Data and Metadata. 2025; 4:451
doi: 10.56294/dm2025451

ORIGINAL

Design of an information system for the management, visibility, and scientific positioning in research centers: CRIS-AGROSAVIA System study case

Diseño de un sistema de información para la gestión, visibilidad y posicionamiento científico en centros de investigación: caso de estudio, Sistema CRIS-AGROSAVIA

Óscar Javier Vásquez-Casallas[1,2], Betty Jazmín Gutiérrez-Rodríguez[1,2], Carlos Arturo Bedoya Sánchez[1], Diego Hernando Flórez Martínez[1]

[1]Corporación Colombiana de Investigación Agropecuaria, Departamento de Inteligencia y Divulgación científica y Tecnológica. Mosquera, Colombia.
[2]Universitat Oberta de Catalunya. Barcelona, España.

Cite as: Vásquez-Casallas ÓJ, Gutiérrez-Rodríguez BJ, Bedoya Sánchez CA, Flórez Martínez DH. Design of an information system for the management, visibility, and scientific positioning in research centers: CRIS-AGROSAVIA System study case. Data and Metadata. 2025; 4:451. https://doi.org/10.56294/dm2025451

Submitted: 25-02-2024 **Revised:** 20-07-2024 **Accepted:** 30-11-2024 **Published:** 01-01-2025

Editor: Adrián Alejandro Vitón-Castillo

Corresponding author: Betty Jazmín Gutiérrez-Rodríguez

ABSTRACT

Introduction: this research focused on the use of technology to facilitate the management of resources, products, and knowledge services in research, development, and innovation (R+D+i) organizations. Specifically, it highlighted how Current Research Information Systems (CRIS) could be employed for this purpose.
Objective: to develop and implement a CRIS information system at the AGROSAVIA Research Center, with an emphasis on integrating the system with institutional repositories and external/internal systems to manage scientific and technological knowledge assets effectively.
Methods: the process of creating the CRIS involved several stages: planning, requirements analysis, system design, development, and implementation. Key elements included the deployment of system interfaces for the target audience (stakeholders of the National System of Science, Technology, and Innovation) and the use of a recognized data model (CERIF standard) to enhance metadata generation, ensure standardization, and enable interoperability with external and internal systems.
Results: the system was designed with two primary interfaces: a public version for the external scientific community and a corporate version for internal users of the research center. The CERIF-based data model facilitated repository structuring and the loading of an initial data baseline, supporting effective data management and decision-making processes.
Conclusion: this study provides a valuable case for those looking to build information systems for knowledge management. The CRIS developed at AGROSAVIA acted as a tool for process evaluation, scientific communication, and dissemination, offering key insights into the technological architecture, data management model, and technological deployment required for such systems.

Keywords: Current Agricultural Research Information System; Data Management; Software Development.

RESUMEN

Introducción: esta investigación se centró en el uso de la tecnología para facilitar la gestión de recursos, productos y servicios de conocimiento en organizaciones de investigación, desarrollo e innovación (I+D+i). En concreto, se puso de relieve cómo los Sistemas de Información de la Investigación Actual (CRIS) podrían emplearse con este fin.

Objetivo: desarrollar e implantar un sistema de información CRIS en el Centro de Investigación AGROSAVIA, haciendo hincapié en la integración del sistema con repositorios institucionales y sistemas externos/internos para gestionar eficazmente los activos de conocimiento científico y tecnológico.
Métodos: el proceso de creación del CRIS constó de varias fases: planificación, análisis de requisitos, diseño del sistema, desarrollo e implantación. Los elementos clave incluyeron el despliegue de interfaces del sistema para el público objetivo (partes interesadas del Sistema Nacional de Ciencia, Tecnología e Innovación) y el uso de un modelo de datos reconocido (norma CERIF) para mejorar la generación de metadatos, garantizar la normalización y permitir la interoperabilidad con sistemas externos e internos.
Resultados: el sistema se diseñó con dos interfaces principales: una versión pública para la comunidad científica externa y una versión corporativa para los usuarios internos del centro de investigación. El modelo de datos basado en CERIF facilitó la estructuración del repositorio y la carga de una línea de base de datos inicial, apoyando la gestión eficaz de los datos y los procesos de toma de decisiones.
Conclusión: este estudio proporciona un caso valioso para quienes buscan construir sistemas de información para la gestión del conocimiento. El CRIS desarrollado en AGROSAVIA actuó como herramienta para la evaluación de procesos, la comunicación científica y la divulgación, ofreciendo una visión clave de la arquitectura tecnológica, el modelo de gestión de datos y el despliegue tecnológico necesarios para este tipo de sistemas.

Palabras clave: Sistema de Información de la Investigación Agraria Actual; Gestión de Datos; Desarrollo de Software.

INTRODUCTION

Research related to the use of technology that supports the knowledge management process is a widely discussed challenge. Among the most relevant topics in knowledge management systems (KMS) research, the development of capabilities that enable the capture, organization, sharing, and utilization of knowledge in versatile interaction forms for different target audiences, comprises a key goal. To achieve this, it is necessary to reduce the effort required to collect and transform data into knowledge insights and increase the interconnection of the information sources that feed a KMS in an R+D+i institution,[1,2] which centralizes and structures knowledge assets through the interaction between institutional information repositories of research activities.[3]

Institutional repositories provide the means to properly preserve and disseminate scientific and academic output, thereby supporting knowledge management (KM) with open access technological tools.[4] However, many systems are limited to managing a specific type of content and do not connect with the frameworks in which they were created, such as researchers, organizations, funders, projects, context, and services.[5]

The relationship between research-derived products is linked to the original source, and their creation is tracked and traced through metadata synchronization via identifiers across different repositories, either manually or automatically.[6] Ideally, these identifiers should be persistent and function across multiple repositories within an institution and outside it. However, the reality is different, while codes are used to distinguish records from an information source, it is uncommon for them to match with another information system or repository (lack of equivalence structures).[7]

Instead, research data and metadata are not exclusively sourced from internal repositories; they are also recovered from scientific database systems like Scopus, Web of Science, Google Scholar, among others. Therefore, a KMS for scientific production must also be interoperable with external indexing systems that provide visibility and impact assessment of research. Thus, a KMS contributes to research trends analysis, leads to the integration of researchers with internal and external stakeholders, and associates them with the community by presenting their experts, units, and infrastructure.[8,9]

