envolvem a modelagem, integração e gerenciamento de dados agrícolas. Esta revisão oferece uma visão abrangente das metodologias mais recentes, apresentando trabalhos que exploram modelos de dados específicos para a agricultura, com ênfase na área de melhoramento vegetal. Além disso, técnicas e métodos disponíveis na literatura que podem ser utilizados na construção de bancos de dados — considerando as necessidades específicas para a tomada de decisão sobre dados agrícolas — são destacados.

Termos para indexação: modelos de banco de dados, modelo relacional, integração de banco de dados.

Herica Fernanda de Sousa Carvalho ✉
Renato Archer, Information Technology Center (CTI), Campinas, SP, Brazil.
E-mail: hfscarvalho@cti.gov.br

Luiza Maria Capanema Bezerra
Agronomic Institute (IAC), Campinas, SP, Brazil.
E-mail: luiza.bezerra@sp.gov.br

Ferrucio de Franco Rosa
Renato Archer, Information Technology Center (CTI), Campinas, SP, Brazil.
E-mail: ferrucio.rosa@cti.gov.br

✉ Corresponding author

Received
August 28, 2024

Accepted
June 2, 2025

Published
December 19, 2025

How to cite
CARVALHO, H.F. de S.; BEZERRA, L.M.C.; ROSA, F. de F. Database Models in Agriculture: a systematic review. **Cadernos de Ciência & Tecnologia**, v.42, e27757, 2025. DOI: https://doi.org/10.35977/0104-1096.cct2025.v42.27757.

INTRODUCTION

Currently, the ability to collect, store, process, and analyze data has become essential to support decision making in many areas, such as economics, society, the environment, health, education, security, and even the daily life of citizens. This is exemplified most recently by artificial intelligence tools, which generally offer solutions based on large databases. Agriculture also benefits from this advance. The use of information organized in databases in the agricultural sector is not new – agrometeorological data, for example, have long been collected and processed to support decision making.

The collection, organization, and dissemination of this type of information began in the 1970s with the climatology unit of the Instituto Agronômico (IAC) (Romani et al., 2016). Although it may seem simple, the choice of sowing or harvest time can significantly impact the yield and income of an agricultural production unit. To build a collection of detailed information on a certain crop, it is necessary to adopt categorically hierarchical methodologies (Souza Filho et al., 2008). Thus, the way the data are related and stored in databases must follow a structure understandable to the user. This is only possible through building a well-structured database.

Databases are built up based on some fundamental principles to establish relationships during the steps of collecting, storing, processing, and retrieving the information, ensuring that it is consistent and reliable (Martinez-Mosquera et al., 2020). According to Mendes et al. (2023), modern digital agriculture depends on increasingly interconnected and organized data that can be accessed by a range of agricultural innovation agents to benefit various stages of crop and livestock production. Therefore, due to the large amount of interrelated data involved, increasingly sophisticated systems are required to support organization of and access to this information (Whairit et al., 2023).

A database is a source of information fed by data. A database follows one or several data models, i.e., how and which data can be structured and, at times, which operations are available for data retrieval and handling (Taipalus, 2024a). A crucial element for structuring a database is the data model, which includes a collection of tools for semantic description of the entities, as well as their relationships and restrictions.

The models represent logical organization of the data, including a structure for their storage, methods for retrieval, and management within a given system. They are classified as conceptual, physical, and logical, along with their subdivisions. An example of a conceptual model is one known as the Entity-Relationship Model (Chen, 1976), which describes the interrelationships and characteristics of the data. A logical data model (Codd, 1970) characterizes how the data are stored and presented (Taipalus, 2024a). Models with a non-relational logic (NoSQL) are widely used in the context of big data.

However, despite the increasing adoption of data technologies in agriculture, there is a significant gap in the literature regarding the systematization of knowledge about the different database models applied to agriculture. Therefore, this review's contribution is to provide a clear and updated overview of the topic, highlighting the existing gaps and offering directions for future research, to strengthen the integration between data technology and agricultural practices.

Therefore, this study was motivated by the need to know which database models are most effective in making information available in the agricultural sector, considering structure and content. The aim is to answer the following research question: *What types of data models, methods, and techniques are being applied in creating agricultural databases?*

The main contribution of this review is to classify and synthesize state-of-the-art studies that discuss data methods, techniques, and models applied to the field of agriculture.

The aim of this study is to identify, analyze, and show the evolution of articles in agriculture

Cadernos de Ciência & Tecnologia, Brasília, v. 42, e27757, 2025
DOI: 10.35977/0104-1096.cct2025.v42.27757

and key modeling techniques being applied to agricultural databases.

METHODOLOGY

The systematic review presented in this article is qualitative and informative. It discusses studies on database models developed to organize and use data in the agricultural sector and it covers a period from 2010 to 2023.

A search was conducted in January 2024 in four scientific databases: Web of Science, Scopus, IEEE Xplore, and ACM Digital Library. These databases were selected due to their relevance and coverage of the areas of computer engineering and database development, and due to the recognized quality and volume of information they provide (Pestana et al., 2024). Search filters were applied to titles and keywords using the following search strings: ("*Database*" OR "*Relational Database*" OR "*Database Management System*" OR "*Data Model*" OR "*Data Integration*") AND ("*Crop*" OR "Agriculture" OR "*Farm*" OR "*Plant Biology*"). This strategy was adopted to find studies that address the convergence between database development and applications in agriculture, precision agriculture, and plant biology. The search returned 351 scientific articles.

A total of 99 duplicate documents were excluded. The abstracts and methodologies of the documents were then read to confirm their alignment with the aim of the study, considering the following inclusion criteria: i) use of agricultural data and ii) presentation of information that identifies the data model. A total of 227 documents did not meet the inclusion criteria.

A final selection resulted in 25 articles for examination of methods and techniques considering the eligibility criteria as well as their direct relationship with the aim of the study. A diagram of the article selection process is shown in Figure 1.