A Current Research Information System (CRIS) is a new trend that integrates all platforms supporting KM, enabling faster value creation from the growing amount of knowledge assets, and sometimes even the variety and richness of data related to research management.[10] It also provides standardized alternatives for evaluating results. A CRIS system provides the workflow for provenance control of products, processes, and technological services as well project portfolio improvement; and facilitates data discovery that strengthens the planning and execution of new research projects.[11]

CRIS systems supports the governance of innovation management activities and are situated in technologically advanced environments and highly integrated systems to meet new research challenges and increase their level of visibility.[12] For CRIS systems to be interoperable, accessible, and to ensure data exchange, it is necessary to use standard formats that include data frameworks or models, that facilitate the use and reuse of data, information, and knowledge. These should be based on open science principles with special focus on open access aims[13] that benefit both institutions and individual researchers, regardless of data heterogeneity.[14]

For this research the European CERIF standard created in 1991 was selected. CERIF comprises a metadata catalog of data dimensions and entities that generalizes the specificity and granularity for research data warehouse, and it also establishes the hierarchies and relationships between these dimensions and entities.[11] A dimension is understood as a characteristic that defines an entity, which can be temporal, locational, or another attribute that can be discretized within a range of possible values. Entities are groupings of a type of data associated with research. For example, entities that represent outcomes such as publications, multimedia products, research data, technological developments, and patents,[15] or entities that represent the origins of these outcomes, such as funding sources, events, teams, services, and organizational units, including strategic partners and researchers.[16]

In this context, the Scientific and Technological Information Management System of Agrosavia (CRIS-AGROSAVIA https://vivo.agrosavia.co/) design and implementation embraces the complete data lifecycle of the research process, through sequential phases. CRIS deployment starts with the requirements analysis phase of the organization as a research center, where technological products are mainly generated through projects.[17] Second, the conceptual design phase highlights the relevance of the CERIF standard, which ensures quality and usability compared to other data model standards.[18] Third, the logical design phase is based on a relational database structure, following a proposed general scheme that includes all the key actors and objects involved in the research process, represented as entities in the Entity-Relationship scheme.[19,20]

The implementation and presentation to the end-user phase feature two versions, one for internal users and another for external users, considering the operation and maintenance processes of the information system. This is done through ETL (Extraction, Transformation, and Load) processes that resolve conflicts from various data sources, ensuring data quality during integration.[21] Additionally, to facilitate interoperability with other systems and ensure efficient access to information, the system includes web services that expose data, allowing other platforms to consume and utilize the scientific and technological information generated.

METHOD

The methodological design was based on the phases defined for information systems development, combined with the SCRUM methodology specifically for the implementation and development phase. Thus, the results were a hybrid methodology known as Water Scrum Fall (figure 1).

Phase 1 – Planning

This phase involves gathering information and analyzing the needs of the information system and potential stakeholders/user, along with a feasibility study to determine whether the project is viable from a technical, economic, and operational (TI infrastructure) standpoint. The system is proposed as a custom development where its data model must encompass five components: i) data sources, considering quality, structure, and types of analysis; ii) technological architecture, including software and hardware for data processing, visualization, and reporting (interface); iii) human resources, comprising data and information generators and producers, users, system administrators, and technical support; iv) system processes, covering identification, encoding, storage, transfer, access, and usage flows, as well as process modeling standards, v) share and display of information through levels of aggregation and access permissions in a data model designed for the odds of user profile.

These components are structured into three feasibility studies—technical, economic, and operational, which enabled the creation of a roadmap to guide the system's development. This roadmap includes a detailed schedule of activities, resource allocation, and a risk management plan that identifies potential contingencies and defines mitigation strategies

Phase 2 – Requirements Analysis

Identifying, collecting, and recording the system's requirements using techniques like interviews with end users, observing existing workflows, and data related processes. It also includes the creation of a baseline that specifies the functional and non-functional requirements of the application, detailing what the system must do and how it should behave. The solution must include specialized search systems, customized and up-to-date reports based on the metadata structure of the main entities as functional requirements. This creates the need for intra- and inter-repository articulation, both internal and external, persistent identifiers, and stable information flows that strengthen the vertical scalability of the storage system and a relational data model that ensures consistency and integrity, as non-functional requirements.

Phase 3 - System Design

This stage focuses on the design of the scalability of the information system by adjusting the granularity of the data source and the standard, the implementation of reporting services and data downloads enriched with analysis and metrics established in the requirements phase, support for intellectual property and the protection of sensitive institutional data, data security, and the technological architecture of software and hardware applications.

The design of the CRIS information system is a critical stage in the development process, as it aims to create a solid and efficient structure for effectively managing information. This step involves defining the system architecture and the required technologies for its implementation. It also includes constructing the data model, creating data flow diagrams, entity-relationship diagrams, class models, and the user interface sketch.

For this last step, a preliminary sketch of the user interface (UI) was created to ensure the system is intuitive and easy to use. This stage included the layout of visual elements, navigation, and usability, always considering the final user experience. The initial prototypes allowed for design validation with key users before proceeding to full development, ensuring that both functionality and user experience expectations were met.

Phase 4 - Implementation and Development

In this phase, the development approach was defined to guide the implementation of the CRIS system functionalities, ensuring alignment with previously established requirements. This included selecting the tools, programming languages, and platforms best suited to the project's needs and AGROSAVIA's technological infrastructure. Key factors considered were compatibility with existing systems (Planview, Dinamix Ax, internal developments such as SIM, intra research data), scalability, and ease of maintenance.

During this phase, the system functionalities were coded, the centralized relational database was configured, and the various components making up the architecture defined in the previous phase were integrated.