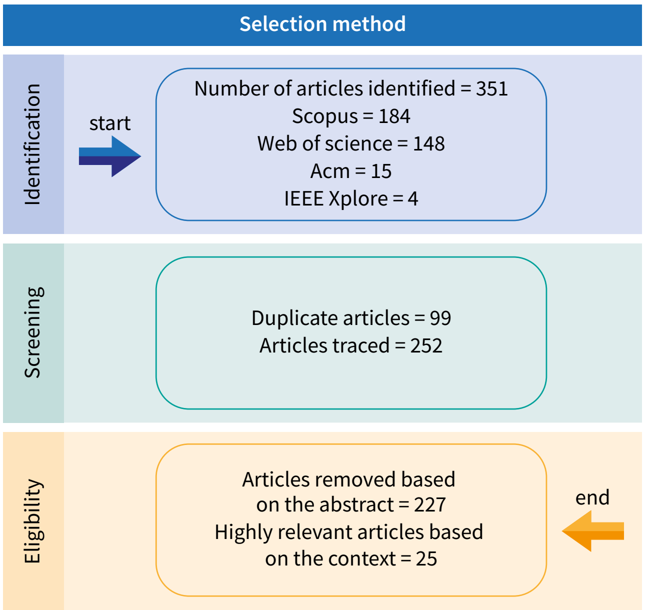

Figure 1. Diagram of the article selection process adapted from Walters & Light (2021).

FINDINGS

In general, the documents analyzed include articles published in periodicals, conference proceedings, and book chapters. The studies are mainly within the areas of agricultural and biological sciences and computer engineering. Analysis of the number of publications in general per periodical over the period analyzed shows growth in research results, and the Scopus database is the most prominent, with 52.5% of the documents, as shown in Figure 2. This increase can be attributed to growth in the volume of data generated in recent years.

Whereas older database formats fall out of use or have limited applications, more modern web tools are becoming predominant and incorporating functions of similar systems. According to Stonebraker & Çetintemel (2005), the adoption of web-based technologies allows greater flexibility and scalability, which are essential characteristics for efficient management of agricultural data. These modern tools, such as NoSQL databases and cloud data management systems, offer significant advantages in terms of performance and ability to handle large volumes of data (Stonebraker & Çetintemel, 2005).

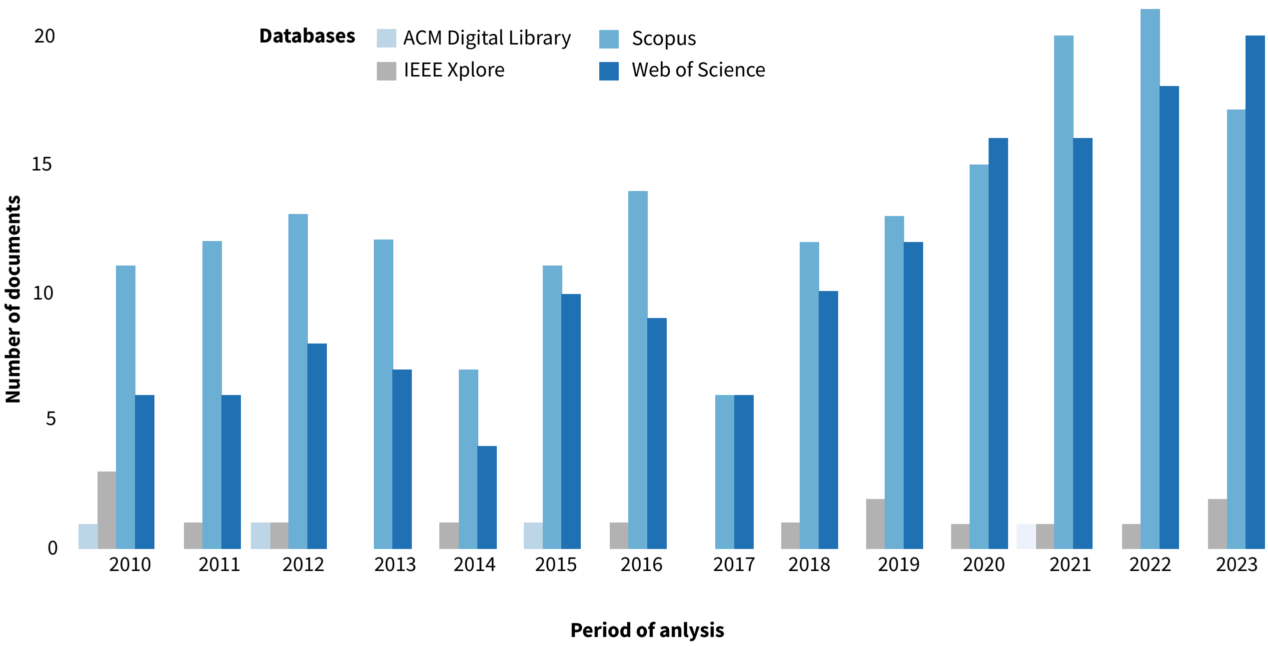

Figure 2. Total number of documents per year in each scientific database.

In addition, integration of artificial intelligence techniques with agricultural databases has the potential to transform agricultural practices, improving efficiency and decision making (Kamilaris et al., 2018). Therefore, interdisciplinary use of data from the areas of biological sciences, computing, statistics, and mathematics allows significant advances in creating new technologies, such as interactive databases.

Figure 3 presents a global distribution of eligible publications by country, considering the corresponding author. It can be observed that the country with the highest number of publications on the study's theme is Switzerland, totaling seven articles. Following Switzerland, the Netherlands, the United Kingdom, China, and the United States delivered considerable contributions, reflecting the growing interest in data models to optimize agriculture, especially given the vast agricultural areas and the importance of these nations for global food production. On the other hand, other countries presented a reduced quantity of publications. This may indicate a concentration of more recent or less developed studies on this topic or even an issue still under development in these regions.

To present the results section, we present four literature review studies together with this review (Table 1). These studies discuss methods and techniques related to data models and their applications in the agricultural domain. They can be classified according to their respective objectives — Literature Review (LR), Database Construction (DBC), Database Models (DBM), Database Integration (DBI — and application subdomains — Plant breeding, Data modeling, and Agricultural Management.

The articles that presented a LR have some points in common regarding the more recent need for managing large volumes of data, integrating databases (especially those that store information on plant genetic resources), and developing platforms that are user-friendly and accessible for different user profiles. For example, in the review of Lai et al. (2012), the authors analyzed characteristics of germplasm database repositories for plant breeding, highlighting specific resources for breeding of individual crops and the potential future use of these databases. Dhanapal & Govindaraj (2015) examined terms with fundamental definitions and considered it necessary to create and use agricultural databases, especially genome

Cadernos de Ciência & Tecnologia, Brasília, v. 42, e27757, 2025
DOI: 10.35977/0104-1096.cct2025.v42.27757

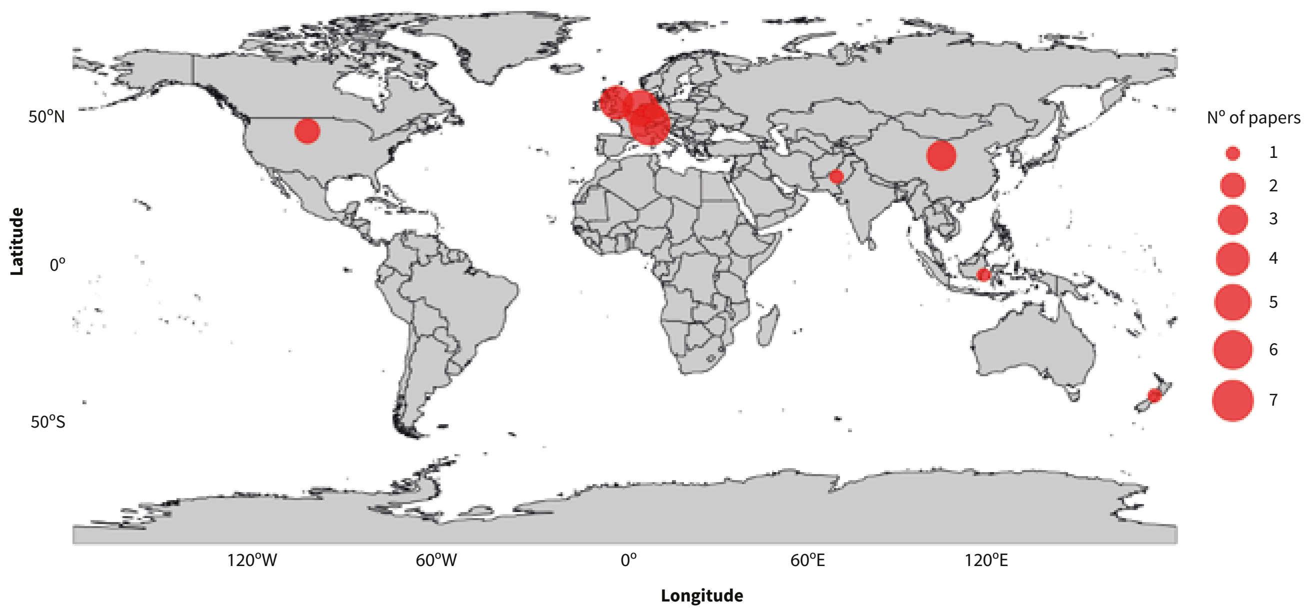

Figure 3. Global distribution of eligible papers considering the corresponding author.

database platforms. Both studies emphasize that a common challenge in agricultural biotechnology is managing the increasing volume of data and presenting the data in user-accessible interfaces.

Cabrera & Fadul-Pacheco (2021) analyzed the main databases reported in the scientific literature, focusing on innovative technologies for data integration. Their analysis revealed that most studies use various artificial intelligence algorithms. However, they emphasized that data integration is particularly complex in the agricultural context due to the heterogeneity of the data. Despite this difficulty, the authors highlighted that the move toward integration is essential for improving decision-making, especially given the increasing volume of available data.

Chao et al. (2023) emphasized the importance of integrating databases on plant breeding. Challenges associated with data integration were discussed, enhancing how this process is essential for comprehensive understanding of the biological processes that affect plant traits and their interactions in the growing environment. The authors underlined that effective integration of genetic databases is essential for advancement of plant breeding, since analysis and interpretation of these data require advanced computational tools.

Databases are essential technological tools that consolidate information from multiple sources, providing a unified and accessible view (Taipalus, 2024b). Creating and managing databases involve practices, methods, techniques, and modeling essential for effective operation of modern information systems. In this context, database models (DBM) play a fundamental role in decision making in agriculture.

Therefore, the articles within the scope of this review will be presented, addressing topics regarding database models applied to agriculture, followed by a section on database integration. Twenty-one articles noted below were analyzed, as well as classification in subdomains, and they can be observed in Table 2. These articles were classified according to their respective objectives — Database Construction (DBC), Database Models (DBM), Database Integration (DBI), Management Systems (MS) — and application subdomains — Plant breeding, Precision agriculture, Agricultural Management, Data modeling, and System development.

Table 1. Synthesis of reported literature review studies.

Authors	Objective				Domain of application		
	LR	DBC	DBM	DBI	1	2	3
Cabrera & Fadul-Pacheco (2021)	x			x			x
Chao et al. (2023)	x			x	x		
Dhanapal & Govindaraj (2015)	x	x			x		x
Lai et al. (2012)	x	x			x		
This study	x		x	x		x	x

Objectives: Literature Review (LR), Database Construction (DBC), Database Models (DBM), Database Integration (DBI); Application domains: Plant breeding (1), Data modeling (2), Agricultural management (3).

Table 2. Synthesis of the studies analyzed within the scope of this research.

Authors	Objective				Application domain		
	DBC	DBM	DBI	MS	1	2	3
Awad et al. (2019)	x	x				x	x
Bhawna et al. (2015)	x	x			x		
Bhermana & Susilawati (2020)	x			x	x		x
Chao et al. (2020)	x				x		
Duhan et al. (2020)	x				x		x
Duhan & Kaundal (2021)	x				x		
Gomez-Cano et al. (2020)	x	x	x		x		
Gong et al. (2022)	x	x	x		x		
Hai-Ping et al. (2013)		x	x				x
Haiping et al. (2010)					x		x
Ke et al. (2015)				x	x		
San Emeterio de la Parte et al. (2023)		x	x			x	
Moot et al. (2021)				x			x
Mochida et al. (2013)	x				x		
Ngo & Kechadi (2020)		x				x	
Nizar et al. (2021)	x	x				x	
Rezník et al. (2015)		x				x	
Shi et al. (2023)							
Timlin et al. (2023)		x		x		x	
Wu et al. (2011)	x		x				x
Zheng et al. (2019)	x				x		

Objectives: Database Construction (DBC), Database Models (DBM), Database Integration (DBI), Management System (MS).; Application domains: Plant breeding (1), Precision agriculture (2), System development (3).