Testing was conducted, starting with unit tests on each individual module or component to ensure they functioned correctly. Once the individual parts were validated, tests were carried out with researchers, management support roles, and other stakeholders, allowing them to interact with the system to verify that functionalities met the requirements and expectations established in previous phases. These tests helped identify errors, inconsistencies, or areas for improvement, leading to adjustments in the code or interface

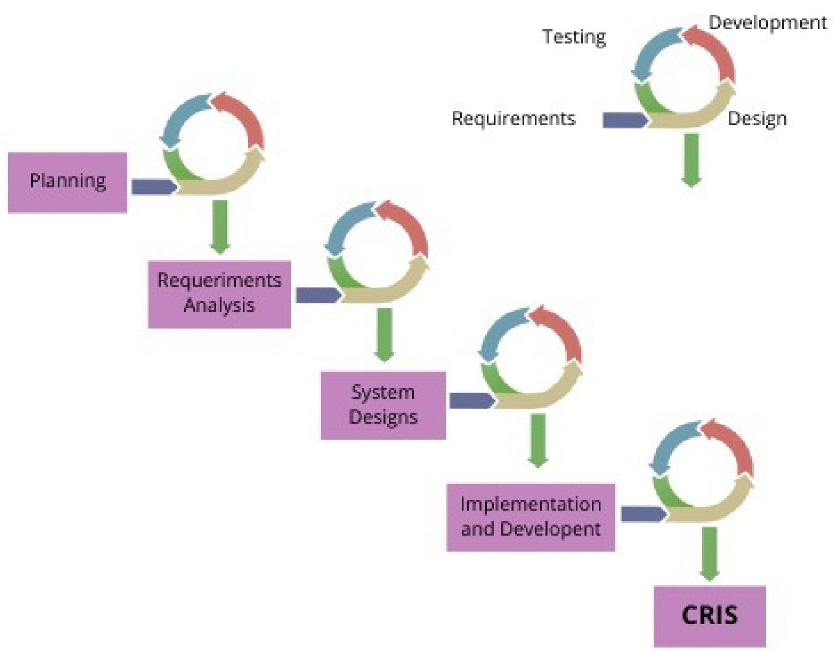

Figure 1. Methodological Design - Water-Scrum-Fall

RESULTS

On this section results are presented according to the methodological phases described on the previous section.

Phase 1 – Planning

Information gathering and requirement definition were conducted, involving various activities to understand and document the specific functional and non-functional needs and requirements of the information system. A feasibility study was conducted to determine whether the project was viable from a technical, economic, and operational standpoint.

Table 1. Roadmap by phases for the development of the information system CRIS AGROSAVIA

Phase	Activity - Milestones
Planning	Definition of the roadmap
	Inventory and analysis of data from involved information systems (internal and external)
	Feasibility study
	Consolidation of the work team
Require-ments Analysis	Meeting and interviews with the information managers of each department
	Definition of system use cases, functional and non-functional specifications
	Synchronization of internal and external data with the CERIF standard
	Baseline document
System Design	Data design document
	Architecture document, UML diagrams
	Database model construction, ERD Diagram
	Interface definition, Web application and VIVO software
	Design the internal and external data flow
	Define data protection and intellectual property guidelines: user profiles, data license, open-source repositories
Implementation and Develop-ment	ETL processes (extraction, transformation, and load) into the CRIS data ware-house - base information
	Web services exposing CRIS system information
	Installation, configuration, customization, and data loading for the vivo platform (public CRIS)
	Outputs and visualizations of the system (corporate CRIS)
	Security and audit (corporate CRIS)
	Design and development of external interoperability
	Link and implement metrics in system reports
	Testing and quality control
	Deployment and go-live

Phase 2 – Requirements Analysis

Meetings and interviews with potential stakeholders/users were held with researchers and the various departments responsible for supplying information to the system to gain a detailed understanding of their needs and expectations. Questionnaires were used to address key topics related to research management, project tracking, collaboration, and dissemination of results. The information gathered during the meetings was documented in the form of use cases, flow diagrams, and requirement specifications. These documents served as the foundation for defining the functional and non-functional requirements of the CRIS information system and for developing the baseline document.

To ensure that the requirements were clear and complete, review and validation sessions were conducted by the development team and key users involved in the process. Feedback and additional suggestions were requested to refine and improve the identified needs. Once the requirements were gathered and validated, they were prioritized and classified based on their importance and feasibility. Requirements related to modeling, open data model and data integration features were the most relevant, mainly due to the several information systems available for interoperability.[22] Finally, a product backlog was created, to approach the project's development and implementation using the SCRUM methodology.

Phase 3 - System Design

Information System Architecture

The system was defined as a multi-layer architecture, with different technologies used for the implementation of each layer:

- Data Layer: in the data layer, the repository was designed and built using a relational data model based on the CERIF metadata model, which is a standard for describing research-related information. The repository was implemented using the PostgreSQL database engine, version 14.

- Application layer: the application layer was implemented through ETL (Extraction, Transformation, and Load) processes. These processes were responsible for extracting data from various primary sources (both internal and external), categorizing the data by type, cleaning and normalizing it, and finally loading it into the repository. The ETL processes were developed using Python notebooks with the Jupyter framework.
- Presentation layer: in the presentation layer, two main interfaces were defined for information output. First, a public version of the CRIS was designed to allow access to any user via a web browser. Second, a corporate version was created to provide access exclusively to individuals belonging to AGROSAVIA, with authentication through credentials defined in the corporate Active Directory of human resources.

Data Sources

Based on the CERIF standard, ten categories of metadata were proposed, aligned with the specificities of corporate systems, the R&D&I process of the corporation, and the external resources requiring interoperability.

The data sources for the CRIS information system embraces various primary sources, such as internal databases, the human resource management system (Microsoft Dynamics AX), scientific publications (Biblioteca Agropecuaria de Colombia - BAC Digital Repository), project data (PlanView), budget management and alliances (SIM). These sources were identified during the information gathering stage and selected based on their relevance to the research center.

ETL Processes

Integration processes consist of Extraction, Transformation, and Load activities (ETL, executed through an open-source Python routine. First, metadata from each internal or external information source were collected. Data related to individuals and research projects was obtained from the corporation's own API. Data on research group associations, publications, and the national categorization of researchers were gathered by scanning the Minciencias website. Other associated data sources were consulted in specific departments or external databases such as Scopus and ORCID, using flat files in CSV format.

The recovered data was filtered and normalized according to the CERIF standard, with each field mapped to its equivalent in the standard based on the data source. Some data, such as funders, were extracted through institutional agreements specifying the roles of funders. In general, a primary entity was composed of two or more data sources. In addition to the data generated as specified in the standard, the system also gathered information from other secondary entities created for this specific data model, which were related and aimed at strengthening the system's Entity-Relationship model. This approach enriched the database and provided a more comprehensive and detailed view of scientific and technological information within AGROSAVIA.

Deduplication techniques were implemented to eliminate duplicates using persistent identifiers. When the use of identifiers was not possible, natural language processing (NLP) techniques were applied (using NLTK in Python 3.9). This involved extracting stop-words, removing accents, applying stemming (reducing a word to its base or root form by removing prefixes and suffixes),[23] vectorizing the text, and removing texts with 100 % similarity. Metadata was enriched with external data sources such as Crossref, Agriperfiles, Scienti, Google Scholar, and internal information systems like Coupa, Planview, and SharePoint. Storage of each entity and their interactions was based on a relational database in PostgreSQL and disseminated through VIVO software.[24] The data flow and its relationships can be detailed in figure 2.

System Interfaces

Once the repository was consolidated and the initial data was loaded, two main interfaces were established for information output. These interfaces are distinguished as the public version of the system and the corporate version, each offering specific access and presentation of information.

Phase 4 - Implementation and Development

The development of the CRIS information system used an iterative and incremental approach with SCRUM methodology, making it a hybrid model known as Water-Scrum-Fall. This approach allowed continuous delivery of functionalities, fostering adaptability and user engagement. A multidisciplinary team, including software, database, and information analysis experts, was involved in the development. The process was organized into two-week sprints, with planning meetings at the start of each sprint to select and estimate features. Daily stand-up meetings were held during the sprints, and sprint review meetings were conducted to showcase the developed functionalities. User involvement was encouraged through testing sessions and feedback to assess features and identify potential improvements.

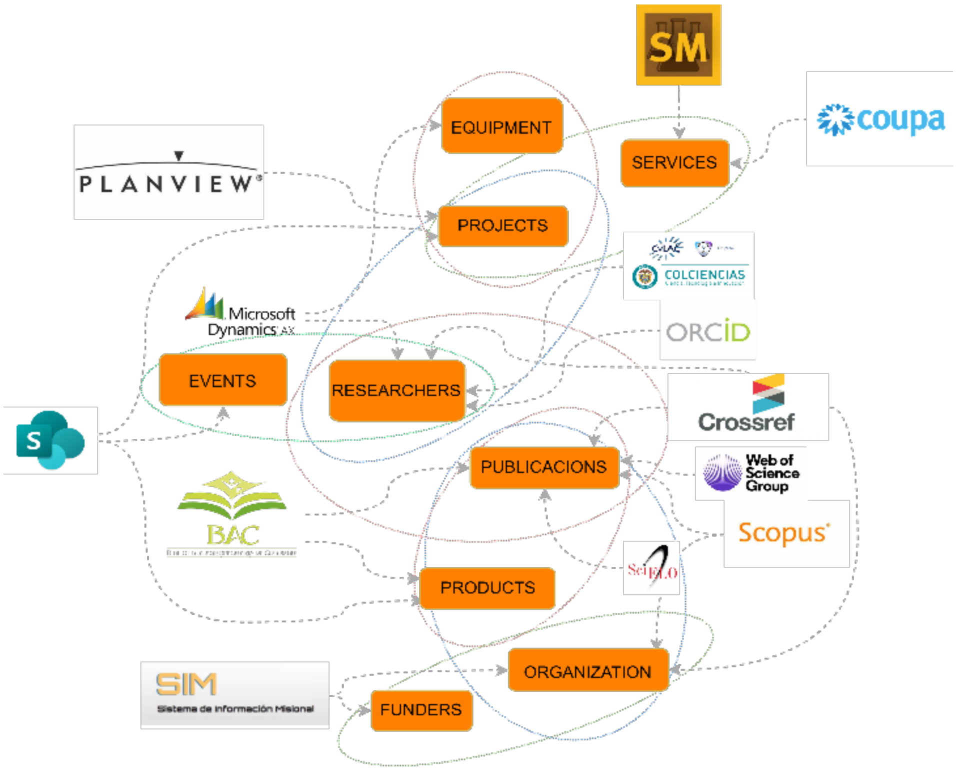

Figure 2. ETL Processes for Data Warehouse Construction

Within the CERIF standard, ten basic entities relevant to OPENAIRE are defined. The CRIS-AGROSAVIA system addressed all these entities comprehensively, including:

- 1 960 projects
- 303 publishing institutions
- 868 funders and co-executors of research projects
- 403 researchers
- 1 048 technical and scientific events
- 7 patents
- 9 458 publications
- 3 305 datasets
- 167 completed or released technological offers (grouped as products)
- 275 laboratory services
- 3 377 specialized research equipment

From these objects, additional metadata was derived, which was normalized and structured to be related to the main entities. These results are key factors that impact on the organization for the ability to manage innovation through the developments in scientific and technological innovation (STI) for supporting strategic decision-making related to undeveloped areas.[25]

- 1 071 journals
- 13 568 publication authorities
- 7 584 thematic areas, supported by the FAO AGROVOC thesaurus

Public Version of the CRIS AGROSAVIA

This version was designed to be accessible to the public, without concerns about intellectual property or business-sensitive data. The public version focuses on disseminating scientific achievements, highlighting notable projects, sharing relevant publications, and covering other aspects of interest to a broader audience. In this way, it promoted the dissemination and transparency of scientific research.

The public version of CRIS AGROSAVIA was implemented using VIVO software, which offers a broad range of features and benefits that enhance the user experience (figure 3). Additionally, VIVO enables the generation of statistics and key metrics for performance and productivity assessment of researchers, departments, and work teams. Those metrics provide valuable information for strategic decision-making and planning of future research.