Cadernos de Ciência & Tecnologia, Brasília, v. 42, e27757, 2025
DOI: 10.35977/0104-1096.cct2025.v42.27757

Articles dealing with DBM generally present applications based on the relational model, which will be presented below. These documents highlight thematic information on genomics, plant breeding, and molecular markers; and this last item is an important tool for research in both genomics and in plant breeding. Databases that offer solutions for improving agricultural yield indices also stand out, the main objective of which is to support decision making within the sphere of the agricultural property.

Moot et al. (2021) present a web repository (called AgYields) with the aim of improving agricultural production systems by adding information on plants and offering guidelines for future data collection through appropriate models. The authors emphasize the need to create databases and standardize data collection to maximize the value of information stored in common formats.

Bhermana & Susilawati (2020) propose a database system as an integral part of a Decision Support System (DSS) in Kalimantan City, Indonesia. Using a relational database model, all data were initially classified according to crop group, i.e. food crops, fruit crops, ornamental crops, and medicinal crops. Species classification was then determined in each group, followed by addition of relevant information to provide useful information regarding general description, original habitat, spatial distribution, recommendation for ex situ and in situ conservation and propagation area, cultivation techniques and benefits value (for medicinal crops).

Awad et al. (2019) present a proposal for an interactive database tool (CSSIT – Crop Spectral Signatures Database Interactive Tool) to store spectral signatures of diverse crops. The aim was to refine the crop mapping and, consequently, facilitate yield estimates. The relational model was used in the database. Using a relational model, Duhan et al. (2020) built a database called citSATdb, a web-genome resource for citrus crop species, storing the data on MySQL. A graphic interface was developed using the PHP, HTML5, and JQuery languages. Chao et al. (2020) likewise present a database on gene expression profile for oilseed seeds (BrassicaEDB), using the relational model.

Duhan & Kaundal (2021) present a database for legume crops (LegumeSSRdb), which contains molecular markers of simple sequence repeats (SSRs). Zheng et al. (2019) developed a genomic database (CuGenDB) for crops of the *Cucurbitaceae* family. The authors used the open-source Tripal Toolkit (Sanderson et al., 2013), which uses the relational DBM. Tripal is a toolkit for building online portals of community biological databases (genetics, genomics, etc.), and it is used to create genome sites. Gomez-Cano et al. (2020) propose the structured database CamRegBase, using an entity-relationship diagram to integrate genetic resource data of a wheat species. An open-source toolkit was used in building the web database. Data in CamRegBase are free of charge to users.

Shi et al. (2021) propose a database using crop portraits, containing information extracted from 15 crops, 28 pesticides, and 93 diseases and pests. To do so, Neo4j (Nikam et al., 2020) was used as a graphed database management system. The crop portraits, a kind of property graph, model the characteristics of real-world crops based on data, allowing networked management of agricultural knowledge. According to the authors, the results were positive for agricultural analysis and have practical applications, as well as perspectives for theoretical research in the field of big data analysis.

Bhawna et al. (2015) propose a database of molecular markers for plants of the Cucurbitaceae family with the aim of identifying and developing a dataset using the relational model to provide information widely used in population genetics for mapping and identifying individuals. Ke et al. (2015) developed the ocsESTdb database to store information on molecular markers of oilseed seeds. The development process involved organizing the data, establishing logical relationships among the data, and creating a web-based user interface. The platform was developed using the Perl/CGI, Python, and JavaScript languages running on an Apache web server hosted on the CentOS operating system, using a MySQL database management system. The database is divided into sections and is available for research and download of results.

San Emeterio de la Parte et al. (2023) propose a semantic DBM for management of precision agriculture data. The aim is to allow insertion and retrieval of information in real time, using minimal resources in managing large volumes of data. The proposal focused on modeling non-relational time series database highlighting that NoSQL databases offer greater versatility and performance in terms of scalability of the data volume. The study discussed the structure, the DBM, and the management system of databases.

Nizar et al. (2021) designed and developed an accessible database to store information on underutilized crops. The aim was to present a DBM and describe its implementation and possible uses. The authors chose a relational DBM due to its robustness in prototyping and building user interfaces to govern and disseminate data. Additionally, an automated application programming interface (API) and a graphical user interface (GUI) were developed to allow online access to the database by users and administrators. Timlin et al. (2023) present a graphical user interface (CLASSIM) to manage the input and output (I/O) of data in a database. The I/O data were managed in a relational database (SQLite). Furthermore, a database structure with SQL was used to facilitate input, storage, and retrieval of data, also allowing more advanced analyses. The outputs of the model are presented in tables and graphs, allowing visualization and analysis of the results.

As computation in agriculture has advanced, several diverse heterogeneous agricultural databases have emerged, requiring greater integration among them. Sharing agricultural information has become a trend in the development of agriculture in recent years, requiring more integrated and interoperable solutions (Shi et al., 2021). Concerning that trend, documents related to database integration (DBI) have shown greater attention to integration and sharing of heterogeneous agricultural databases, including information on omics data, precision agriculture, environmental monitoring, and contamination of soil, water, and air by agricultural inputs. Wu et al. (2011) developed a platform (system) for sharing heterogeneous agricultural databases. The system was implemented using XML, allowing users to access the heterogeneous databases using a web browser. According to the authors, the platform proved to be stable, reliable, and capable of meeting the requirements of information sharing in agriculture.

Rezník et al. (2015) developed an open data model for precision agriculture applications and agricultural pollution monitoring using various methods, including the relational data model. The open data model developed for these applications was registered in the GEOSS (Global Earth Observation System of Systems) to meet a wide range of demands, such as water pollution by chemical elements.