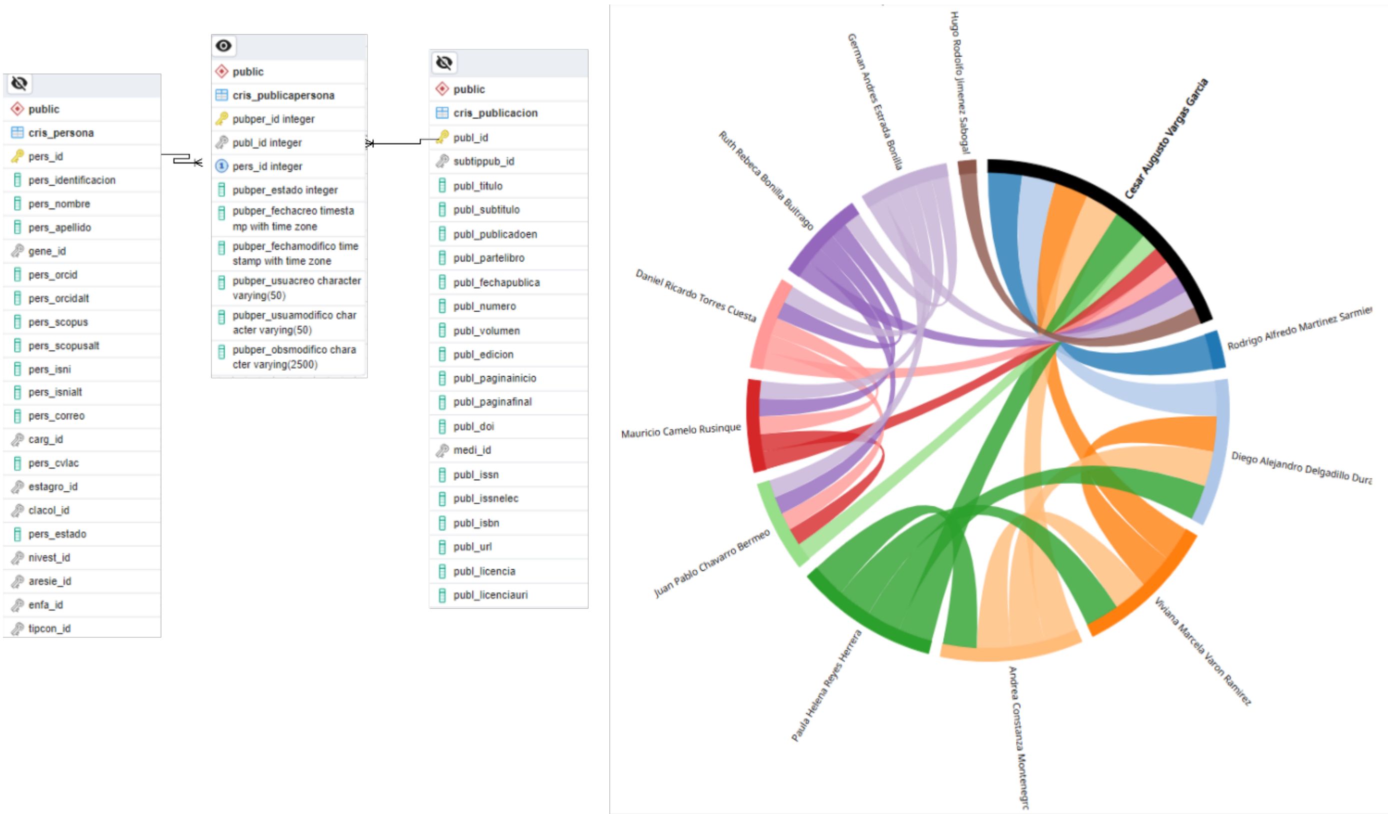

Figure 3. ERD person and publication tables, Co-authorship relationships of researchers in publications

Corporate Version of CRIS AGROSAVIA

For the development of the user interface for the internal version, .NET Core was used as a cross-platform framework, for the development of a web environment and the projection of the solution based on Android and iOS platforms tools. Additionally, this framework facilitates the organization and management of the project, the handling of necessary services, both those that feed the CRIS and those required to communicate with other corporate solutions, enabling the exchange of information with other entities in academic and research environments.

The development has adopted the MVC (Model-View-Controller) pattern, a widely recognized and proved architectural approach in web application development, which provides a clear separation of responsibilities in the CRIS architecture. This structure offers greater modularity, scalability, and maintenance features for the code.[26]

For the graphical interface, a side navigation panel was created that allows access to information on the main entities. For each entity, a dashboard has been added to highlight the most relevant details, such as the number of researchers, their areas of development, and profiles on scientific networks, as well as the complete list of information that comprises each entity in the data warehouse, for example, projects (Figure 4). Additionally, users can expand the information to find more detailed profiles, allowing them to identify the groups and scientific networks to which the researchers belong, their location according to the center where they work, and information about their formal education. The model for presenting general data and detailed information was maintained for each of the 10 entities displayed in the development.

The technologies used in the presentation layer include HTML, CSS, JavaScript, and libraries such as Bootstrap, jQuery, and Leaflet; Power BI dash boards were included for reports. The primary foundation was implemented with the free Gentella Alela template, which includes other tools such as Datatables, ChartJs, DateJS, Flot, Moment, Nprogress, and pdfMake.

Additionally, several design patterns were included to address common problems in software development and to offer best practices, thereby achieving more robust and efficient applications: Dependency Injection, Configuration Pattern, Authentication Pattern, and Object Transfer Pattern.

On the other hand, the corporate version is intended for internal use within the organization. This interface provides full access to the database and all the information stored in the repository. Researchers, decision-makers, and other authorized users can access this version to gain a more detailed and comprehensive view of research activities. Additional data, metrics, and analytical tools are presented here to facilitate strategic decision-making and the internal evaluation of research outcomes.

An example of these visualizations is the Project Execution dashboard (figure 4). The primary activity, which requires the most resources and generally demands the most effort from other administrative departments, is the execution of projects, which can be research, transfer, or capacity-building projects. These projects are assigned to a network and a center according to their technical and operational capacities, making it important for efficient planning to have up-to-date and detailed information on the workload carried by each center and network. Compared to ISIS in China,[34] CRIS AGROSAVIA favored the collection, organization, analysis and dissemination of data related to research and development projects, encouraging greater transparency and control over funding and scientific results.

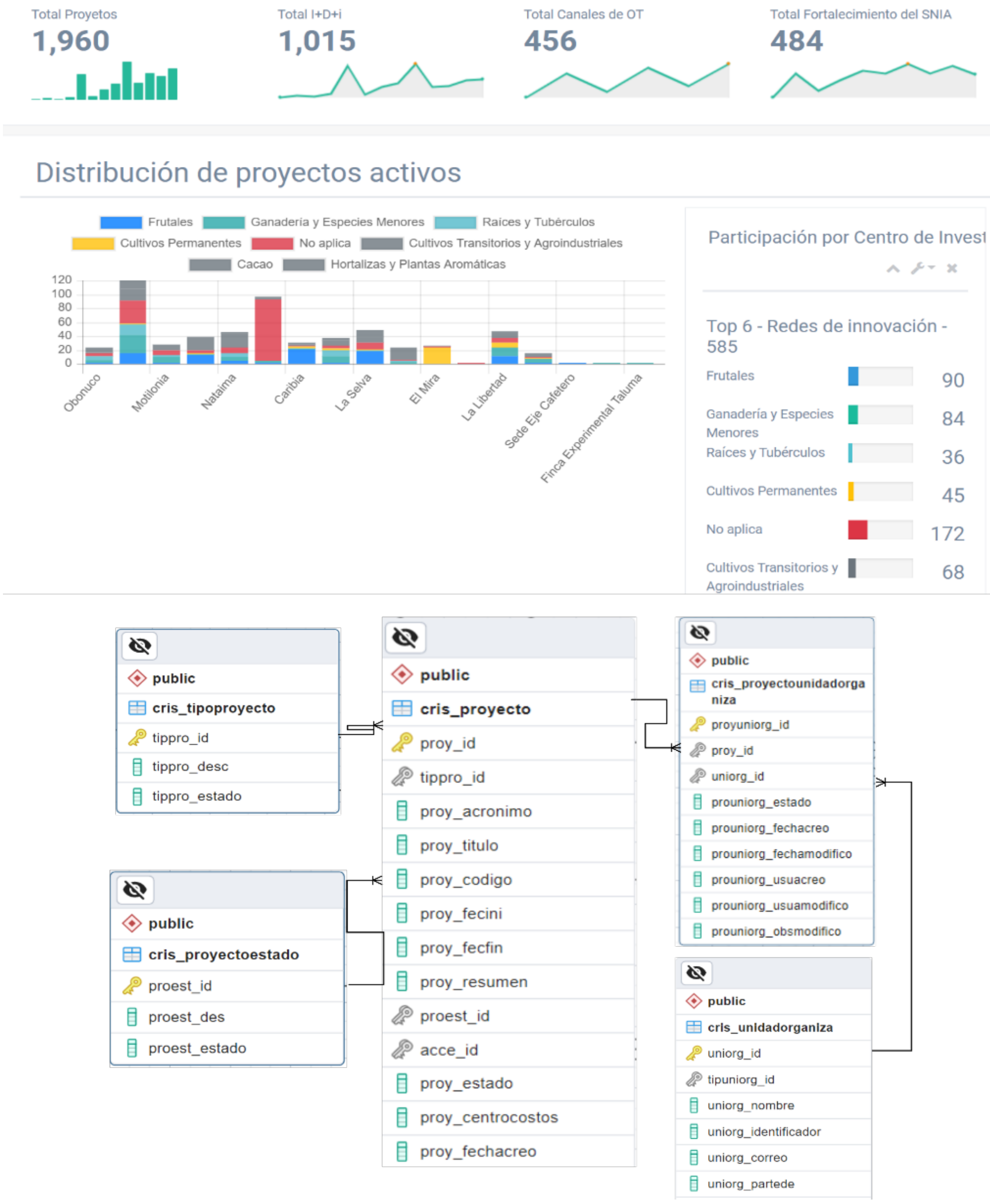

Figure 4. Project Execution, ERD project, organizational unit's tables

DISCUSSION

The institutionalization of CRIS is closely related to national and institutional quality control governance mechanisms. Furthermore, the firm establishment of an institutional CRIS infrastructure is vital for achieving integration between open science repositories and research data at the sectorial and national level.[27] This integration showcases research performance and has become essential for the national and international competitiveness of research institutions, and it is increasingly important for strategic decision-making at the executive level.[16]

The architecture of a CRIS project is composed of the following systems: Institutional Repository (IR); Research Data Repository (e-Repository); and the CRIS System, which operates on a data model based on the CERIF standard[12] These three systems interoperate with each other and with others. In our case, our IR was the BAC, the research data resides in various internal systems that are not interoperable with each other, and the model aimed to achieve an initial integration at the institutional level.

The implementation of CRIS in research institutions is technically demanding. Practical implementation first requires access to the necessary database to build the research data repository, which is often dispersed across different databases. It also requires access to internal organizational processes to collect and integrate the data, ensuring the quality and comparability of transformed inputs.[27]

According to other studies, a relational database management system (RDBMS) was used for storage, implemented with PostgreSQL software.[28] Although this model provides data, it presents a disadvantage in storing time series data due to the lack of assignment of temporal foreign keys. In our case, this is managed by changing the state of the records since the CRIS AGROSAVIA only presents information from a single institution.

Similarly, the CRIS AGROSAVIA system can be assessed in terms of data quality according to the six dimensions,[29] where accuracy and precision are related to the number of metadata and their truthfulness, respectively. These are ensured by direct connection to the original data source. A survey has been conducted to identify the persistent identifiers of researchers, which have been cross-referenced with the data registered on the Minciencias, Scopus, and ORCID platforms, managed directly by CRIS editors. This ensures data quality using different, reliable, and efficient tools for data control and correction without direct interaction with the end user until now.[11]

Another dimension is integrity, which is a strong point of the CRIS system. For the 10 entities proposed in the CERIF standard, complete and related data were recorded. Consistency, another dimension, is an area for improvement, as the heterogeneity of various data sources within the same entity prevents automatic synchronization for some entities, such as Organizational Units, and their relationships with others. Finally, reputation and timeliness are dimensions that are under constant evaluation because, although they met the initial quality standards during the design phase, they are vulnerable to changes depending on specific references in each public context, such as researchers, funders, or executive staff. To improve data quality issues in the repository context, future work will consider other topics such as the optimization of ETL processes and semantic text analysis.[14]