The heterogeneity of these large volumes of data and the difficulty of organizing them has limited progress in the field of plant breeding. To respond to this problem, Gong et al. (2022) proposed a plant database architecture with multiomics features, that is, the inclusion and integration of multiple types of omics data. The GpemDB, the database developed by the authors, is based on the entity-relationship model and is divided into four main aspects: genotype, phenotype, environment, and management. In addition, a platform for viewing this information was developed to ensure management and access to users.

Cadernos de Ciência & Tecnologia, Brasília, v. 42, e27757, 2025
DOI: 10.35977/0104-1096.cct2025.v42.27757

Hai-Ping et al. (2013) analyzed data integration solutions and the technical characteristics of the enterprise service bus (ESB). The authors proposed service-oriented architecture (SOA) and an implementation approach that provides a theoretical foundation and technical support for integration and sharing of agricultural data. Mochida et al. (2013) conducted a data survey for six tree species for integration into a database called TreeTFDB, with a web-based user interface. The web database was built using MySQL, with a web interface using Perl and JavaScript running on an Apache web server.

Hainping et al. (2010) developed an Enterprise Service Bus (ESB)-based architecture for the integration and sharing of data related to the investigation of crop germplasm resources. The purpose of the research was to address issues of information fragmentation, known as 'information islands,' and to facilitate the construction of an efficient information system for the management and analysis of these agricultural resources. Ngo & Kechadi (2020) propose a method for integrating agricultural data using a constellation arrangement designed to be flexible enough to incorporate other datasets and big data models. The authors applied a relational model to identify relationships between the information and improvement in crop yield.

In Figure 4 (a), we present a text representation that formed the word cloud, showing the frequency of words that appear in the articles analyzed. On the set of documents analyzed, some words stand out, such as "database", "data", "crop", and "breeding". The word cloud generated from the literature review articles highlights essential terms that are recurrent in the field of study. Each of these words reflects concepts and areas of focus crucial for understanding and developing the domain.

Figure 4 (b) shows the annual frequency of the most frequently used keywords in the articles analyzed during the systematic review. Years with fewer than two keywords were excluded from the figure. A variation was observed over the years, with certain terms – for example, genetics, cultivation, and database – remaining consistent across most years, indicating their prominent role in scientific discussions in the field. Other terms – for example, data integration, heterogeneous database, and data model – show a recent increase, possibly reflecting technological advances or the emergence of new methodological approaches.

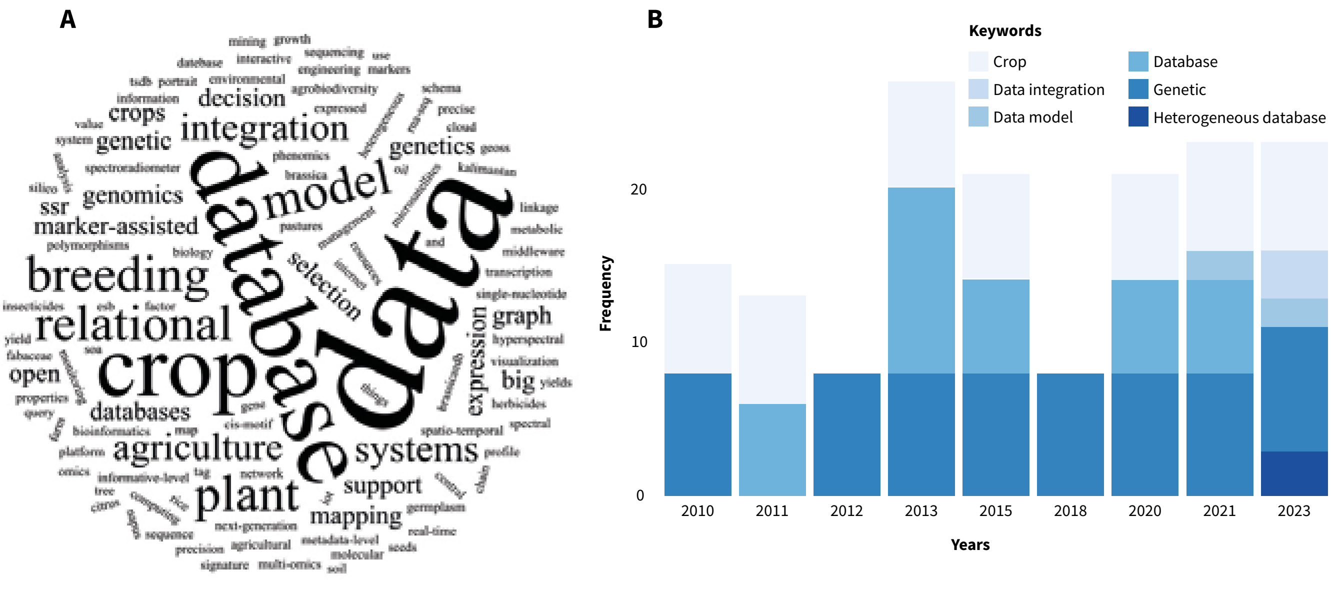

Figure 4. (A) Word cloud. (B) Temporal distribution of the articles' most used keywords.

DISCUSSION

Choosing a suitable data representation is a crucial task in developing information systems to serve the agricultural domain. The studies evaluated here dealt with different methods for creating and integrating databases to decision-making in the agricultural sector, notably in the subdomains of plant breeding, precision agriculture, agricultural management, and data modeling.

We highlight the increasing number of articles over time on the subject of this research in the studied databases. This increase may reflect the interest and evolution of the context, especially in the field of data modeling applied to agriculture. In addition, it can be considered an indicator of the maturity of the research area and institutional collaborations.

The number of articles increased over time, and the corresponding author information pointed to Switzerland having the highest number of researchers who published on database modeling in agriculture. This country may be a relevant centre of research in this area, which may be associated with its strong research and innovation system, especially in the field of data science applied to agriculture. The other countries also indicated considerable participation in the field of research studied, which may reflect the growing application of data science technologies in agriculture in countries with developed scientific and agricultural infrastructure.