The outputs generated by a CRIS are associated with its interrelation with research management activities, primarily: visibility of scientific and academic output; long-term preservation of documentary heritage; channels for self-management of information; measurement and evaluation of research quality; generating reports on the current state of research within the institution; and, defining research priorities and designing research lines and approaches.[30,31] It can be concluded that the value generated by a CRIS emerge from the data model and processing data protocols, along with an interoperability framework; from having efficient tools that enable data analysis, reporting, access, and usability for decision-making, as well as addressing the specific needs of the users within the organization. In this regard, two initial versions for the CRIS AGROSAVIA were proposed:

Public version: This version focused on the visibility and positioning of the research process and offers functionalities and visualizations such as the co-authorship network, which allows for intuitive and detailed exploration and analysis of co-authors and co-investigators networks, the application web enable monitoring of the scientific competencies of researches based on an practice evaluation of published scientific results, projects performance and recent capabilities and research topics.[32]

Corporate version: This version focused on the management of knowledge assets associated with the research process. It provides customized visualizations of the data, prioritizing the specific query needs of different departments within the organization. These visualizations are presented dynamically and interactively, allowing detailed, filtered, or summarized data reports, supports the decision-making process within business analytics, it can be categorized into descriptive, predictive and prescriptive analytics.[33]

CONCLUSIONS

This system facilitates the collection, organization, centralization, and access to essential data related to projects, publications, researchers, and research outcomes, among others. It enables more informed and effective data-driven decision-making, enhancing efficiency supported by integration, organization, and usability.

Internal interoperability aids clear data flow and the creation of web services and APIs to automate data loading and connection, allowing for continuous updates and improvements that ensure the system's long-term effectiveness and utility. Additionally, using the CERIF standard will enable data exchange and flow with other CRIS systems or commercial solutions like PURE systems, promoting knowledge exchange.

In this regard, the public version functions as a tool for external exchange, given VIVO's versatility and the uniform level of detail or aggregation among all institutions that have integrated it. Meanwhile, the internal development is customized for internal reports, exposing relationships between departments and research actors, managing the interconnection of physical and intellectual capacities to drive technological developments for agriculture and livestock in Colombia.

REFERENCES

1. Jeffery KG. The New Technologies: Can CRISs Benefit? In: “Putting the Sparkle in the Knowledge Society”: Proceedings of the 7th International Conference on Current Research Information Systems (2004) [Internet]. Antwerp: euroCRIS; 2004 [cited 2024 Mar 5].

2. SMITH M, BUSI M, BALL P, VAN DER MEER R. FACTORS INFLUENCING AN ORGANISATION’S ABILITY TO MANAGE INNOVATION: A STRUCTURED LITERATURE REVIEW AND CONCEPTUAL MODEL. International Journal of Innovation Management. 2008;12(04):655–76. DOI: 10.1142/S1363919608002138

3. Iskandar K, Jambak MI, Kosala R, Prabowo H. Current Issue on Knowledge Management System for future research: A Systematic Literature Review. Procedia Comput Sci. Elsevier B.V.; 2017;116:68–80. DOI: 10.1016/J.PROCS.2017.10.011

4. González LI, Ramírez-Montoya MS, García-Peñalvo FJ, Gibrán H, Juárez EA. RITEC & CRIS: Interoperabilidad para visibilidad y medición del impacto de la producción científica energética. In: Ramírez-Montoya MS, Mendoza-Domínguez A, editors. Madrid, España; 2018 [cited 2023 Jun 13]. Available from: https://gredos.usal.es/bitstream/handle/10366/139459/Cap2.pdf?sequence=1

5. Ntini-Kounoudes A, Zervas M. The Transformation of the Ktisis Repository into a Current Research Information System (CRIS). Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices - Proceedings of the 21st International Conference on Electronic Publishing. IOS Press; 2017;227–34. DOI: 10.3233/978-1-61499-769-6-227

6. Kovačević A, Ivanović D, Milosavljević B, Konjović Z, Surla D. Automatic extraction of metadata from scientific publications for CRIS systems. Program. 2011;45(4):376–96. DOI: 10.1108/00330331111182094

7. Penca V, Nikolić S, Ivanović D, Konjović Z, Surla D. SRU/W-based CRIS systems search profile. Program. Emerald Group Publishing Ltd.; 2014;48(2):140–66. DOI: 10.1108/PROG-07-2012-0040

8. Carr-Wiggin M, Rothfus M, Barrett A, Bourne-Tyson D. Implementing a Current Research Information System (CRIS) in Canada. In: Annual IATUL Conference [Internet]. Purdue University; 2019 [cited 2023 May 10].

9. Lappalainen Y, Lassila M, Heikkilä T, Nieminen J, Lehtilä T. Migrating 120,000 Legacy Publications from Several Systems into a Current Research Information System Using Advanced Data Wrangling Techniques. Publications. 2023;11(4):49. DOI: 10.3390/publications11040049

10. Azeroual O, Schöpfel J, Ivanovic D, Nikiforova A. Combining Data Lake and Data Wrangling for Ensuring Data Quality in CRIS. Procedia Comput Sci. 2022;211:3–16. DOI: 10.1016/j.procs.2022.10.171

11. Schöpfel J, Prost H, Rebouillat V. Research Data in Current Research Information Systems. Procedia Comput Sci. Elsevier; 2017;106:305–20. DOI: 10.1016/J.PROCS.2017.03.030

12. Sales L, Sayão LF. Cyberinfrastructure information for research: A proposal for architecture for integrating repositories and CRIS systems. Informação & Sociedade [Internet]. 2015;25(3):163–84. Available from: http://www.eurocris.org/cerif/main-features-cerif

13. Medina González A, Torres Ponjuan D. Sistema de información para la investigación en el Instituto Nacional de Deportes, Educación Física y Recreación (INDER). Information system for research at the National Institute of Sports, Physical Education and Recreation (INDER). 2018 [cited 2023 Jun 12];