We noted that one of the countries with a developed research infrastructure lead, in terms of publications. This may indicate a concentration of more recent or less developed studies on this topic or an area still under development. In addition, Switzerland focuses on innovation and technological research, where the academic publisher MDPI (Multidisciplinary Digital Publishing Institute) is located, specializing in the publication of interdisciplinary scientific journals, covering various research areas.

Overall, the geographical diversification of interest in this research area is noteworthy, with a considerable representation of countries from Europe, Asia, and Oceania indicating that database modeling applied to agriculture is a globally relevant research area. The variation in the number of publications can also be attributed to different developmental levels of agricultural research in each country.

Most of the studies aimed to develop or integrate databases to enhance research in different agricultural fields, as well as to manage agricultural practices to improve yield indices. The relational database model was most cited in the articles analyzed as it is useful for developing systems to support decision making. The relational model has significant advantages, e.g., the ease of data validation and storage (Nizar et al., 2021).

Although most of the studies analyzed use the relational model, contributing a broader understanding of the relationships between entities in agricultural databases, it is necessary to adopt an integrated approach that considers both the basic elements of traditional (relational) data modeling and the requirements of greater-volume non-relational databases. The studies analyzed indicate that integrating and relating information available in different databases is a challenge that requires a multidimensional approach (Dhanapal & Govindaraj, 2015; Cabrera & Pacheco, 2021).

Regarding the temporal distribution of the most frequently occurring keywords in the articles, the data reveal the predominant themes over the analyzed period (e.g., the development of databases addressing plant breeding), and suggest emerging trends. Identifying these keywords over the years helps to understand how agricultural data modeling has evolved, highlighting gaps in the literature and providing directions for future research.

Cadernos de Ciência & Tecnologia, Brasília, v. 42, e27757, 2025
DOI: 10.35977/0104-1096.cct2025.v42.27757

FINAL CONSIDERATIONS

We present a systematic review study with recent articles on database models in the area of agriculture. We identified approaches that involved modeling, integration, and management activities of agricultural data.

This study provides a comprehensive overview of the most recent methodologies, highlighted works that explore data models specific to agriculture. Thus, methods and techniques for building databases were presented and compared, considering the specific needs for decision making in agriculture.

The studies remain concentrated in the academic field of plant breeding, especially in genetics, as evidenced by the recurring presence of related keywords. This focus reveals a narrow perspective, highlighting the need to explore other approaches within the domain. We identified additional directions for future research that could further enhance database modeling in agriculture, due to its broad scope. Integrating databases (e.g., relational and noSQL) can introduce innovative technologies, methodologies, and approaches, offering new insights into the agriculture domain and strengthening the applicability of models.

In general, the results suggest that there is still a gap in studies that present methodologies regarding modeling of databases for agricultural data, such as yield; soil-water, plant, and atmosphere interaction parameters; and environment. Studies are still focused on the academic field of plant breeding (genetic studies). This shows the way to proceed in studies developed in research on this topic. Studies on database models developed in Brazil were not found among the articles selected.

Given this scenario of challenges and opportunities, new possibilities arise for future studies. We plan to develop a structured database for a certain type of crop based on a relational database model. From this model, it will be possible to build a system to govern and disseminate data in the form of an operational dashboard made available on the web. This system will allow online access to relevant information for scientific researchers and other interested parties.

ACKNOWLEDGMENTS

This article presents a result of the research project “Results and Impacts Evaluation of the Bean Genetic Improvement Program of the Agronomic Institute”, supported by the São Paulo Research Foundation (Fapesp), Government of the São Paulo State, Brazil – Grant 2020/01721-7, and the research project “Data modeling of bean crop production using entity-relationship models”, supported by the National Council for Scientific and Technological Development (CNPq), Brazil – Grant 444303/2018-9.

REFERENCES

AWAD, M.M.; ALAWAR, B.; JBEILY, R. A new crop spectral signatures database interactive tool (CSSIT). **Data**, v.4, art.77, 2019. DOI: https://doi.org/10.3390/data4020077.

BHAWNA; CHADUVULA, P.K.; BONTHALA, V.S.; MANJUSHA, V.; SIDDIQ, E.A.; POLUMETLA, A.K.; PRASAD, G.M.N.V. CmMDb: a versatile database for Cucumis melo microsatellite markers and other horticulture crop research. **PLoS ONE**, v.10, e0118630, 2015. DOI: https://doi.org/10.1371/journal.pone.0118630.

BHERMANA, A.; SUSILAWATI. Decision support system based on database system of genetic resources for Central Kalimantan local crops to develop ex situ and in situ conservation. **IOP Conference Series: Earth and Environmental Science**, v.482, art.012012, 2020. DOI: https://doi.org/10.1088/1755-1315/482/1/012012.

CABRERA, V.E.; FADUL-PACHECO, L. Future of dairy farming from the Dairy Brain perspective: data integration, analytics, and applications. **International Dairy Journal**, v.121, art.105069, 2021. DOI: https://doi.org/10.1016/j.idairyj.2021.105069.

CHAO, H.; LI, T.; LUO, C.; HUANG, H.; RUAN, Y.; LI, X.; NIU, Y.; FAN, Y.; SUN, W.; ZHANG, K.; LI, J.; QU, C.; LU, K. BrassicaEDB: a gene expression database for Brassica crops. **International Journal of Molecular Sciences**, v.21, art.5831, 2020. DOI: https://doi.org/10.3390/ijms21165831.

CHAO, H.; ZHANG, S.; HU, Y.; NI, Q.; XIN, S.; ZHAO, L.; IVANISENKO, V.A.; ORLOV, Y.L.; CHEN, M. Integrating omics databases for enhanced crop breeding. **Journal of Integrative Bioinformatics**, v.20, art.20230012, 2023. DOI: https://doi.org/10.1515/jib-2023-0012.

CHEN, P.P.-S. The entity-relationship model—toward a unified view of data. **ACM Transactions on Database Systems**, v.1, p.9-36, 1976. DOI: https://doi.org/10.1145/320434.320440.

CODD, E.F. A relational model of data for large shared data banks. **Communications of the ACM**, v.13, p.377-387, 1970. DOI: https://doi.org/10.1145/362384.362685.

DHANAPAL, A.P.; GOVINDARAJ, M. Unlimited thirst for genome sequencing, data interpretation, and database usage in genomic era: the road towards fast-track crop plant improvement. **Genetics Research International**, v.2015, art.684321, 2015. DOI: https://doi.org/10.1155/2015/684321.

DUHAN, N.; KAUNDAL, R. LegumeSSRdb: a comprehensive microsatellite marker database of legumes for germplasm characterization and crop improvement. **International Journal of Molecular Sciences**, v.22, art.11350, 2021. DOI: https://doi.org/10.3390/ijms222111350.

DUHAN, N.; MESHRAM, M.; LOAIZA, C.D.; KAUNDAL, R. citSATdb: genome-wide simple sequence repeat (SSR) marker database of Citrus species for germplasm characterization and crop improvement. **Genes**, v.11, art.1486, 2020. DOI: https://doi.org/10.3390/genes11121486.

GOMEZ-CANO, F.; CAREY, L.; LUCAS, K.; NAVARRETE, T.G.; MUKUNDI, E.; LUNDBACK, S.; LUNDBACK, S.; SCHNELL, D.; GROTEWOLD, E. CamRegBase: a gene regulation database for the biofuel crop, Camelina sativa. **Database**, v.2020, art.baaa075, 2020. DOI: https://doi.org/10.1093/database/baaa075.

GONG, L.; LOU, Q.; YU, C.; CHEN, Y.; HONG, J.; WU, W.; FAN, S.; CHEN, L.; LIU, C. GpemDB: a scalable database architecture with the multi-omics entity-relationship model to integrate heterogeneous big-data for precise crop breeding. **Frontiers in Bioscience-Landmark**, v.27, art.159, 2022. DOI: https://doi.org/10.31083/j.fbl2705159.

HAI-PING, S.; LEI, S.; HONG-BO, Q.; SHAO-JUN, Q.; WEI, F. Method for agriculture data integration and sharing based on SOA. **Information Technology Journal**, v.12, p.5673-5676, 2013. DOI: https://doi.org/10.3923/itj.2013.5673.5676.

HAIPING, S.; WEI, F.; PENG, T.; YONGSHENG, C. ESB-based architecture for data integration and sharing of crop germplasm resources investigation. In: INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING SYSTEMS, 2., 2010. Dalian: IEEE, 2010. p.v3-301-v3-305. DOI: https://doi.org/10.1109/ICSPS.2010.5555808.

KAMILARIS, A.; KARTAKOULLIS, A.; PRENAFETA-BOLDÚ, F.X. A review on the practice of big data analysis in agriculture. **Computers and Electronics in Agriculture**, v.143, p.23-37, 2017. DOI: https://doi.org/10.1016/j.compag.2017.09.037.

KE, T.; YU, J.; DONG, C.; MAO, H.; HUA, W.; LIU, S. ocsESTdb: a database of oil crop seed EST sequences for comparative analysis and investigation of a global metabolic network and oil accumulation metabolism. **BMC Plant Biology**, v.15, art.19, 2015. DOI: https://doi.org/10.1186/s12870-014-0399-8.

LAI, K.; LORENC, M.T.; EDWARDS, D. Genomic databases for crop improvement. **Agronomy**, v.2, p.62-73, 2012. DOI: https://doi.org/10.3390/agronomy2010062.

MARTINEZ-MOSQUERA, D.; NAVARRETE, R.; LUJAN-MORA, S. Modeling and management big data in databases—a systematic literature review. **Sustainability**, v.12, art.634, 2020. DOI: https://doi.org/10.3390/su12020634.

MENDES, C.I.C.; MARANHÃO, J. de S. de A.; BERTIN, P.R.B.; MONDO, V.H.V.; PIRES, F.C. Governança de dados para a pesquisa agrícola: segurança jurídica e autorregulação. **Cadernos de Ciência & Tecnologia**, v.40, e27209, 2023. DOI: https://doi.org/10.35977/0104-1096.cct2023.v40.27209.

MOCHIDA, K.; YOSHIDA, T.; SAKURAI, T.; YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K.; TRAN, L.-S.P. TreeTFDB: an Integrative Database of the transcription factors from six economically important tree crops for functional predictions and comparative and functional genomics. **DNA Research**, v.20, p.151-162, 2013. DOI: https://doi.org/10.1093/dnares/dss040.

MOOT, D.J.; GRIFFITHS, W.M.; CHAPMAN, D.F.; DODD, M.B.; TEIXEIRA, C.S.P. AgYields – a national database for collation of past, present and future pasture and crop yield data. **Journal of New Zealand Grasslands**, v.83, p.15-24, 2021. DOI: https://doi.org/10.33584/jnzg.2021.83.3512.

NGO, V.M.; KECHADI, M.-T. Crop knowledge discovery based on agricultural big data integration. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND SOFT COMPUTING, 4., 2020, Haiphong City. **Proceedings**. New York: Association for Computing Machinery, 2020. p.46-50.

NIKAM, P.; BHOITE, S.; SHENOY, A. Neo4j graph database implementation for LinkedIn. **International Journal of Scientific Research in Computer Science, Engineering and Information Technology**, v.6, p.339-342, 2020. DOI: https://doi.org/10.32628/CSEIT206665.

NIZAR, N.M.M.; JAHANSHIRI, E.; THARMANDRAM, A.S.; SALAMA, A.; SININ, S.S.M.; ABDULLAH, N.J.; ZOLKEPLI, H.; WIMALASIRI, E.M.; SUHAIRI, T.A.S.T.M.; HUSSIN, H.;

GREGORY, P.J.; AZAM-ALI, S.N. Underutilised crops database for supporting agricultural diversification. **Computers and Electronics in Agriculture**, v.180, art.105920, 2021. DOI: https://doi.org/10.1016/j.compag.2020.105920.

PESTANA, H.; BONACIN, R.; ROSA, F. de F.; DAMETTO, M. A review on the use of machine learning for pharmaceutical formulations. In: INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY-NEW GENERATIONS, 21., 2024, Cham. **Proceedings**. Cham: Springer, 2024. p.339-344. Editor: Shahram Latifi. ITNG 2024. DOI: https://doi.org/10.1007/978-3-031-56599-1_44.