14. Azeroual O, Saake G, rgen Wastl J, Wastl JuergenWastl J. Data measurement in research information systems: metrics for the evaluation of data quality. Scientometrics. 2018;115:1271–90. DOI: 10.1007/s11192-018-2735-5

15. Dvořák J, Bollini A, Rémy L, Schirrwagen J. OpenAIRE Guidelines Release 1.1.1 OpenAIRE. 2018 [cited 2023 Jun 14]. Available from: https://doi.org/10.5281/zenodo.2316420

16. Scholze F, Maier J. Establishing a Research Information System as Part of an Integrated Approach to Information Management: Best Practice at the Karlsruhe Institute of Technology (KIT). LIBER Quarterly. 2012;21. DOI: 10.18352/lq.8019

17. de Castro P, Shearer K, Summann F. The Gradual Merging of Repository and CRIS Solutions to Meet Institutional Research Information Management Requirements. Procedia Comput Sci. 2014;33:39–46. DOI: 10.1016/j.procs.2014.06.007

18. Riechert MT. RIDAL – A Language for Research Information Definition Argumentation. Data Sci J. 2017;16. DOI: 10.5334/dsj-2017-005

19. Bailo D, Jeffery KG. EPOS: A Novel Use of CERIF for Data-intensive Science. Procedia Comput Sci. 2014;33:3–10. DOI: 10.1016/j.procs.2014.06.002

20. Dvořák J, Brasse V. CERIF Tutorial: Autumn 2019 Münster strategic membership meeting. In: euroCRIS [Internet]. 2019 [cited 2024 Mar 5].

21. Azeroual O, Saake G, Abuosba M. ETL Best Practices for Data Quality Checks in RIS Databases. Informatics. 2019;6(1):10. DOI: 10.3390/informatics6010010

22. Carrillo de Gea JM, Nicolás J, Fernández Alemán JL, Toval A, Ebert C, Vizcaíno A. Requirements engineering tools: Capabilities, survey and assessment. Inf Softw Technol. 2012;54(10):1142–57. DOI: 10.1016/j.infsof.2012.04.005

23. Atwan J, Mohd M, Kanaan G. Enhanced Arabic Information Retrieval: Light Stemming and Stop Words. 2013;219–28. DOI: 10.1007/978-3-642-40567-9_19

24. Salinas I, Lera I, Guerrero C, Juiz C, Sukic E. TOWARDS A SEMANTIC RESEARCH INFORMATION SYSTEM BASED ON RESEARCHERS' CV DOCUMENTS. International Journal for Quality Research. 2019;13(1):131–44. DOI: 10.24874/IJQR13.01-08

25. Liu W, Shi X, Zheng J, Li R. Characteristics of the knowledge graph of scientific and technological innovation in Gansu Province. Environ Dev Sustain. 2023;26(4):10827–43. DOI: 10.1007/s10668-023-03124-0

26. Mannaert H, Verelst J, Ven K. Towards evolvable software architectures based on systems theoretic stability. Softw Pract Exp. 2012;42(1):89–116. DOI: 10.1002/spe.1051

27. Biesenbender S, Hornbostel S. The Research Core Dataset for the German science system: developing standards for an integrated management of research information. Scientometrics. 2016;108(1):401–12. DOI: 10.1007/s11192-016-1909-2

28. Kremenjaš D, Udovičić P, Orel O. Adapting CERIF for a national CRIS: A case study. 43rd International Convention on Information, Communication and Electronic Technology (MIPRO). 2020;1633–8. DOI: 10.23919/MIPRO48935.2020.9245344.

29. Edris Abadi R, Ershadi MJ, Niaki STA. A clustering approach for data quality results of research information systems. Inf Discov Deliv. Emerald Publishing; 2022; ahead-of-print(ahead-of-print). DOI: 10.1108/IDD-07-2022-0063/FULL/PDF

30. Rybinski H, Skonieczny L, Koperwas J, Struk W, Stepniak J, Kubrak W. Integrating IR with CRIS – a novel researcher-centric approach. Program. 2017;51(3):298–321. DOI: 10.1108/PROG-04-2017-0026

31. Flórez Martínez DH, Sanchez-Torres JM, Rodríguez-Romero CA. Gestión de agendas de investigación, desarrollo tecnológico e innovación a través de procesos de gestión de conocimiento, un estudio de caso. Proceedings of the 16th LACCEI International Multi-Conference for Engineering, Education, and Technology: "Innovation in Education and Inclusion." Latin American and Caribbean Consortium of Engineering Institutions; 2018; DOI: 10.18687/LACCEI2018.1.1.1

32. Ivanovic L, Dimic Surla B, Surla D, Ivanovic D, Konjovic Z, Rudic G. Improving the discoverability of PhD

student work through a CRIS system. The Electronic Library. 2018;36(3):471-86. DOI: 10.1108/EL-05-2016-0104

33. Delen D, Demirkan H. Data, information and analytics as services. Decis Support Syst. 2013;55(1):359-63. DOI: 10.1016/j.dss.2012.05.044

34. Tian Q, Ma J, Liu O. A hybrid knowledge and model system for R&D project selection. Expert Syst Appl. 2002;23(3):265-71. DOI: 10.1016/S0957-4174(02)00046-5

FINANCING

The source of financing from Ministerio de Agricultura y Desarrollo Rural de Colombia MADR, agreement TV24

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Diego Flórez.
Data curation: Óscar Vásquez, Betty Gutiérrez.
Formal analysis: Óscar Vásquez, Betty Gutiérrez, Carlos Bedoya.
Research: Óscar Vásquez, Betty Gutiérrez, Carlos Bedoya, Diego Florez.
Methodology: Óscar Vasquéz, Carlos Bedoya.
Project management: Óscar Vásquez, Diego Florez.
Resources: Diego Florez.
Software: Óscar Vásquez, Betty Gutiérrez, Carlos Bedoya
Supervision: Diego Florez.
Validation: Óscar Vásquez, Betty Gutiérrez, Carlos Bedoya, Diego Florez.
Display: Óscar Vásquez, Carlos Bedoya.
Drafting - original draft: Óscar Vásquez, Betty Gutiérrez.
Writing - proofreading and editing: Óscar Vásquez, Betty Gutiérrez, Carlos Bedoya, Diego Florez.