REZNÍK, T.; CHARVÁT, K.; LUKAS, V.; CHARVÁT JR., K.; HORÁKOVÁ, Š.; KEPKA, M. Open data model for (precision) agriculture applications and agricultural pollution monitoring. In: INTERNATIONAL CONFERENCE ON INFORMATICS FOR ENVIRONMENTAL PROTECTION, 29.; INTERNATIONAL CONFERENCE ON ICT FOR SUSTAINABILITY, 3., 2015, Copenhagen. **Proceedings**. Dordrecht: Atlantis Press, 2015. p.97-107. ENVIROINFO 2015; ICT4S 2015. DOI: https://doi.org/10.2991/ict4s-env-15.2015.12.

ROMANI, L.A.S.; BAMBINI, M.D.; COLTRI, P.P.; LUCHIARI JUNIOR, A.; KOENIGKAN, L.V.; OTAVIAN, A.F.; EVANGELISTA, S.R.M.; ZULLO JR., J.; PINTO, H.S.; ASSAD, E.D. Sistema de Monitoramento Agrometeorológico - Agritempo: inovação em rede apoiando políticas públicas e a tomada de decisão agrícola. **Agrometeoros**, v.24, p.29-40, 2016. DOI: https://doi.org/10.31062/agrom.v24i1.24879.

SAN EMETERIO DE LA PARTE, M.; SERRANO, S.L.; ELDUAYEN, M.M.; MARTÍNEZ-ORTEGA, J.-F. Spatio-temporal semantic data model for precision agriculture IoT networks. **Agriculture**, v.13, art.360, 2023. DOI: https://doi.org/10.3390/agriculture13020360.

SANDERSON, L.-A.; FICKLIN, S.P.; CHENG, C.-H.; JUNG, S.; FELTUS, F.A.; BETT, K.E.; MAIN, D. Tripal v1.1: a standards-based toolkit for construction of online genetic and genomic databases. **Database**, v.2013, art.bat075, 2013. DOI: https://doi.org/10.1093/database/bat075.

SHI, Y.-X.; ZHANG, B.-K.; WANG, Y.-X.; LUO, H.-Q.; LI, X. Constructing crop portraits based on graph databases is essential to agricultural data mining. **Information**, v.12, art.227, 2021. DOI: https://doi.org/10.3390/info12060227.

SOUZA FILHO, T.A.; TAMADA, M.M.; PASSOS, R.M.; ABI-ABIB, N.P. Modelagem de banco de dados de geoprocessamento aplicado na agricultura. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE ECONOMIA, ADMINISTRAÇÃO E SOCIOLOGIA RURAL, 46., 2008, Rio Branco. [**Anais**]. Rio Branco: SOBER, 2008. DOI: https://doi.org/10.22004/ag.econ.109673.

STONEBRAKER, M.; ÇETINTEMEL, U. "One size fits all": an idea whose time has come and gone. In: INTERNATIONAL CONFERENCE ON DATA ENGINEERING, 21., 2005, Tokyo. **Proceedings**. Tokyo: IEEE, 2005. p.2-11. ICDE 2005. DOI: https://doi.org/10.1109/ICDE.2005.1.

TAIPALUS, T. Database management system performance comparisons: a systematic literature review. **Journal of Systems and Software**, v.208, art.111872, 2024a. DOI: https://doi.org/10.1016/j.jss.2023.111872.

TAIPALUS, T. Vector database management systems: fundamental concepts, use-cases, and current challenges. **Cognitive Systems Research**, v.85, art.101216, 2024b. DOI: https://doi.org/10.1016/j.cogsys.2024.101216.

TIMLIN, D.; FLEISHER, D.; TOKAY, M.; PAFF, K.; SUN, W.; BEEGUM, S.; LI, S.; WANG, Z.; REDDY, V. CLASSIM: a relational database driven crop model interface. **Smart Agricultural Technology**, v.5, art.100281, 2023. DOI: https://doi.org/10.1016/j.atech.2023.100281.

WALTERS, J.R.; LIGHT, K. The Australian digital online farm trials database increases the quality of systematic reviews and meta-analyses in grains crop research. **Crop & Pasture Science**, v.72, p.789-800, 2021. DOI: https://doi.org/10.1071/CP20534.

WHAIRIT, T.; PHADERMROD, B.; ATTASENA, V. JINDEX: JSON and index search system for plant germplasm database. **Journal of King Saud University - Computer and Information Sciences**, v.35, art.101701, 2023. DOI: https://doi.org/10.1016/j.jksuci.2023.101701.

WU, Q.; SUN, Y.; YANG, X.; LIANG, Y.; GENG, X. Study on XML-based heterogeneous agriculture database sharing platform. In: LI, D.; LIU, Y.; CHEN, Y. (Ed.). **Computer and Computing Technologies in Agriculture IV**: 4th IFIP TC 12 Conference, CCTA 2010: Nanchang, China, October 22-25, 2010: Selected Papers, Part III. Nanchang: Springer, 2011. p.1-7. DOI: https://doi.org/https://doi.org/10.1007/978-3-642-18354-6.

ZHENG, Y.; WU, S.; BAI, Y.; SUN, H.; JIAO, C.; GUO, S.; ZHAO, K.; BLANCA, J.; ZHANG, Z.; HUANG, S.; XU, Y.; WENG, Y.; MAZOUREK, M.; REDDY, U.K.; ANDO, K.; MCCREIGHT, J.D.; SCHAFFER, A.A.; BURGER, J.; TADMOR, Y.; KATZIR, N.; TANG, X.; LIU, Y.; GIOVANNONI, J.J.; LING, K.-S.; WECHTER, W.P.; LEVI, A.; GARCIA-MAS, J.; GRUMET, R.; FEI, Z. Cucurbit Genomics Database (CuGenDB): a central portal for comparative and functional genomics of cucurbit crops. **Nucleic Acids Research**, v.47, p.D1128-D1136, 2019. DOI: https://doi.org/10.1093/nar/gky944.